

VIRAL VECTORS ENCODING RECOMBINANT FVIII VARIANTS WITH INCREASED EXPRESSION FOR GENE THERAPY OF HEMOPHILIA A

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to United States Provisional Patent Application No. 62/255,317, filed November 13, 2015, the content of which is hereby incorporated by reference in its entirety for all purposes.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 9, 2016, is named 008073_5107_WO_Sequence_Listing.txt and is 353,479 bytes in size.

BACKGROUND OF THE DISCLOSURE

[0003] Blood coagulation proceeds through a complex and dynamic biological pathway of interdependent biochemical reactions, referred to as the coagulation cascade. Coagulation Factor VIII (FVIII) is a key component in the cascade. Factor VIII is recruited to bleeding sites, and forms a Xase complex with activated Factor IX (FIXa) and Factor X (FX). The Xase complex activates FX, which in turn activates prothrombin to thrombin, which then activates other components in the coagulation cascade to generate a stable clot (reviewed in Saenko *et al.*, *Trends Cardiovasc. Med.*, 9:185-192 (1999); Lenting *et al.*, *Blood*, 92:3983-3996 (1998)).

[0004] Hemophilia A is a congenital X-linked bleeding disorder characterized by a deficiency in Factor VIII activity. Diminished Factor VIII activity inhibits a positive feedback loop in the coagulation cascade. This causes incomplete coagulation, which manifests as bleeding episodes with increased duration, extensive bruising, spontaneous oral and nasal bleeding, joint stiffness and chronic pain, and possibly internal bleeding and anemia in severe cases (Zhang *et al.*, *Clinic. Rev. Allerg. Immunol.*, 37:114-124 (2009)).

[0005] Conventionally, hemophilia A is treated by Factor VIII replacement therapy, which consists of administering Factor VIII protein (e.g., plasma-derived or recombinantly-produced Factor VIII) to an individual with hemophilia A. Factor VIII is administered prophylactically to prevent or reduce frequency of bleeding episodes, in response to an acute bleeding episode, and/or perioperatively to manage bleeding during surgery. However, there are several undesirable features of Factor VIII replacement therapy.

[0006] First, Factor VIII replacement therapy is used to treat or manage hemophilia A, but does not cure the underlying Factor VIII deficiency. Because of this, individuals with hemophilia A require Factor VIII replacement therapy for the duration of their lives. Continuous treatment is expensive and requires the individual to maintain strict compliance, as missing only a few prophylactic doses can have serious consequences for individuals with severe hemophilia A.

[0007] Second, because Factor VIII has a relatively short half-life in vivo, conventional prophylactic Factor VIII replacement therapy requires administration every second or third day. This places a burden on the individual to maintain compliance throughout their life. While third generation “long-acting” Factor VIII drugs may reduce the frequency of administration, prophylactic Factor FVIII replacement therapy with these drugs still requires monthly, weekly, or more frequent administration in perpetuity. For example, prophylactic treatment with ELOCTATE™ [Antihemophilic Factor (Recombinant), Fc Fusion Protein] requires administration every three to five days (ELOCTATE™ Prescribing Information, Biogen Idec Inc., (2015)). Moreover, the long-term effects of chemically modified biologics (e.g., pegylated polypeptides) are not yet fully understood.

[0008] Third, between 15% and 30% of all individuals receiving Factor VIII replacement therapy form anti-Factor VIII inhibitor antibodies, rendering the therapy inefficient. Factor VIII bypass therapy (e.g., administration of plasma-derived or recombinantly-produced prothrombin complex concentrates) can be used to treat hemophilia in individuals that form inhibitor antibodies. However, Factor VIII bypass therapy is less effective than Factor VIII replacement therapy (Mannucci P.M., J Thromb Haemost., 1(7):1349-55 (2003)) and may be associated with an increased risk of cardiovascular complication (Luu and Ewenstein, Haemophilia, 10 Suppl. 2:10-16 (2004)).

[0009] Somatic gene therapy holds great promise for the treatment of hemophilia A because it would remedy the underlying under-expression functional Factor VIII activity (e.g., due to missense or nonsense mutations), rather than provide a one-time dose of Factor VIII activity to the individual. Because of this difference in the mechanism of action, as compared to Factor VIII replacement therapy, one-time administration of a Factor VIII gene therapy vector may provide an individual with Factor VIII for several years, reducing the cost of treatment and eliminating the need for continued patient compliance.

[0010] Coagulation Factor IX (FIX) gene therapy has been used effectively to treat individuals with hemophilia B, a related blood coagulation condition characterized by diminished Factor IX activity (Manno C.S., et al., Nat Med., 12(3):342-47 (2006)). However, Factor VIII gene therapy presents several unique challenges. For example, the full-length, wild-type Factor VIII polypeptide (2351 amino acids; UniProt accession number P00451) is five times larger than the full-length, wild-type Factor IX polypeptide (461 amino acids; UniProt accession number P00740). As such, the coding sequence of wild-type Factor VIII is 7053 base pairs, which is too large to be packaged in conventional AAV gene therapy vectors. Further, reported recombinant expression of B-domain deleted variants of Factor VIII (BDD-FVIII) has been poor. As such, several groups have attempted to alter the codon usage of BDD-FVIII constructs, with limited success.

SUMMARY

[0011] Accordingly, there is a need for Factor VIII variants whose coding sequences are more efficiently packaged into, and delivered via, gene therapy vectors. There is also a need for synthetic, codon-altered nucleic acids which express Factor VIII more efficiently. Such Factor VIII variants and codon-altered nucleic acids allow for improved treatment of Factor VIII deficiencies (e.g., hemophilia A). The above deficiencies and other problems associated with the treatment of Factor VIII deficiencies (e.g., hemophilia A) are reduced or eliminated by the disclosed codon-altered Factor VIII variants.

[0012] In accordance with some examples, the present disclosure provides nucleic acids encoding Factor VIII variants that have high sequence identity to the disclosed codon-altered sequences of the Factor VIII heavy chain (e.g., CS01-HC-NA, CS04-HC-NA, or CS23-HC-NA)

and light chain (CS01-LC-NA, CS04-LC-NA, or CS23-LC-NA). In some examples, these nucleic acids further include a sequence encoding a linker sequence that replaces the native Factor VIII B-domain (e.g., a linker sequences comprising a furin cleavage site), between the sequences coding for the Factor VIII heavy and light chains.

[0013] In one example, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS04-HC-NA (SEQ ID NO: 3). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS04-LC-NA (SEQ ID NO: 4). The polypeptide linker comprises a furin cleavage site.

[0013a] In a first aspect, there is provided a polynucleotide comprising a nucleotide sequence encoding a Factor VIII polypeptide, the Factor VIII polypeptide comprising a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain, wherein the heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 3; wherein the light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 4; and wherein the polypeptide linker comprises a furin cleavage site and a glycosylation peptide having an amino acid sequence of SEQ ID NO:55.

[0014] In one example of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO04 (SEQ ID NO: 6).

[0015] In one example, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS01-HC-NA (SEQ ID NO: 24). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS01-LC-NA (SEQ ID NO: 25). The polypeptide linker comprises a furin cleavage site.

[0016] In one example of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO01 (SEQ ID NO: 5).

[0017] In one example, the disclosure provides a polynucleotide including a nucleotide sequence encoding a Factor VIII polypeptide. The Factor VIII polypeptide includes a light chain, a heavy

chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 95% identity to CS23-HC-NA (SEQ ID NO: 22). The light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 95% identity to CS23-LC-NA (SEQ ID NO: 23). The polypeptide linker comprises a furin cleavage site.

[0018] In one example of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to BDLO23 (SEQ ID NO: 7).

[0019] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 96% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 96% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0020] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 97% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 97% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0021] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 98% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 98% identity to the respective light chain

sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0022] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0023] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.5% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.5% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0024] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide has at least 99.9% identity to the respective heavy chain sequence (e.g., CS04-HC-NA (SEQ ID NO: 3), CS01-HC-NA (SEQ ID NO: 24), or CS23-HC-NA (SEQ ID NO: 22)), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide has at least 99.9% identity to the respective light chain sequence (e.g., CS04-LC-NA (SEQ ID NO: 4), CS01-LC-NA (SEQ ID NO: 25), or CS23-LC-NA (SEQ ID NO: 23)).

[0025] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS04-HC-NA (SEQ ID NO: 3), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS04-LC-NA (SEQ ID NO: 4).

[0026] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS01-HC-NA (SEQ ID NO: 24), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS01-LC-NA (SEQ ID NO: 25).

[0027] In one example of the polynucleotides described above, the first nucleotide sequence encoding the heavy chain of the Factor VIII polypeptide is CS23-HC-NA (SEQ ID NO: 22), and the second nucleotide sequence encoding the light chain of the Factor FVIII polypeptide is CS23-LC-NA (SEQ ID NO: 23).

[0028] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0029] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0030] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-FL-NA, wherein the polynucleotide encodes a Factor VIII polypeptide.

[0031] In one example of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0032] In one example of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0033] In one example of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0034] In one example of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0035] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0036] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-FL-NA (SEQ ID NO: 1), CS01-FL-NA (SEQ ID NO: 13), or CS23-FL-NA (SEQ ID NO: 20)).

[0037] In one example of the polynucleotides described above, the nucleotide sequence is CS04-FL-NA (SEQ ID NO: 1).

[0038] In one example of the polynucleotides described above, the nucleotide sequence is CS01-FL-NA (SEQ ID NO: 13).

[0039] In one example of the polynucleotides described above, the nucleotide sequence is CS23-FL-NA (SEQ ID NO: 20).

[0040] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 95% identity to CS04-FL-AA (SEQ ID NO: 2).

[0041] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 96% identity to CS04-FL-AA (SEQ ID NO: 2).

[0042] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 97% identity to CS04-FL-AA (SEQ ID NO: 2).

[0043] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 98% identity to CS04-FL-AA (SEQ ID NO: 2).

[0044] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99% identity to CS04-FL-AA (SEQ ID NO: 2).

[0045] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2).

[0046] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising an amino acid sequence having at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2).

[0047] In one example of the polynucleotides described above, the polynucleotide encodes a Factor VIII polypeptide comprising the amino acid sequence of CS04-FL-AA (SEQ ID NO: 2).

[0048] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0049] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0050] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0051] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0052] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0053] In one example, the disclosure provides a polynucleotide comprising a nucleotide sequence having at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29), wherein the polynucleotide encodes a single-chain Factor VIII polypeptide.

[0054] In one example of the polynucleotides described above, the nucleotide sequence has at least 96% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0055] In one example of the polynucleotides described above, the nucleotide sequence has at least 97% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0056] In one example of the polynucleotides described above, the nucleotide sequence has at least 98% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0057] In one example of the polynucleotides described above, the nucleotide sequence has at least 99% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0058] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0059] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.9% identity to the respective full-length polynucleotide sequence (e.g., CS04-SC1-NA (SEQ ID NO: 9), CS04-SC2-NA (SEQ ID NO: 11), CS01-SC1-NA (SEQ ID NO: 26), CS01-SC2-NA (SEQ ID NO: 27), CS23-SC1-NA (SEQ ID NO: 28), or CS23-SC2-NA (SEQ ID NO: 29)).

[0060] In one example of the polynucleotides described above, the nucleotide sequence is CS04-SC1-NA (SEQ ID NO: 9).

[0061] In one example of the polynucleotides described above, the nucleotide sequence is CS04-SC2-NA (SEQ ID NO: 11).

[0062] In one example of the polynucleotides described above, the nucleotide sequence is CS01-SC1-NA (SEQ ID NO: 26).

[0063] In one example of the polynucleotides described above, the nucleotide sequence is CS01-SC2-NA (SEQ ID NO: 27).

[0064] In one example of the polynucleotides described above, the nucleotide sequence is CS23-SC1-NA (SEQ ID NO: 28).

[0065] In one example of the polynucleotides described above, the nucleotide sequence is CS23-SC2-NA (SEQ ID NO: 29).

[0066] In one example of the polynucleotides described above, the nucleotide sequence has at least 95% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA,

CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0067] In one example of the polynucleotides described above, the nucleotide sequence has at least 96% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-

NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0068] In one example of the polynucleotides described above, the nucleotide sequence has at least 97% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA,

CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0069] In one example of the polynucleotides described above, the nucleotide sequence has at least 98% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-

NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0070] In one example of the polynucleotides described above, the nucleotide sequence has at least 99% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0071] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0072] In one example of the polynucleotides described above, the nucleotide sequence has at least 99.5% identity to a sequence selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA,

CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0073] In one example of the polynucleotides described above, the nucleotide sequence is selected from the group consisting of CS01-FL-NA, CS01-HC-NA, CS01-LC-NA, CS04-FL-NA, CS04-HC-NA, CS04-LC-NA, CS23-FL-NA, CS23-HC-NA, CS23-LC-NA, CS01m1-FL-NA, CS01m2-FL-NA, CS01m3-FL-NA, CS01m4-FL-NA, CS01m12-FL-NA, CS01m13-FL-NA, CS01m23-FL-NA, CS01m24-FL-NA, CS01m34-FL-NA, CS01m123-FL-NA, CS01m234-FL-NA, CS04m1-FL-NA, CS04m2-FL-NA, CS04m3-FL-NA, CS04m4-FL-NA, CS04m12-FL-NA, CS04m13-FL-NA, CS04m23-FL-NA, CS04m24-FL-NA, CS04m34-FL-NA, CS04m123-FL-NA, CS04m234-FL-NA, CS23m1-FL-NA, CS23m2-FL-NA, CS23m3-FL-NA, CS23m4-FL-

NA, CS23m12-FL-NA, CS23m13-FL-NA, CS23m23-FL-NA, CS23m24-FL-NA, CS23m34-FL-NA, CS23m123-FL-NA, CS23m234-FL-NA, CS01-SC1-NA, CS04-SC1-NA, CS23-SC1-NA, CS01m1-SC1-NA, CS01m2-SC1-NA, CS01m3-SC1-NA, CS01m4-SC1-NA, CS01m12-SC1-NA, CS01m13-SC1-NA, CS01m23-SC1-NA, CS01m24-SC1-NA, CS01m34-SC1-NA, CS01m123-SC1-NA, CS01m234-SC1-NA, CS04m1-SC1-NA, CS04m2-SC1-NA, CS04m3-SC1-NA, CS04m4-SC1-NA, CS04m12-SC1-NA, CS04m13-SC1-NA, CS04m23-SC1-NA, CS04m24-SC1-NA, CS04m34-SC1-NA, CS04m123-SC1-NA, CS04m234-SC1-NA, CS23m1-SC1-NA, CS23m2-SC1-NA, CS23m3-SC1-NA, CS23m4-SC1-NA, CS23m12-SC1-NA, CS23m13-SC1-NA, CS23m23-SC1-NA, CS23m24-SC1-NA, CS23m34-SC1-NA, CS23m123-SC1-NA, CS23m234-SC1-NA, CS01-SC2-NA, CS04-SC2-NA, CS23-SC2-NA, CS01m1-SC2-NA, CS01m2-SC2-NA, CS01m3-SC2-NA, CS01m4-SC2-NA, CS01m12-SC2-NA, CS01m13-SC2-NA, CS01m23-SC2-NA, CS01m24-SC2-NA, CS01m34-SC2-NA, CS01m123-SC2-NA, CS01m234-SC2-NA, CS04m1-SC2-NA, CS04m2-SC2-NA, CS04m3-SC2-NA, CS04m4-SC2-NA, CS04m12-SC2-NA, CS04m13-SC2-NA, CS04m23-SC2-NA, CS04m24-SC2-NA, CS04m34-SC2-NA, CS04m123-SC2-NA, CS04m234-SC2-NA, CS23m1-SC2-NA, CS23m2-SC2-NA, CS23m3-SC2-NA, CS23m4-SC2-NA, CS23m12-SC2-NA, CS23m13-SC2-NA, CS23m23-SC2-NA, CS23m24-SC2-NA, CS23m34-SC2-NA, CS23m123-SC2-NA, and CS23m234-SC2-NA.

[0074] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide comprises a glycosylation polypeptide positioned between two consecutive amino acids.

[0075] In one example of the polynucleotides described above, the encoded polypeptide linker includes a glycosylation peptide with an amino acid sequence having at least 92% identity to a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0076] In one example of the polynucleotides described above, the encoded polypeptide linker comprises a glycosylation peptide with an amino acid sequence selected from the group

consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[0077] In one example of the polynucleotides described above, the glycosylation peptide is encoded by a polynucleotide with a nucleotide sequence having at least 95% identity to a sequence selected from the group consisting of NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA, NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA and NGV-NA.

[0078] In one example of the polynucleotides described above, the glycosylation peptide is encoded by a polynucleotide with a nucleotide sequence selected from one of NG1-NA, NG4-NA, NG5-NA, NG6-NA, NG7-NA, NG9-NA, NG10-NA, NG16-NA, NG17-NA, NG18-NA, NG19-NA, NG20-NA, NG21-NA and NGV-NA.

[0079] In one example of the polynucleotides described above, the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity to a sequence selected from the group consisting of BDLNG1-NA, BDLNG3-NA, BDLNG5-NA, BDLNG6-NA, BDLNG9-NA, BDLNG10-NA, BDLNG16-NA, BDLNG17-NA, BDLNG18-NA, BDLNG19-NA, BDLNG20-NA and BDLNG21-NA.

[0080] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes an F328S (SPI, F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0081] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes I105V, A127S, G151K, M166T, and L171P (SPI; I86V, A108S, G132K, M147T, and L152P, SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0082] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some examples (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the

encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0083] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S (SPI; F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) C1918G and C1922G (SPI; C1899G and C1903 SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0084] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S (SPI; F309S SPE) amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), and b) I105V, A127S, G151K, M166T, and L171P (SPI; I86V, A108S, G132K, M147T, and L152P, SPE, respectively) amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0085] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and c) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some examples (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0086] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), b) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and c) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some examples (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0087] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), and c) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0088] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), b) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), c) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and d) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some examples (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0089] In one example of the polynucleotides described above, the encoded Factor VIII polypeptide includes a) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), b) an F328S amino acid substitution, relative to FVIII-FL-AA (SEQ ID NO: 19), c) C1918G and C1922G amino acid substitutions, relative to FVIII-FL-AA (SEQ ID NO: 19), d) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and e) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19). In some examples (e.g., where the encoded FVIII molecule includes a portion of the N-terminal region of the wild-type B-domain), the encoded Factor VIII polypeptide also includes a deletion of amino acids SF760-761, relative to FVIII-FL-AA (SEQ ID NO: 19).

[0090] In one example of the polynucleotides described above, the polynucleotide also includes a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0091] In one example of the polynucleotides described above, the polynucleotide also includes an enhancer element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0092] In one example of the polynucleotides described above, the polynucleotide also includes a polyadenylation element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

[0093] In one example of the polynucleotides described above, the polynucleotide also includes an intron operatively linked to the nucleotide sequence encoding the Factor VIII polypeptide.

[0094] In one example of the polynucleotides described above, the intron is positioned between a promoter element and the translation initiation site (e.g., the first coding ATG) of the nucleotide sequence encoding a Factor VIII polypeptide.

[0095] In another example, the disclosure provides a mammalian gene therapy vector including a polynucleotide as described above.

[0096] In one example of the mammalian gene therapy vector described above, the mammalian gene therapy vector is an adeno-associated virus (AAV) vector.

[0096a] In a second aspect, there is provided an adeno-associated virus (AAV) vector comprising a polynucleotide according to the first aspect.

[0096b] In a third aspect, there is provided adeno-associated virus (AAV) particle comprising a polynucleotide according to the first aspect.

[0096c] In a fourth aspect, there is provided a host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to the first aspect.

[0096d] In a fifth aspect, there is provided a method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to the first aspect into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

[0096e] In a sixth aspect, there is provided use of an adeno-associated virus (AAV) particle according to the third aspect in the manufacture of a medicament for treating hemophilia A in a patient in need thereof.

[0096f] In a seventh aspect, there is provided an *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to the third aspect.

[0097] In one example of the mammalian gene therapy vector described above, the AAV vector is an AAV-8 vector.

[0098] In another example, the disclosure provides a method for treating hemophilia A including administering, to a patient in need thereof, a mammalian gene therapy vector as described above.

[0099] In another example, the disclosure provides a mammalian gene therapy vector as described above for treating hemophilia A.

[00100] In another example, the disclosure provides the use of a mammalian gene therapy vector as described above for the manufacture of a medicament for treating hemophilia A.

[00101] In another example, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to

the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm23. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm23. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19).

[00102] In another example, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm123. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm123. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) an F328S amino acid substitution.

[00103] In another example, the disclosure provides a Factor VIII polypeptide including a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-HC-AAm234. The light chain of the Factor VIII polypeptide has a sequence at least 95% identical to the sequence CS01-LC-AAm234. The polypeptide linker of the Factor VIII polypeptide includes a furin cleavage site. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) F328S/C1918G/C1922G amino acid substitutions.

[00104] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 96% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 96% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00105] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 97% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 97% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00106] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 98% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 98% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00107] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 99% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 99% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00108] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence at least 99.5% identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the Factor FVIII polypeptide has a sequence at least 99.5% identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00109] In one example of the Factor VIII polypeptides described, the heavy chain of the Factor VIII polypeptide has a sequence identical to the respective heavy chain sequence (e.g., CS01-HC-AAm23, CS01-HC-AAm123, or CS01-HC-AAm234), and the light chain of the

Factor FVIII polypeptide has a sequence identical to the respective light chain sequence (e.g., CS01-LC-AAm23, CS01-LC-AAm123, or CS01-LC-AAm234).

[00110] In one example of the Factor VIII polypeptides described above, the polypeptide linker has at least 95% identity to BDL-SQ-AA (SEQ ID NO: 30).

[00111] In one example of the Factor VIII polypeptides described above, the polypeptide linker has the amino acid sequence of BDL-SQ-AA (SEQ ID NO: 30).

[00112] In one example of the Factor VIII polypeptides described above, the polypeptide linker includes a glycosylation peptide with an amino acid sequence having at least 92% identity to a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[00113] In one example of the Factor VIII polypeptides described above, the polypeptide linker includes a glycosylation peptide selected from the group consisting of NG1-AA, NG4-AA, NG5-AA, NG6-AA, NG7-AA, NG9-AA, NG10-AA, NG16-AA, NG17-AA, NG18-AA, NG19-AA, NG20-AA, NG21-AA and NGV-AA.

[00114] In one example of the Factor VIII polypeptides described above, the polypeptide linker has an amino acid sequence having at least 95% identity to a sequence selected from the group consisting of BDLNG1-AA, BDLNG3-AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-AA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA and BDLNG21-AA.

[00115] In one example of the Factor VIII polypeptides described above, the polypeptide linker has an amino acid sequence selected from the group consisting of BDLNG1-AA, BDLNG3-AA, BDLNG5-AA, BDLNG6-AA, BDLNG9-AA, BDLNG10-AA, BDLNG16-AA, BDLNG17-AA, BDLNG18-AA, BDLNG19-AA, BDLNG20-AA and BDLNG21-AA.

[00116] In another example, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm23 (SEQ ID NO: 104). The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid

substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), and iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19).

[00117] In another example, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm123. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) an F328S amino acid substitution.

[00118] In another example, the disclosure provides a Factor VIII polypeptide having an amino acid sequence with at least 95% identity to CS40-FL-AAm234. The Factor VIII polypeptide includes i) I105V, A127S, G151K, M166T, and L171P amino acid substitutions, ii) a deletion of amino acids AIEPR755-759, relative to FVIII-FL-AA (SEQ ID NO: 19), iii) an insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19), and iv) F328S/C1918G/C1922G amino acid substitutions.

[00119] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 96% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00120] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 97% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00121] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 98% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00122] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 99% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00123] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence at least 99.5% identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00124] In one example of the Factor VIII polypeptides described, the Factor VIII polypeptide has a sequence identical to the respective full-length sequence (e.g., CS40-FL-AAm23 (SEQ ID NO: 104), CS40-FL-AAm123, or CS40-FL-AAm234).

[00124a] In an eighth aspect, there is provided a polynucleotide comprising a nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO:1, wherein the polynucleotide encodes a Factor VIII polypeptide having I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19.

[0124b] In a ninth aspect, there is provided an adeno-associated virus (AAV) vector comprising a polynucleotide according to the eighth aspect.

[0124c] In a tenth aspect, there is provided an adeno-associated virus (AAV) particle comprising a polynucleotide according to the eighth aspect.

[0124d] In an eleventh aspect, there is provided a host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to the eighth aspect.

[0124e] In a twelfth aspect, there is provided a method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to the eighth aspect into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

[0124f] In a thirteenth aspect, there is provided use of an adeno-associated virus (AAV) particle according to the tenth aspect in the manufacture of a medicament for treating hemophilia A in a patient in need thereof.

[0124g] In a fourteenth aspect, there is provided an *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to the tenth aspect.

[0124h] In a fifteenth aspect, there is provided a polynucleotide comprising a nucleotide sequence encoding a Factor VIII polypeptide, the Factor VIII polypeptide comprising a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain, the Factor VIII polypeptide having I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19, wherein the heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 3; wherein the light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 4; and wherein the polypeptide linker comprises a furin cleavage site and a glycosylation peptide having an amino acid sequence of SEQ ID NO:55.

[0124i] In a sixteenth aspect, there is provided an adeno-associated virus (AAV) vector comprising a polynucleotide according to the fifteenth aspect.

[0124j] In a seventeenth aspect, there is provided an adeno-associated virus (AAV) particle comprising a polynucleotide according to the fifteenth aspect.

[0124k] In an eighteenth aspect, there is provided a host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to the fifteenth aspect.

[0124l] In a nineteenth aspect, there is provided a method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to the fifteenth aspect into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

[0124m] In a twentieth aspect, there is provided use of an adeno-associated virus (AAV) particle according to the seventeenth aspect in the manufacture of a medicament for treating hemophilia A in a patient in need thereof.

[0124n] In a twenty first aspect, there is provided an *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to the seventeenth aspect.

BRIEF DESCRIPTION OF DRAWINGS

[00125] Figure 1 shows schematic illustrations of the wild-type and ReFacto-type human Factor VIII protein constructs.

[00126] Figures 2A and 2B show the CS04 codon-altered nucleotide sequence (SEQ ID NO: 1) encoding a Factor VIII variant in accordance with some embodiments (“CS04-FL-NA” for full-length coding sequence).

[00127] Figure 3 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 2) encoded by the CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-FL-AA” for full-length amino acid sequence).

[00128] Figure 4 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 3) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS04-HC-NA”).

[00129] Figure 5 shows the portion of the CS04 codon-altered nucleotide sequence (SEQ ID NO: 4) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS04-LC-NA”).

[00130] Figure 6 shows exemplary coding sequences (SEQ ID NOS 5-7 and 36-48, respectively, in order of appearance) for B-domain substituted linkers in accordance with some embodiments. BDLO01 (SEQ ID NO: 5), BDLO04 (SEQ ID NO: 6), and BDLO23 (SEQ ID NO: 7) are the respective portions of the CS01, CS04, and CS23 codon-altered nucleotide sequences that encode a B-domain substituted linker, respectively.

[00131] Figures 7A, 7B, and 7C show an AAV vector sequence (SEQ ID NO: 8) containing an CS04 codon-altered nucleotide sequence in accordance with some embodiments (“CS04-AV-NA”).

[00132] Figures 8A and 8B show the CS01m1 codon-altered nucleotide sequence (SEQ ID NO: 49) encoding a Factor VIII variant with an F328S amino acid substitution in accordance with some embodiments (“CS01m1-FL-NA”).

[00133] Figures 9A and 9B show the CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 9) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC1-NA”).

[00134] Figure 10 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 10) encoded by the CS01Δ(760-1667) (SPI; CS01Δ(741-1648), SPE), CS04Δ(760-1667) (SPI; CS04Δ(741-1648), SPE), and CS23Δ(760-1667) (SPI; CS23Δ(741-1648), SPE) codon-altered nucleotide sequences in accordance with some embodiments (“CS01-SC1-AA,” “CS04-SC1-AA,” and “CS23-SC1-AA,” respectively).

[00135] Figures 11A and 11B show the CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 11) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS04-SC2-NA”).

[00136] Figure 12 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 12) encoded by the CS01Δ(772-1667) (SPI; CS01Δ(753-1648), SPE), CS04Δ(772-1667) (SPI; CS04Δ(753-1648), SPE), and CS23Δ(772-1667) (SPI; CS23Δ(753-1648), SPE) codon-altered nucleotide sequence in accordance with some embodiments (“CS01-SC2-AA,” “CS04-SC2-AA,” and “CS23-SC2-AA,” respectively).

[00137] Figure 13A and 13B show amino acid and nucleotide sequences for exemplary glycosylation peptides that are inserted into the B-domain substituted linker in accordance with some embodiments. “NG1” or NG1-AA” is the code for the amino acid sequence, shown in the top line. “NG1-NA” is the code for the nucleic acid sequence, shown in the bottom line for each set. Figure 13A and 13B disclose the amino acid sequences as SEQ ID NOS 51, 53, 55, 57, 59,

61, 63, 65, 67, 69, 71, 73, and 75, and the nucleotide sequences as SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, all respectively, in order of appearance.

[00138] Figure 14 shows the results of in silico prediction of in vivo N-glycosylation of the wild-type Factor VIII B-domain. Figure discloses SEQ ID NOS 76 and 76-82, respectively, in order of appearance.

[00139] Figure 15 shows the results of in silico prediction of in vivo N-glycosylation of the V3 peptide linker. Figure discloses SEQ ID NOS 83 and 83-89, respectively, in order of appearance.

[00140] Figures 16A and 16B show the CS01 codon-altered nucleotide sequence (SEQ ID NO: 13) encoding a Factor VIII variant in accordance with some embodiments (“CS01-FL-NA”).

[00141] Figures 17A and 17B show the CS08 codon-altered nucleotide sequence (SEQ ID NO: 14) encoding a Factor VIII variant in accordance with some embodiments (“CS08-FL-NA”).

[00142] Figures 18A and 18B show the CS10 codon-altered nucleotide sequence (SEQ ID NO: 15) encoding a Factor VIII variant in accordance with some embodiments (“CS10-FL-NA”).

[00143] Figures 19A and 19B show the CS11 codon-altered nucleotide sequence (SEQ ID NO: 16) encoding a Factor VIII variant in accordance with some embodiments (“CS11-FL-NA”).

[00144] Figures 20A and 20B show the CS40 wild-type ReFacto coding sequence (SEQ ID NO: 17), in accordance with some embodiments (“CS40-FL-NA”).

[00145] Figures 21A and 21B show the CH25 codon-altered nucleotide sequence (SEQ ID NO: 18) encoding a Factor VIII variant in accordance with some embodiments (“CH25-FL-NA”).

[00146] Figure 22 shows a wild-type human Factor VIII amino acid sequence (SEQ ID NO: 19), in accordance with some embodiments (“FVIII-FL-AA”).

[00147] Figure 23 illustrates the scheme for cloning the pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, and pCh25 constructs, by inserting synthetic Refacto-type BDD-FVIII DNA sequences into the vector backbone pCh-BB01 via *AscI* and *NotI* restriction sites.

[00148] Figure 24 shows the integrity of AAV vector genome preparations, as analyzed by agarose gel electrophoresis. Lane 1, DNA marker; lane 2, vCS40; lane 3, vCS01; lane 4, vCS04. The AAV vectors have all the same-sized genomes, migrating at approximately 5 kb (arrow, right side). The scale on the left side indicates size of the DNA fragments in kilobases (kb).

[00149] Figure 25 shows the protein analysis of AAV vector preparations by PAGE and silver staining. Lane 1, protein marker (M); lane 2, vCS40, lane 3, vCS01; and lane 4, vCS04. The constructs all have the same AAV8 capsids consisting of VP1, VP2, and VP3 (arrows right side). The scale on the left side indicates size of the protein marker in kilodaltons (kDa).

[00150] Figure 26A and 26B show the CS23 codon-altered nucleotide sequence (SEQ ID NO: 20) encoding a Factor VIII variant in accordance with some embodiments ("CS23-FL-NA").

[00151] Figure 27 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 21) encoded by the CS23 codon-altered nucleotide sequence in accordance with some embodiments ("CS23-FL-AA").

[00152] Figure 28 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 22) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments ("CS23-HC-NA").

[00153] Figure 29 shows the portion of the CS23 codon-altered nucleotide sequence (SEQ ID NO: 23) encoding the light chain of a Factor VIII variant in accordance with some embodiments("CS23-LC-NA").

[00154] Figures 30A and 30B show the CS01m13 codon-altered nucleotide sequence (SEQ ID NO: 90) encoding a Factor VIII variant with m1 (F328S) and m3 amino acid substitutions in accordance with some embodiments ("CS01-FL-NA-m13").

[00155] Figures 31A and 31B show the CS01m23 codon-altered nucleotide sequence (SEQ ID NO: 91) encoding a Factor VIII variant with the m2 and m3 mutation sets in accordance with some embodiments (“CS01-FL-NA-m23”).

[00156] Figures 32A and 32B show the CS01m3 codon-altered nucleotide sequence (SEQ ID NO: 92) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m3”).

[00157] Figures 33A and 33B show the CS01m2 codon-altered nucleotide sequence (SEQ ID NO: 93) encoding a Factor VIII variant with the m2 mutation set (I105V/A127S/G151K/M166T/L171P (SPI)) amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m2”).

[00158] Figures 34A and 34B show the CS04m2 codon-altered nucleotide sequence (SEQ ID NO: 94) encoding a Factor VIII variant with the m2 mutants (I105V/A127S/G151K/M166T/L171P (SPI)) amino acid substitutions in accordance with some embodiments (“CS01-FL-NA-m2”).

[00159] Figures 35A and 35B show the CS04m3 codon-altered nucleotide sequence (SEQ ID NO: 95) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m3”).

[00160] Figures 36A and 36B show the CS04m23 codon-altered nucleotide sequence (SEQ ID NO: 96) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P (SPI)) and m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m23”).

[00161] Figures 37A and 37B show the CS04m1 codon-altered nucleotide sequence (SEQ ID NO: 97) encoding a Factor VIII variant with an m1 (F328S) amino acid substitution in accordance with some embodiments (“CS04-FL-NA-m1”).

[00162] Figures 38A and 38B show the CS04m13 codon-altered nucleotide sequence (SEQ ID NO: 98) encoding a Factor VIII variant with m1 and m3 amino acid substitutions in accordance with some embodiments (“CS04-FL-NA-m13”).

[00163] Figures 39A and 39B show the CS23m13 codon-altered nucleotide sequence (SEQ ID NO: 99) encoding a Factor VIII variant with m1 and m3 amino acid substitutions in accordance with some embodiments (“CS23m13-FL-NA”)

[00164] Figures 40A and 40B show the CS23m3 codon-altered nucleotide sequence (SEQ ID NO: 100) encoding a Factor VIII variant with m3 amino acid substitutions in accordance with some embodiments (“CS23-FL-NA-m3”)

[00165] Figures 41A and 41B show the CS23m2 codon-altered nucleotide sequence (SEQ ID NO: 101) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P amino acid substitutions) in accordance with some embodiments (“CS23-FL-NA-m2”).

[00166] Figures 42A and 42B show the CS23m1 codon-altered nucleotide sequence (SEQ ID NO: 102) encoding a Factor VIII variant with an m1 (F328S) amino acid substitution in accordance with some embodiments (“CS23-FL-NA-m1”).

[00167] Figures 43A and 43B show the CS23m23 codon-altered nucleotide sequence (SEQ ID NO: 103) encoding a Factor VIII variant with the m2 mutant set (I105V/A127S/G151K/M166T/L171P) and m3 amino acid substitutions in accordance with some embodiments (“CS23-FL-NA-m23”).

[00168] Figure 44 depicts cloning of the pCS constructs, done by inserting synthetic Refacto-type BDD-FVIII carrying different mutations (see inserted table) into the vector backbone pCh-BB01 via AscI and NotI restriction sites.

[00169] Figure 45 depicts the protein analysis of AAV vector preparations by PAGE and silver staining. Lane 1, protein marker (M); lane 2, vCS01, lane 3, vCS17; lane 4, vCS19; lane 5, vCS20; lane 6, vCS40; lane 7, vCS04; lane 8, vCS17; lane 9, vCS24 construct. The constructs have all the same AAV8 capsids consisting of VP1, VP2 and VP3 (arrows right side). The scale on the left side indicates size of the protein marker in kilo Daltons (kDa).

[00170] Figure 46 depicts the integrity of AAV vector genome preparations analyzed by agarose gel electrophoresis. Lane 1, DNA marker (M); lane 2, vCS04, lane 3, vCS17; lane 4,

vCS20; lane 5, vCS24; lane 6, vCS16; lane 7, vCS40 construct. Vector load is 1.5E10 vg per lane. The AAV vectors have the same-sized genomes, migrating at approximately 5 kb (arrow, right side). The scale on the left side indicates size of the DNA fragments in kilobases (kb).

[00171] Figure 47 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 24) encoding the heavy chain of a Factor VIII variant in accordance with some embodiments (“CS01-HC-NA”).

[00172] Figure 48 shows the portion of the CS01 codon-altered nucleotide sequence (SEQ ID NO: 25) encoding the light chain of a Factor VIII variant in accordance with some embodiments (“CS01-LC-NA”).

[00173] Figures 49A and 49B show the CS01 Δ (760-1667) (SPI; CS01 Δ (741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 26) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS01-SC1-NA”).

[00174] Figures 50A and 50B show the CS01 Δ (772-1667) (SPI; CS01 Δ (753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 27) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS01-SC2-NA”).

[00175] Figures 51A and 51B show the CS23 Δ (760-1667) (SPI; CS23 Δ (741-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 28) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS23-SC1-NA”).

[00176] Figures 52A and 52B show the CS23 Δ (772-1667) (SPI; CS23 Δ (753-1648), SPE) codon-altered nucleotide sequence (SEQ ID NO: 29) encoding a single-chain Factor VIII variant in accordance with some embodiments (“CS23-SC2-NA”).

[00177] Figure 53 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 104) encoded by the CS01m23 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m23-FL-AA”).

[00178] Figure 54 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 105) encoded by the CS04m3 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m23-FL-AA”).

[00179] Figure 55 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 106) encoded by the CS01m12 codon-altered nucleotide sequence in accordance with some embodiments (“CS01m12-FL-AA”).

[00180] Figure 56 shows the Factor VIII variant amino acid sequence (SEQ ID NO: 107) encoded by the CS04m12 codon-altered nucleotide sequence in accordance with some embodiments (“CS04m12-FL-AA”).

[00181] Figures 57A and 57B show the CS01m12 codon-altered nucleotide sequence (SEQ ID NO: 108) encoding a Factor VIII variant with m1 (F328S) and m2 amino acid substitutions in accordance with some embodiments (“CS01-FL-NAm12”).

[00182] Figures 57A and 57B show the CS04m12 codon-altered nucleotide sequence (SEQ ID NO: 109) encoding a Factor VIII variant with m1 (F328S) and m2 amino acid substitutions in accordance with some embodiments (“CS04-FL-NAm12”).

DETAILED DESCRIPTION OF DISCLOSURE

I. Introduction

[00183] AAV-based gene therapy holds great promise for the treatment of hemophiliacs. For hemophilia B, first clinical data are encouraging in that FIX levels of about 10% can be maintained in at least some patients for more than 1 year. For hemophilia A however, achieving therapeutic expression levels of 5-10% with AAV vectors remains challenging for various reasons. First, the Factor VIII coding sequence is too large for conventional AAV-based vectors. Second, engineered B-domain deleted or truncated Factor VIII constructs suffer from poor expression in vivo, even when codon-optimized. Third, these B-domain deleted or truncated Factor VIII variant constructs have short half-lives in vivo, exacerbating the effects of poor expression. Fourth, even when expressed, FVIII is not efficiently secreted from cells, as are other coagulation factors, such as Factor IX.

[00184] Moreover, these challenges cannot be addressed by simply administering higher doses of the gene therapy construct. According to current knowledge, the vector dose of an AAV-based gene therapy vector should be increased above 2×10^{12} vg/kg bodyweight. This is

because at such high doses a T cell immune response is triggered, which destroys transduced cells and, as a consequence, transgene expression is reduced or even eliminated. Therefore, strategies to improve the expression of FVIII are needed to make FVIII gene therapy a viable therapeutic option for hemophilia A patients.

[00185] The present disclosure relates to the discovery of codon-altered Factor VIII variant coding sequences that solve these and other problems associated with Factor VIII gene therapy. For example, the polynucleotides disclosed herein provide markedly improved expression in mammalian cells, and display improved virion packaging due to stabilized packing interactions. In some implementations, these advantages are realized by using coding sequences for the heavy and light chains of Factor VIII with high sequence identity to the codon altered CS01, CS04, and CS23 constructs (e.g., with high sequence identity to one of the CS01-HC, CS04-HC, and CS23-HC heavy chain coding sequences and high sequence identity to one of the CS01-LC, CS04-LC, and CS23-LC light chain coding sequences).

[00186] In some implementations, the Factor VIII molecules encoded by the polynucleotides described herein have been shortened by truncating, deleting, or replacing the wild-type B-domain. As such, the polynucleotides are better suited for expressing Factor VIII via conventional gene therapy vectors, which inefficiently express larger polypeptides, such as the wild-type Factor VIII.

[00187] Advantageously, it is shown herein that the CS01, CS04, and CS23 codon-altered Factor VIII variant coding sequences provide superior expression of a B-domain deleted Factor VIII construct in vivo. For example, it is demonstrated in Example 2 and Example 4 that intravenous administration of AAV-based gene therapy vectors having the CS01 (SEQ ID NO: 13), CS04 (SEQ ID NO: 1), and CS23 (SEQ ID NO: 20) coding sequence provide 18-fold, 74-fold, and 30-fold increases in Factor VIII expression, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence (SEQ ID NO: 17), in Factor VIII knock-out mice (**Table 4** and **Table 7**).

[00188] Further, it also shown herein that the CS01 and CS04 codon-altered Factor VIII variant coding sequences provide superior virion packaging and virus production. For example, it is demonstrated in Example 1 that AAV vector constructs containing the CS01 and CS04

constructs provided 5 to 7-fold greater viral yield, relative to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence, when isolated from the same amount of cell pellet.

[00189] Advantageously, Applicants also found that the improved Factor VIII activity generated from the CS01, CS04, and CS23 codon altered sequences could be further enhanced by introducing mutations into the underlying Factor VIII polypeptide sequence. For example, as demonstrated in Example 4, the F328S, X5, and X1 mutations, alone and in combination with one another, further increased FVIII activity when expressed in vivo in the CS01 or CS04 codon altered background 2 to 7-fold, relative to the wild type, codon altered constructs (**Table 7**). More strikingly, these codon altered sequences, encoding the mutant Factor VIII mutants, provided up to 246-fold greater increase as compared to the corresponding CS40 construct encoded with the wild-type polynucleotide sequence (**Table 7**).

II. Definitions

[00190] As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[00191] As used herein, the terms “Factor VIII” and “FVIII” are used interchangeably, and refer to any protein with Factor VIII activity (e.g., active FVIII, often referred to as FVIIIa) or protein precursor (e.g., pro-protein or pre-pro-protein) of a protein with Factor VIII activity, particularly Factor IXa cofactor activity. In an exemplary embodiment, a Factor VIII polypeptide refers to a polypeptide that has sequences with high sequence identity (e.g., at least 70%, 75%, 80%, 85%, 90%, 95%, 99%, or more) to the heavy and light chains of a wild type Factor VIII polypeptide. In some embodiments, the B-domain of a Factor VIII polypeptide is deleted, truncated, or replaced with a linker polypeptide to reduce the size of the polynucleotide encoding the Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-1457 of SEQ ID NO: 2 constitute a Factor VIII polypeptide.

[00192] Non-limiting examples of wild type Factor VIII polypeptides include human pre-pro-Factor VIII (e.g., GenBank accession nos. AAA52485, CAA25619, AAA58466, AAA52484, AAA52420, AAV85964, BAF82636, BAG36452, CAI41660, CAI41666,

CAI41672, CAI43241, CAO03404, EAW72645, AAH22513, AAH64380, AAH98389, AAI11968, AAI11970, or AAB61261), corresponding pro-Factor VIII, and natural variants thereof; porcine pre-pro-Factor VIII (e.g., UniProt accession nos. F1RZ36 or K7GSZ5), corresponding pro-Factor VIII, and natural variants thereof; mouse pre-pro-Factor VIII (e.g., GenBank accession nos. AAA37385, CAM15581, CAM26492, or EDL29229), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII (e.g., GenBank accession no. AAQ21580), corresponding pro-Factor VIII, and natural variants thereof; rat pre-pro-Factor VIII; and other mammalian Factor VIII homologues (e.g., monkey, ape, hamster, guinea pig, etc.).

[00193] As used herein, a Factor VIII polypeptide includes natural variants and artificial constructs with Factor IX cofactor activity. As used in the present disclosure, Factor VIII encompasses any natural variants, alternative sequences, isoforms, or mutant proteins that retain some basal Factor IX cofactor activity (e.g., at least 5%, 10%, 25%, 50%, 75%, or more of the corresponding wild type activity). Examples of Factor VIII amino acid variations (relative to FVIII-FL-AA (SEQ ID NO: 19)) found in the human population include, without limitation, S19R, R22T, Y24C, Y25C, L26P/R, E30V, W33G, Y35C/H, G41C, R48C/K, K67E/N, L69P, E72K, D75E/V/Y, P83R, G89D/V, G92A/V, A97P, E98K, V99D, D101G/H/V, V104D, K108T, M110V, A111T/V, H113R/Y, L117F/R, G121S, E129V, G130R, E132D, Y133C, D135G/Y, T137A/I, S138R, E141K, D145H, V147D, Y155H, V159A, N163K, G164D/V, P165S, C172W, S176P, S179P, V181E/M, K185T, D186G/N/Y, S189L, L191F, G193R, L195P, C198G, S202N/R, F214V, L217H, A219D/T, V220G, D222V, E223K, G224W, T252I, V253F, N254I, G255V, L261P, P262L, G263S, G266F, C267Y, W274C, H275L, G278R, G280D, E284K, V285G, E291G/K, T294I, F295L, V297A, N299I, R301C/H/L, A303E/P, I307S, S308L, F312S, T314A/I, A315V, G323E, L326P, L327P/V, C329F, I331V, M339T, E340K, V345A/L, C348R/S/Y, Y365C, R391C/H/P, S392L/P, A394S, W401G, I405F/S, E409G, W412G/R, K427I, L431F/S, R437P/W, I438F, G439D/S/V, Y442C, K444R, Y450D/N, T454I, F455C, G466E, P470L/R/T, G474E/R/V, E475K, G477V, D478N, T479R, F484C, A488G, R490G, Y492C/H, Y492H, I494T, P496R, G498R, R503H, G513S/V, I522Y, K529E, W532G, P540T, T541S, D544N, R546W, R550C/G/H, S553P, S554C/G, V556D, R560T, D561G/H/Y, I567T, P569R, S577F, V578A, D579A/H, N583S, Q584H/K/R, I585R/T, M586V, D588G/Y, L594Q,

S596P, N601D/K, R602G, S603I/R, W604C, Y605H/S, N609I, R612C, N631K/S, M633I, S635N, N637D/I/S, Y639C, L644V, L650F, V653A/M, L659P, A663V, Q664P, F677L, M681I, V682F, Y683C/N, T686R, F698L, M699T/V, M701I, G705V, G710W, N713I, R717L/W, G720D/S, M721I/L, A723T, L725Q, V727F, E739K, Y742C, R795G, P947R, V1012L, E1057K, H1066Y, D1260E, K1289Q, Q1336K, N1460K, L1481P, A1610S, I1698T, Y1699C/F, E1701K, Q1705H, R1708C/H, T1714S, R1715G, A1720V, E1723K, D1727V, Y1728C, R1740G, K1751Q, F1762L, R1768H, G1769R, L1771P, L1775F/V, L1777P, G1779E/R, P1780L, I1782R, D1788H, M1791T, A1798P, S1799H, R1800C/G/H, P1801A, Y1802C, S1803Y, F1804S, L1808F, M1842I, P1844S, T1845P, E1848G, A1853T/V, S1858C, K1864E, D1865N/Y, H1867P/R, G1869D/V, G1872E, P1873R, L1875P, V1876L, C1877R/Y, L1882P, R1888I, E1894G, I1901F, E1904D/K, S1907C/R, W1908L, Y1909C, A1939T/V, N1941D/S, G1942A, M1945V, L1951F, R1960L/Q, L1963P, S1965I, M1966I/V, G1967D, S1968R, N1971T, H1973L, G1979V, H1980P/Y, F1982I, R1985Q, L1994P, Y1998C, G2000A, T2004R, M2007I, G2013R, W2015C, R2016P/W, E2018G, G2022D, G2028R, S2030N, V2035A, Y2036C, N2038S, 2040Y, G2045E/V, I2051S, I2056N, A2058P, W2065R, P2067L, A2070V, S2082N, S2088F, D2093G/Y, H2101D, T2105N, Q2106E/P/R, G2107S, R2109C, I2117F/S, Q2119R, F2120C/L, Y2124C, R2135P, S2138Y, T2141N, M2143V, F2145C, N2148S, N2157D, P2162L, R2169C/H, P2172L/Q/R, T2173A/I, H2174D, R2178C/H/L, R2182C/H/P, M2183R/V, L2185S/W, S2192I, C2193G, P2196R, G2198V, E2200D, I2204T, I2209N, A2211P, A2220P, P2224L, R2228G/L/P/Q, L2229F, V2242M, W2248C/S, V2251A/E, M2257V, T2264A, Q2265R, F2279C/I, I2281T, D2286G, W2290L, G2304V, D2307A, P2319L/S, R2323C/G/H/L, R2326G/L/P/Q, Q2330P, W2332R, I2336F, R2339T, G2344C/D/S, and C2345S/Y. Factor VIII proteins also include polypeptides containing post-translational modifications.

[00194] Generally, polynucleotides encoding Factor VIII encode for an inactive single-chain polypeptide (e.g., a pre-pro-protein) that undergoes post-translational processing to form an active Factor VIII protein (e.g., FVIIIa). For example, referring to Figure 1, the wild type human Factor VIII pre-pro-protein is first cleaved to release the encoded signal peptide (not shown), forming a first single-chain pro-protein (shown as “human wild-type FVIII). The pro-protein is then cleaved between the B and A3 domains to form a first polypeptide that includes

the Factor VIII heavy chain (e.g., the A1 and A2 domains) and B-domain, and a second polypeptide that includes the Factor VIII light chain (e.g., including the A3, C1, and C3 domains). The first polypeptide is further cleaved to remove the B-domain, and also to separate the A1 and A2 domains, which remain associated with the Factor VIII light chain in the mature Factor VIIIa protein. For review of the Factor VIII maturation process, see Graw et al., *Nat Rev Genet.*, 6(6):488-501 (2005), the content of which is incorporated herein by reference in its entirety for all purposes.

[00195] However, in some embodiments, the Factor VIII polypeptide is a single-chain Factor VIII polypeptide. Single-chain Factor VIII polypeptides are engineered to remove natural cleavage sites, and optionally remove, truncate, or replace the B-domain of Factor VIII. As such, they are not matured by cleavage (other than cleavage of an optional signal and/or leader peptide), and are active as a single chain. Non-limiting examples of single-chain Factor VIII polypeptides are described in Zollner et al. (*Thromb Res*, 134(1):125-31 (2014)) and Donath et al. (*Biochem J.*, 312(1):49-55 (1995)), the disclosures of which are hereby incorporated by reference in their entireties for all purposes.

[00196] As used herein, the terms “Factor VIII heavy chain,” or simply “heavy chain,” refers to the aggregate of the A1 and A2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 20-759 of CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII heavy chain.

[00197] As used herein, the term “Factor VIII light chain,” or simply “light chain,” refers to the aggregate of the A3, C1, and C2 domains of a Factor VIII polypeptide. In an exemplary embodiment, amino acids 774-1457 CS04-FL-AA (SEQ ID NO: 2) constitute a Factor VIII light chain. In some embodiments, a Factor VIII light chain excludes the acidic a3 peptide, which is released during maturation in vivo.

[00198] Generally, Factor VIII heavy and light chains are expressed as a single polypeptide chain, e.g., along with an optional B-domain or B-domain substituted linker. However, in some embodiments, a Factor VIII heavy chain and Factor VIII light chain are expressed as separate polypeptide chains (e.g., co-expressed), and reconstituted to form a Factor VIII protein (e.g., in vivo or in vitro).

[00199] As used herein, the terms “B-domain substituted linker” and “Factor VIII linker” are used interchangeably, and refer to truncated versions of a wild type Factor VIII B-domain (e.g., amino acids 760-1667 of FVIII-FL-AA (SEQ ID NO: 19)) or peptides engineered to replace the B-domain of a Factor VIII polypeptide. As used herein, a Factor VIII linker is positioned between the C-terminus of a Factor VIII heavy chain and the N-terminus of a Factor VIII light chain in a Factor VIII variant polypeptide in accordance with some embodiments. Non-limiting examples of B-domain substituted linkers are disclosed in U.S. Patent Nos. 4,868,112, 5,112,950, 5,171,844, 5,543,502, 5,595,886, 5,610,278, 5,789,203, 5,972,885, 6,048,720, 6,060,447, 6,114,148, 6,228,620, 6,316,226, 6,346,513, 6,458,563, 6,924,365, 7,041,635, and 7,943,374; U.S. Patent Application Publication Nos. 2013/024960, 2015/0071883, and 2015/0158930; and PCT Publication Nos. WO 2014/064277 and WO 2014/127215, the disclosures of which are hereby incorporated by reference, in their entireties, for all purposes.

[00200] Unless otherwise specified herein, the numbering of Factor VIII amino acids refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in Figure 22. As such, when referring to an amino acid substitution in a Factor VIII variant protein disclosed herein, the recited amino acid number refers to the analogous (e.g., structurally or functionally equivalent) and/or homologous (e.g., evolutionarily conserved in the primary amino acid sequence) amino acid in the full-length, wild-type Factor VIII sequence. For example, a T2105N amino acid substitution refers to a T to N substitution at position 2105 of the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA; SEQ ID NO: 19), a T to N substitution at position 1211 of the Factor VIII variant protein encoded by CS04 (CS04-FL-AA; SEQ ID NO: 2), and a T to N substitution at position 1212 of the Factor VIII variant encoded by CS04m3 (CS04m3-FL-AA; SEQ ID NO: 105).

[00201] As described herein, the Factor VIII amino acid numbering system is dependent on whether the Factor VIII signal peptide (e.g., amino acids 1-19 of the full-length, wild-type human Factor VIII sequence) is included. Where the signal peptide is included, the numbering is referred to as “signal peptide inclusive” or “SPI”. Where the signal peptide is not included, the numbering is referred to as “signal peptide exclusive” or “SPE.” For example, F328S is SPI numbering for the same amino acid as F309S, in SPE numbering. Unless otherwise indicated, all

amino acid numbering refers to the corresponding amino acid in the full-length, wild-type human Factor VIII sequence (FVIII-FL-AA), presented as SEQ ID NO: 19 in Figure 22.

[00202] As described herein, the codon-altered polynucleotides provide increased expression of transgenic Factor VIII in vivo (e.g., when administered as part of a gene therapy vector), as compared to the level of Factor VIII expression provided by a natively-coded Factor VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term “increased expression” refers to an increased level of transgenic Factor VIII activity in the blood of an animal administered the codon-altered polynucleotide encoding Factor VIII, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively-coded Factor VIII construct. The activity levels can be measured using any Factor VIII activity known in the art. An exemplary assay for determining Factor VIII activity is the Technochrome FVIII assay (Technoclone, Vienna, Austria).

[00203] In some embodiments, increased expression refers to at least 25% greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide. In some embodiments, increased expression refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, at least 20-fold greater, at least 25-fold greater, at least 30-fold greater, at least 40-fold greater, at least 50-fold greater, at least 60-fold greater, at least 70-fold greater, at least 80-fold greater, at least 90-fold greater, at least 100-fold greater, at least 125-fold greater, at least 150-fold greater, at least 175-fold greater, at least 200-fold greater, at least 225-fold greater, or at least 250-fold greater transgenic Factor VIII activity in the blood of an animal administered the codon-altered Factor VIII polynucleotide, as compared to the level of transgenic Factor VIII activity in the blood of an animal administered a natively coded Factor VIII polynucleotide.

[00204] As described herein, the codon-altered polynucleotides provide increased vector production, as compared to the level of vector production provided by a natively-coded Factor

VIII construct (e.g., a polynucleotide encoding the same Factor VIII construct using the wild-type human codons). As used herein, the term "increased virus production" refers to an increased vector yield in cell culture (e.g., titer per liter culture) inoculated with the codon-altered polynucleotide encoding Factor VIII, as compared to the vector yield in cell culture inoculated with a natively-coded Factor VIII construct. The vector yields can be measured using any vector titer assay known in the art. An exemplary assay for determining vector yield (e.g., of an AAV vector) is qPCR targeting the AAV2 inverted terminal repeats (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)).

[00205] In some embodiments, increased virus production refers to at least 25% greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture. In some embodiments, increased vector production refers to at least 50% greater, at least 75% greater, at least 100% greater, at least 3-fold greater, at least 4-fold greater, at least 5-fold greater, at least 6-fold greater, at least 7-fold greater, at least 8-fold greater, at least 9-fold greater, at least 10-fold greater, at least 15-fold greater, or at least 20-fold greater codon-altered vector yield, as compared to the yield of a natively-coded Factor VIII construct in the same type of culture.

[00206] As used herein, the term "hemophilia" refers to a group of disease states broadly characterized by reduced blood clotting or coagulation. Hemophilia may refer to Type A, Type B, or Type C hemophilia, or to the composite of all three diseases types. Type A hemophilia (hemophilia A) is caused by a reduction or loss of factor VIII (FVIII) activity and is the most prominent of the hemophilia subtypes. Type B hemophilia (hemophilia B) results from the loss or reduction of factor IX (FIX) clotting function. Type C hemophilia (hemophilia C) is a consequence of the loss or reduction in factor XI (FXI) clotting activity. Hemophilia A and B are X-linked diseases, while hemophilia C is autosomal. Conventional treatments for hemophilia include both prophylactic and on-demand administration of clotting factors, such as FVIII, FIX, including Bebulin®-VH, and FXI, as well as FEIBA-VH, desmopressin, and plasma infusions.

[00207] As used herein, the term "FVIII gene therapy" includes any therapeutic approach of providing a nucleic acid encoding Factor VIII to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. The

term encompasses administering any compound, drug, procedure, or regimen comprising a nucleic acid encoding a Factor VIII molecule, including any modified form of Factor VIII (e.g., Factor VIII variant), for maintaining or improving the health of an individual with hemophilia. One skilled in the art will appreciate that either the course of FVIII therapy or the dose of a FVIII therapeutic agent can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[00208] As used herein, the term "bypass therapy" includes any therapeutic approach of providing non-Factor VIII hemostatic agents, compounds or coagulation factors to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. Non-Factor VIII compounds and coagulation factors include, but are not limited to, Factor VIII Inhibitor Bypass Activity (FEIBA), recombinant activated factor VII (FVIIa), prothrombin complex concentrates, and activated prothrombin complex concentrates. These non-Factor VIII compounds and coagulation factors may be recombinant or plasma-derived. One skilled in the art will appreciate that either the course of bypass therapy or the dose of bypass therapy can be changed, e.g., based upon the results obtained in accordance with the present disclosure.

[00209] As used herein, a "combination therapy" including administration of a nucleic acid encoding a Factor VIII molecule and a conventional hemophilia A therapeutic agent includes any therapeutic approach of providing both a nucleic acid encoding a Factor VIII molecule and a Factor VIII molecule and/or non-Factor VIII hemostatic agent (e.g., bypass therapeutic agent) to a patient to relieve, diminish, or prevent the reoccurrence of one or more symptoms (e.g., clinical factors) associated with hemophilia. The term encompasses administering any compound, drug, procedure, or regimen including a nucleic acid encoding a Factor VIII molecule, including any modified form of factor VIII, which is useful for maintaining or improving the health of an individual with hemophilia and includes any of the therapeutic agents described herein.

[00210] The terms "therapeutically effective amount or dose" or "therapeutically sufficient amount or dose" or "effective or sufficient amount or dose" refer to a dose that produces therapeutic effects for which it is administered. For example, a therapeutically effective amount

of a drug useful for treating hemophilia can be the amount that is capable of preventing or relieving one or more symptoms associated with hemophilia. The exact dose will depend on the purpose of the treatment, and will be ascertainable by one skilled in the art using known techniques (*see, e.g.*, Lieberman, *Pharmaceutical Dosage Forms* (vols. 1-3, 1992); Lloyd, *The Art, Science and Technology of Pharmaceutical Compounding* (1999); Pickar, *Dosage Calculations* (1999); and *Remington: The Science and Practice of Pharmacy*, 20th Edition, 2003, Gennaro, Ed., Lippincott, Williams & Wilkins).

[00211] As used herein, the term "gene" refers to the segment of a DNA molecule that codes for a polypeptide chain (e.g., the coding region). In some embodiments, a gene is positioned by regions immediately preceding, following, and/or intervening the coding region that are involved in producing the polypeptide chain (e.g., regulatory elements such as a promoter, enhancer, polyadenylation sequence, 5'-untranslated region, 3'-untranslated region, or intron).

[00212] As used herein, the term "regulatory elements" refers to nucleotide sequences, such as promoters, enhancers, terminators, polyadenylation sequences, introns, etc, that provide for the expression of a coding sequence in a cell.

[00213] As used herein, the term "promoter element" refers to a nucleotide sequence that assists with controlling expression of a coding sequence. Generally, promoter elements are located 5' of the translation start site of a gene. However, in certain embodiments, a promoter element may be located within an intron sequence, or 3' of the coding sequence. In some embodiments, a promoter useful for a gene therapy vector is derived from the native gene of the target protein (e.g., a Factor VIII promoter). In some embodiments, a promoter useful for a gene therapy vector is specific for expression in a particular cell or tissue of the target organism (e.g., a liver-specific promoter). In yet other embodiments, one of a plurality of well characterized promoter elements is used in a gene therapy vector described herein. Non-limiting examples of well-characterized promoter elements include the CMV early promoter, the β -actin promoter, and the methyl CpG binding protein 2 (MeCP2) promoter. In some embodiments, the promoter is a constitutive promoter, which drives substantially constant expression of the target protein. In other embodiments, the promoter is an inducible promoter, which drives expression of the target

protein in response to a particular stimulus (e.g., exposure to a particular treatment or agent). For a review of designing promoters for AAV-mediated gene therapy, see Gray et al. (Human Gene Therapy 22:1143-53 (2011)), the contents of which are expressly incorporated by reference in their entirety for all purposes.

[00214] As used herein, the term "vector" refers to any vehicle used to transfer a nucleic acid (e.g., encoding a Factor VIII gene therapy construct) into a host cell. In some embodiments, a vector includes a replicon, which functions to replicate the vehicle, along with the target nucleic acid. Non-limiting examples of vectors useful for gene therapy include plasmids, phages, cosmids, artificial chromosomes, and viruses, which function as autonomous units of replication in vivo. In some embodiments, a vector is a viral vehicle for introducing a target nucleic acid (e.g., a codon-altered polynucleotide encoding a Factor VIII variant). Many modified eukaryotic viruses useful for gene therapy are known in the art. For example, adeno-associated viruses (AAVs) are particularly well suited for use in human gene therapy because humans are a natural host for the virus, the native viruses are not known to contribute to any diseases, and the viruses illicit a mild immune response.

[00215] As used herein, the term "CpG island" refers to a region within a polynucleotide having a statistically elevated density of CpG dinucleotides. As used herein, a region of a polynucleotide (e.g., a polynucleotide encoding a codon-altered Factor VIII protein) is a CpG island if, over a 200-base pair window: (i) the region has GC content of greater than 50%, and (ii) the ratio of observed CpG dinucleotides per expected CpG dinucleotides is at least 0.6, as defined by the relationship:

$$\frac{N[CpG]*N[length\ of\ window]}{N[C]*N[G]} \geq 0.6.$$

For additional information on methods for identifying CpG islands, see Gardiner-Garden M. et al., J Mol Biol., 196(2):261-82 (1987), the content of which is expressly incorporated herein by reference, in its entirety, for all purposes.

[00216] As used herein, the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form, and complements thereof. The term encompasses nucleic acids containing known nucleotide analogs or modified

backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, and peptide-nucleic acids (PNAs).

[00217] The term “amino acid” refers to naturally occurring and non-natural amino acids, including amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids include those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Naturally occurring amino acids can include, e.g., D- and L-amino acids. The amino acids used herein can also include non-natural amino acids. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., any carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, or methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally occurring amino acid. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[00218] The nucleotide sequences that encode the mutant Factor VIII constructs herein may be identical to the coding sequence provided herein or may be a different coding sequence, which sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptides as the coding sequences provided herein. One of ordinary skill in the art will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each variation of a nucleic acid which

encodes a same polypeptide is implicit in each described sequence with respect to the expression product, but not with respect to actual gene therapy constructs.

[00219] As to amino acid sequences, one of ordinary skill in the art will recognize that individual substitutions, deletions or additions to a nucleic acid or peptide sequence that alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure.

[00220] Conservative amino acid substitutions providing functionally similar amino acids are well known in the art. Dependent on the functionality of the particular amino acid, e.g., catalytic, structural, or sterically important amino acids, different groupings of amino acid may be considered conservative substitutions for each other. **Table 1** provides groupings of amino acids that are considered conservative substitutions based on the charge and polarity of the amino acid, the hydrophobicity of the amino acid, the surface exposure/structural nature of the amino acid, and the secondary structure propensity of the amino acid.

Table 1. Groupings of conservative amino acid substitutions based on the functionality of the residue in the protein.

Important Feature	Conservative Groupings
Charge/Polarity	1. H, R, and K 2. D and E 3. C, T, S, G, N, Q, and Y 4. A, P, M, L, I, V, F, and W
Hydrophobicity	1. D, E, N, Q, R, and K 2. C, S, T, P, G, H, and Y 3. A, M, I, L, V, F, and W

Structural/Surface Exposure	1. D, E, N, Q, H, R, and K 2. C, S, T, P, A, G, W, and Y 3. M, I, L, V, and F
Secondary Structure Propensity	1. A, E, Q, H, K, M, L, and R 2. C, T, I, V, F, Y, and W 3. S, G, P, D, and N
Evolutionary Conservation	1. D and E 2. H, K, and R 3. N and Q 4. S and T 5. L, I, and V 6. F, Y, and W 7. A and G 8. M and C

[00221] The terms “identical” or percent “identity,” in the context of two or more nucleic acids or peptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e., about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection.

[00222] As is known in the art, a number of different programs may be used to identify whether a protein (or nucleic acid as discussed below) has sequence identity or similarity to a known sequence. Sequence identity and/or similarity is determined using standard techniques known in the art, including, but not limited to, the local sequence identity algorithm of Smith &

Waterman, *Adv. Appl. Math.*, 2:482 (1981), by the sequence identity alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.*, 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. U.S.A.*, 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), the Best Fit sequence program described by Devereux et al., *Nucl. Acid Res.*, 12:387-395 (1984), preferably using the default settings, or by inspection. Preferably, percent identity is calculated by FastDB based upon the following parameters: mismatch penalty of 1; gap penalty of 1; gap size penalty of 0.33; and joining penalty of 30, "Current Methods in Sequence Comparison and Analysis," *Macromolecule Sequencing and Synthesis, Selected Methods and Applications*, pp 127-149 (1988), Alan R. Liss, Inc, all of which are incorporated by reference.

[00223] An example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pair wise alignments. It may also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987); the method is similar to that described by Higgins & Sharp *CABIOS* 5:151-153 (1989), both incorporated by reference. Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

[00224] Another example of a useful algorithm is the BLAST algorithm, described in: Altschul et al., *J. Mol. Biol.* 215, 403-410, (1990); Altschul et al., *Nucleic Acids Res.* 25:3389-3402 (1997); and Karlin et al., *Proc. Natl. Acad. Sci. U.S.A.* 90:5873-5787 (1993), both incorporated by reference. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul et al., *Methods in Enzymology*, 266:460-480 (1996); [http://blast.wustl.edu/blast/ README.html](http://blast.wustl.edu/blast/README.html)]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span = 1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

[00225] An additional useful algorithm is gapped BLAST, as reported by Altschul et al., Nucl. Acids Res., 25:3389-3402, incorporated by reference. Gapped BLAST uses BLOSUM-62 substitution scores; threshold T parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of k a cost of $10+k$; Xu set to 16, and Xg set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

[00226] A % amino acid sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the “longer” sequence in the aligned region. The “longer” sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored). In a similar manner, “percent (%) nucleic acid sequence identity” with respect to the coding sequence of the polypeptides identified is defined as the percentage of nucleotide residues in a candidate sequence that are identical with the nucleotide residues in the coding sequence of the cell cycle protein. A preferred method utilizes the BLASTN module of WU-BLAST-2 set to the default parameters, with overlap span and overlap fraction set to 1 and 0.125, respectively.

[00227] The alignment may include the introduction of gaps in the sequences to be aligned. In addition, for sequences which contain either more or fewer amino acids than the protein encoded by the sequence of Figure 2 (SEQ ID NO:1), it is understood that in one embodiment, the percentage of sequence identity will be determined based on the number of identical amino acids or nucleotides in relation to the total number of amino acids or nucleotides. Thus, for example, sequence identity of sequences shorter than that shown in Figure 2 (SEQ ID NO:1), as discussed below, will be determined using the number of nucleotides in the shorter sequence, in one embodiment. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, etc.

[00228] In one embodiment, only identities are scored positively (+1) and all forms of sequence variation including gaps are assigned a value of “0”, which obviates the need for a weighted scale or parameters as described below for sequence similarity calculations. Percent sequence identity may be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the “shorter” sequence in the aligned region and

multiplying by 100. The “longer” sequence is the one having the most actual residues in the aligned region.

[00229] The term “allelic variants” refers to polymorphic forms of a gene at a particular genetic locus, as well as cDNAs derived from mRNA transcripts of the genes, and the polypeptides encoded by them. The term “preferred mammalian codon” refers a subset of codons from among the set of codons encoding an amino acid that are most frequently used in proteins expressed in mammalian cells as chosen from the following list: Gly (GGC, GGG); Glu (GAG); Asp (GAC); Val (GTG, GTC); Ala (GCC, GCT); Ser (AGC, TCC); Lys (AAG); Asn (AAC); Met (ATG); Ile (ATC); Thr (ACC); Trp (TGG); Cys (TGC); Tyr (TAT, TAC); Leu (CTG); Phe (TTC); Arg (CGC, AGG, AGA); Gln (CAG); His (CAC); and Pro (CCC).

[00230] As used herein, the term codon-altered refers to a polynucleotide sequence encoding a polypeptide (e.g., a Factor VIII variant protein), where at least one codon of the native polynucleotide encoding the polypeptide has been changed to improve a property of the polynucleotide sequence. In some embodiments, the improved property promotes increased transcription of mRNA coding for the polypeptide, increased stability of the mRNA (e.g., improved mRNA half-life), increased translation of the polypeptide, and/or increased packaging of the polynucleotide within the vector. Non-limiting examples of alterations that can be used to achieve the improved properties include changing the usage and/or distribution of codons for particular amino acids, adjusting global and/or local GC content, removing AT-rich sequences, removing repeated sequence elements, adjusting global and/or local CpG dinucleotide content, removing cryptic regulatory elements (e.g., TATA box and CCAAT box elements), removing of intron/exon splice sites, improving regulatory sequences (e.g., introduction of a Kozak consensus sequence), and removing sequence elements capable of forming secondary structure (e.g., stem-loops) in the transcribed mRNA.

[00231] As discussed herein, there are various nomenclatures to refer to components of the disclosure herein. “CS-number” (e.g. “CS04”, “CS01”, “CS23”, etc.) refer to codon altered polynucleotides encoding FVIII polypeptides and/or the encoded polypeptides, including variants. For example, CS01-FL refers to the **F**ull **L**ength codon altered CS01 polynucleotide sequence or amino acid sequence (sometimes referred to herein as “CS01-FL-AA” for the

Amino Acid sequence and “CS01-FL-NA” for the Nucleic Acid sequence) encoded by the CS01 polynucleotide sequence. Similarly, “CS01-LC” refers to either the codon altered nucleic acid sequence (“CS01-LC-NA”) encoding the light chain of a FVIII polypeptide or the amino acid sequence (also sometimes referred to herein as “CS01-LC-AA”) of the FVIII light chain encoded by the CS01 polynucleotide sequence. Likewise, CS01-HC, CS01-HC-AA and CS01-HC-NA are the same for the FVIII heavy chain. As will be appreciated by those in the art, for constructs such as CS01, CS04, CS23, etc., that are only codon-altered (e.g. they do not contain additional amino acid substitutions as compared to Refacto), the amino acid sequences will be identical, as the amino acid sequences are not altered by the codon optimization. Thus, sequence constructs of the disclosure include, but are not limited to, CS01-FL-NA, CS01-FL-AA, CS01-LC-NA, CS01-LC-AA, CS01-HC-AA, CS01-HC-NA, CS04-FL-NA, CS04-FL-AA, CS04-LC-NA, CS04-LC-AA, CS04-HC-AA, CS04-HC-NA, CS23-FL-NA, CS23-FL-AA, CS23-LC-NA, CS23-LC-AA, CS23-HC-AA and CS23-HC-NA.

[00232] This nomenclature also applies to glycosylation peptides as shown in Figure 13, such that “NGA1-AA” refers to the amino acid sequence and NGA1-NA refers to the nucleic acid sequence.

[00233] The disclosure also includes additional new Factor VIII variants, as described below, with the appropriate nomenclature.

III. Codon-Altered Factor VIII Variants

[00234] In some embodiments, the present disclosure provides codon-altered polynucleotides encoding Factor VIII variants. These codon-altered polynucleotides provide markedly improved expression of Factor VIII when administered in an AAV-based gene therapy construct. The codon-altered polynucleotides also demonstrate improved AAV-virion packaging, as compared to conventionally codon-optimized constructs. As demonstrated in Example 2 and Example 4, Applicants have achieved these advantages through the discovery of three codon-altered polynucleotides (CS01-FL-NA, CS04-FL-NA, and CS23-FL-NA) encoding a Factor VIII polypeptide with human wild-type Factor VIII heavy and light chains, and a short, 14 amino acid, B-domain substituted linker (the “SQ” linker) containing a furin cleavage site to facilitate maturation of an active FVIIIa protein in vivo. As further demonstrated in Example 4,

incorporation of various combinations of the F328S, X5, and X1 amino acid mutations into the encoded Factor VIII molecule further increased the *in vivo* expression of Factor VIII activity.

[00235] In one embodiment, a codon-altered polynucleotide provided herein has nucleotide sequences with high sequence identity to at least the sequences within CS01, CS04, or CS23 (SEQ ID NOS 13, 1, and 20, respectively) encoding the Factor VIII heavy chain and Factor VIII light chains. As known in the art, the B-domain of Factor VIII is dispensable for activity *in vivo*. Thus, in some embodiments, the codon-altered polynucleotides provided herein completely lack a Factor VIII B-domain. In some embodiments, the native Factor VIII B-domain is replaced with a short amino acid linker containing a furin cleavage site, e.g., the “SQ” linker consisting of amino acids 760-773 of the CS01, CS04, or CS23 (SEQ ID NOS 2, 2, and 21, respectively) constructs. The “SQ” linker is also referred to as BDLO04, (-AA for the amino acid sequence and -NA for the nucleotide sequence shown in Figure 6).

[00236] In one embodiment, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are human Factor VIII heavy and light chains, respectively. In other embodiments, the Factor VIII heavy and light chains encoded by the codon-altered polynucleotide are heavy and light chain sequences from another mammal (e.g., porcine Factor VIII). In yet other embodiments, the Factor VIII heavy and light chains are chimeric heavy and light chains (e.g., a combination of human and a second mammalian sequence). In yet other embodiments, the Factor VIII heavy and light chains are humanized version of the heavy and light chains from another mammal, e.g., heavy and light chain sequences from another mammal in which human residues are substituted at select positions to reduce the immunogenicity of the resulting peptide when administered to a human.

[00237] The GC content of human genes varies widely, from less than 25% to greater than 90%. However, in general, human genes with higher GC contents are expressed at higher levels. For example, Kudla et al. (PLoS Biol., 4(6):80 (2006)) demonstrate that increasing a gene’s GC content increases expression of the encoded polypeptide, primarily by increasing transcription and effecting a higher steady state level of the mRNA transcript. Generally, the desired GC content of a codon-optimized gene construct is equal or greater than 60%. However, native AAV genomes have GC contents of around 56%.

[00238] Accordingly, in some embodiments, the codon-altered polynucleotides provided herein have a CG content that more closely matches the GC content of native AAV virions (e.g., around 56% GC), which is lower than the preferred CG contents of polynucleotides that are conventionally codon-optimized for expression in mammalian cells (e.g., at or above 60% GC). As outlined in Example 1, CS04-FL-NA (SEQ ID NO: 1), which has a GC content of about 56%, has improved virion packaging as compared to similarly codon-altered coding sequences with higher GC content.

[00239] Thus, in some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 60%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is less than 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is no more than 56%.

[00240] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 59%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 58%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 56% to 57%. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 54% to 56%. In some

embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is from 55% to 56%.

[00241] In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.5\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.4\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.3\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.2\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is $56\pm 0.1\%$. In some embodiments, the overall GC content of a codon-altered polynucleotide encoding a Factor VIII polypeptide is 56%.

A. Factor VIII Amino Acid Substitutions

[00242] To further increase the efficiency of AAV-vector based expression of the Factor VIII constructs described herein, amino acid substitutions known to improve secretion, increase specific activity, and/or enhanced the stability of Factor VIII are further incorporated, in some implementations. A number of potential variants were identified that increase the plasma levels of FVIII activity at a given vector dose. These variants include those with a more efficient signal peptide, amino acid substitutions that prevent BiP interactions, amino acid substitutions resembling more efficiently secreted Factor VIII orthologs (e.g., porcine Factor VIII), single-chain Factor VIII variants, and amino acid substitutions that stabilize Factor VIII and/or reduce subunit dissociation.

[00243] Mutation of residues A108, R121, and L2302 (SPE), located at the interface between the A1 and C2 domains, increases the stability of Factor VIII. For example, the A108I amino acid substitution introduces a hydrophobic residue that better fills the inter-domain space, stabilizing the interaction. Likewise, an R121C/L2302C (SPE) double amino acid substitution introduces a disulfide bond spanning the A1-C2 domains, further stabilizing the interaction. Taken together, all three amino acid substitutions increase the thermal stability of Factor VIII by 3 to 4-fold. For review, see Wakabayashi et al., J Biol Chem. 286(29):25748-55 (2011) and Wakabayashi et al., Thromb Haemost. 10(3):492-95 (2012). Accordingly, in some

embodiments, the encoded Factor VIII polypeptide includes A108I and/or R121C/L2302C amino acid substitutions.

[00244] Mutation of E113 (SPE), located within the calcium binding domain of Factor VIII, increases the specific FVIII clotting activity. For example, E113A appears to increase FXase formation through increased FVIII affinity for Factor IXa. Specifically, the E113A amino acid substitution increases specific FVIII clotting activity two-fold and increases affinity for Factor IXa by four-fold (Biochemistry, 41:8485 (2002); J. Biol. Chem., 279:12677 (2004); and Biochemistry, 44:10298 (2005)). Accordingly, in some embodiments, the encoded Factor VIII polypeptides include an E113A amino acid substitution.

[00245] Substitution of one or more amino acid residues surrounding the Factor VIII APC cleavage site (residues 331-341 (SPE)) reduce Factor VIIIa inactivation by activated protein C, without affecting FVIII activity. For example PQL333-335VDQ (SPE) amino acid substitutions reduce Factor VIII inactivation by 16-fold. Likewise, MKN336-339GNQ amino acid substitutions reduce Factor VIII inactivation by 9-fold. When combined, the two triple amino acid substitutions (e.g., PQLRMKN333-339VDQRGNGN) (SEQ ID NOS 34 and 35, respectively) reduce Factor VIII inactivation by 100-fold (J. Biol. Chem., 282:20264 (2007)). Accordingly, in some embodiments, the encoded Factor VIII polypeptide include PQL333-335VDQ and/or MKN337-339GNQ (SPE) amino acid substitutions.

[00246] Mutations within the A2 domain interface also increase Factor VIII stability. Specifically, mutating charged residues in the A1-A2 and A2-A3 domain interfaces increases stability and retention of the A2 subunit in Factor VIIIa. For example, mutation of D519, E665, and E1984 to V or A yields up to 2-fold increased stability in Factor VIII and up to 5-fold stability in Factor VIIIa. Specifically, D519A/E665V amino acid substitutions provide a 3-fold increase in stability; D519V/E665V amino acid substitutions provide a 2-fold increase in stability, an 8-fold decrease in A2 dissociation, and a 2-4-fold increase in thrombin generation potential; D519V/E1984A amino acid substitutions provide a 2-fold increase in stability; and D519V/E665V/E1984A amino acid substitution provide a 2-fold increase in stability (Blood 112:2761-69 (2008); J. Thromb. Haemost., 7:438-44 (2009)). Accordingly, in some

embodiments, the encoded Factor VIII polypeptides include one or more of D519A/V, E665A/V, and E1984A/V amino acid substitutions.

[00247] Of particular relevance to the present disclosure are a number of specific mutations that can be included separately or in combinations with other variants described herein. These variants are coded as sets herein as follows: “m1” refers to a single amino acid change, “m2” is a set of 5 amino acid variants, “m3” is a combination of a deletion of 7 amino acids and an insertion of six amino acids that span the junction between the polypeptide linker and the heavy chain, “m4” is a combination of the m1 single mutation and the m5 double mutation, and “m5” is a set of two cysteine ablations. These mutations are described below. These can be included in any particular construct alone or in combination with other variants, and they are coded accordingly. For example, “m23” is a combination of the m2 and m3 variants onto a particular scaffold, as outlined herein; thus “CS01m23-FL-NA” or “CS01-FL-NA_{m23}” refers to the CS01 codon-altered polynucleotide sequence with the nucleotides encoding the m2 and m3 mutations included, and “CS01m23-FL-AA” or “CS01-FL-AA_{m23}” refers to the amino acid sequence. As CS01 is codon-altered but does not change the amino acid sequence of Refacto, these can be thought of on the amino acid level as mutations as compared to the Refacto amino acid sequence of CS01-FL-AA (SEQ ID NO: 2).

[00248] In many embodiments, the polypeptides of the disclosure are made with the “m1” variant included. Mutations within an 11 amino acid hydrophobic β -sheet in the A1 domain, which interacts with BiP, increase secretion of Factor VIII. For example, an F328S (SPI, F309S SPE) amino acid substitution within the pocket increased Factor VIII secretion 3-fold. The F328S variant is referred to herein as the “m1” mutation and is within the heavy chain. Again, as described herein, the number of the variants can be done inclusive of the signal peptide, “**Signal Peptide Inclusive**”, or “SPI”, or starting from the processed final protein sequence, “**Signal Peptide Exclusive**”, or “SPE”. Thus, using SPI numbering, the mutation F328S is the same as the F309 SPE mutant. Generally the specification uses the SPI numbering, but as will be appreciated by those in the art, either numbering system results in the same mutation(s).

[00249] Accordingly, included in the present disclosure are polypeptides that include the m1 mutation, including CS01-FL-AA_{m1}, CS01-HC-AA_{m1}, CS04-FL-AA_{m1}, CS04-HC-AA_{m1}

CS23-FL-AAm1, CS23-HC-AAm1, CS40-FL-AAm1 and CS40-HC-AAm1 (all of which encode the same corresponding protein sequences).

[00250] In addition, included in the present disclosure are not only polypeptide sequences that include the m1 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m1 mutation, such as CS01-FL-NA_{m1}, CS01-HC-NA_{m1}, CS04-FL-NA_{m1}, CS04-HC-NA_{m1}, CS23-FL-NA_{m1}, CS23-HC-NA_{m1}, CS40-FL-NA_{m1} and CS40-HC-NA_{m1}.

[00251] In many embodiments, the polypeptides of the disclosure are made with the “m2” variant set included, which is the I105V/A127S/G151K/M166T/L171P mutations (SPI numbering; (SPE numbering is V86I/S108A/K132G/T147M/P152L, respectively). The m2 mutation set is based on the fact that substitution of porcine amino acids 82-176 for the corresponding human amino acids in a B-domain deleted gene therapy construct increased Factor VIII activity when expressed in HEK293 cells (W. Xiao, communication). *Id.* Back-mutation of single porcine amino acids into the human BDD-FVIII construct identified five amino acids within the A1 domain that contribute to this phenomenon: I105V, A127S, G151K, M166T, and L171P (SPI). Introduction of the combination of these mutations into the human construct recapitulated the improved activity of the larger porcine substitution. *Id.* Accordingly, in some embodiments, the encoded Factor VIII polypeptides include one or more amino acid substitutions selected from I105V, A127S, G151K, M166T, and L171P, with the entire 5 amino acid set, m2, finding particular use in many embodiments. As for the m1 mutation, the m2 variants are in the heavy chain, and thus the present disclosure includes polypeptides that include the m2 mutation, including CS01-FL-AAm2, CS01-HC-AAm2, CS04-FL-AAm2, CS04-HC-AAm2, CS23-FL-AAm2, CS23-HC-AAm2, CS40-FL-AAm2 and CS40-HC-AAm2 (all of which encode the same corresponding protein sequences).

[00252] In addition, included in the present disclosure are not only polypeptide sequences that include the m2 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m2 mutation, such as CS01-FL-NA_{m2}, CS01-HC-NA_{m2}, CS04-FL-NA_{m2}, CS04-HC-NA_{m2}, CS23-FL-NA_{m2}, CS23-HC-NA_{m2}, CS40-FL-NA_{m2} and CS40-HC-NA_{m2}.

[00253] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m3 mutations. m3 is the substitution of seven amino acids for six across the HC-B

domain interface that introduces an additional glycosylation site introduced close to the interface. Accordingly, in some embodiments, m3 is the deletion of amino acids AIEPRSF755-761 and the insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19) (e.g., AIEPRSF755-761TTYVNRSL) ("TTYVNRSL" disclosed as SEQ ID NO: 33). Residues AIEPR755-759, relative to SEQ ID NO: 19, fall within the end of the heavy chain, while residues S760 and F761 fall within the B-domain. In some embodiments, where the FVIII B-domain is deleted, truncated, or replaced, residues S760 and F761 may not be present in the underlying amino acid sequence being mutated. Accordingly, in some embodiments, m3 is the deletion of amino acids AIEPR755-759 and the insertion of amino acids TTYVNRSL (SEQ ID NO: 33) after N754, relative to FVIII-FL-AA (SEQ ID NO: 19) (e.g., AIEPR755-759TTYVNRSL ("TTYVNRSL" disclosed as SEQ ID NO: 33)

[00254] The m3 variants are in the junction between the heavy chain and the B domain, and thus the present disclosure includes polypeptides that include the m3 mutation, including CS01-FL-AAm3, CS01-HC-AAm3, CS04-FL-AAm3, CS04-HC-AAm3, CS23-FL-AAm3, CS23-HC-AAm3, CS40-FL-AAm3 and CS40-HC-AAm3 (all of which encode the same corresponding protein sequences).

[00255] In addition, included in the present disclosure are not only polypeptide sequences that include the m3 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m3 mutations, such as CS01-FL-NAAm3, CS01-HC-NAAm3, CS04-FL-NAAm3, CS04-HC-NAAm3, CS23-FL-NAAm3, CS23-HC-NA-m3, CS40-FL-NAAm3 and CS40-HC-NAAm3.

[00256] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m4 mutations. Elimination of the C1899-C1903 disulfide bond in Factor VIII also increased secretion. Moreover, the increases in Factor VIII secretion are additive for the combination of F328S (SPI, F309S SPE) and C1918G/C1922G amino acid substitutions (Miao et al., Blood, 103:3412-19 (2004); Selvaraj et al., J. Thromb. Haemost., 10:107-15 (2012)). Accordingly, in some embodiments, the encoded Factor VIII polypeptides include m4 mutations, which is the F328S (SPI, F309S SPE) and C1918G/C1922G (SPI) amino acid substitutions. As the F328S variant is in the heavy chain and the two cysteine variants are in the light chain, polypeptide sequences that include m4 mutations are CS01-FL-AAm4, CS01-HC-AAm4, CS01-

LC-AAm4, CS04-FL-AAm4, CS04-HC-AAm4, CS04-LC-AAm4, CS23-FL-AAm4, CS23-HC-AAm4 and CS23-LC-AAm4.

[00257] In addition, included in the present disclosure are not only polypeptide sequences that include the m4 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m4 mutations, such as CS01-FL-NA_{m4}, CS01-HC-NA_{m4}, CS01-LC-NA_{m4}, CS04-FL-NA_{m4}, CS04-HC-NA_{m4}, CS04-LC-NA_{m4}, CS23-FL-NA_{m4}, CS23-HC-NA_{m4}, CS23-LC-NA_{m4}, CS40-FL-NA_{m4}, CS40-HC-NA_{m4} and CS40-LC-NA_{m4}.

[00258] In additional embodiments, the polypeptides and polynucleotides of the disclosure include m5 mutations. As above, elimination of the C1899-C1903 disulfide bond in Factor VIII also increased secretion. C1918G/C1922G (SPI) amino acid substitutions, contained within the light chain, referred to herein as the m5 mutation set.

[00259] The m5 variants are in the light chain, and thus the present disclosure includes polypeptides that include the m5 mutation, including CS01-FL-AA_{m5}, CS01-LC-AA_{m5}, CS04-FL-AA_{m5}, CS04-LC-AA_{m5}, CS23-FL-AA_{m5}, CS23-LC-AA_{m5}, CS40-FL-AA_{m5} and CS40-LC-AA_{m5} (all of which encode the same corresponding protein sequences).

[00260] In addition, included in the present disclosure are not only polypeptide sequences that include the m5 mutation, but also those codon-altered polynucleotide sequences that encode proteins with the m5 mutations, such as CS01-FL-NA_{m5}, CS01-LC-NA_{m5}, CS04-FL-NA_{m5}, CS04-LC-NA_{m5}, CS23-FL-NA_{m5}, CS23-LC-NA_{m5}, CS40-FL-NA_{m5} and CS40-LC-NA_{m5}.

[00261] In addition to specific constructs (both amino acid and nucleic acid) that include m1, m2, m3, m4 and m5 individually, combinations of mutation sets can be made as outlined herein. As noted herein, these are noted as “m12”, which is the combination of m1 and m2 sets, or “m123” which is the combination of m1, m2 and m3 sets. Thus, included in the disclosure are dual combinations including m12, m13, m14, m15, m23, m24, m25, m34, m35 and m45. Also included are triple combinations, m123, m124, m125, m234, m235 and m345. Further included are quad combinations, m1234, m1235, m1345 and the m12345 combination.

Of particular interest in some embodiments are the following mutation sets: m1, m2, m3 and m4, m23, m123, and m234.

B. Factor VIII B-domain Substituted Linkers

[00262] In some embodiments, the linkage between the FVIII heavy chain and the light chain (e.g., the B-domain in wild-type Factor VIII) is further altered. Due to size constraints of AAV packaging capacity, B-domain deleted, truncated, and or linker substituted variants should improve the efficacy of the FVIII gene therapy construct. The most conventionally used B-domain substituted linker is that of SQ FVIII, which retains only 14 amino acids of the B domain as linker sequence. Another variant of porcine VIII ("OBI-1," described in U.S. Patent No. 6,458,563) is well expressed in CHO cells, and has a slightly longer linker of 24 amino acids. In some embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an SQ-type B-domain linker sequence. In other embodiments, the Factor VIII constructs encoded by the codon-altered polynucleotides described herein include an OBI-1-type B-domain linker sequence.

[00263] In some embodiments, the encoded Factor VIII polypeptides described herein include an SQ-type B-domain linker, including amino acids 760-762/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Sandberg et al. *Thromb. Haemost.* 85:93 (2001)). In some embodiments, the SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the SQ-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00264] In some embodiments, the encoded Factor VIII polypeptides described herein include a Greengene-type B-domain linker, including amino acids 760/1582-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Oh et al., *Biotechnol. Prog.*, 17:1999 (2001)). In some embodiments, the Greengene-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments,

the Greengene-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the Greengene-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00265] In some embodiments, the encoded Factor VIII polypeptides described herein include an extended SQ-type B-domain linker (SFSQNPPVLKRHQR; BDL-SQ-AA; SEQ ID NO: 30), including amino acids 760-769/1657-1667 of the wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19) (Thim et al., *Haemophilia*, 16:349 (2010)). In some embodiments, the extended SQ-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the extended SQ-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the extended SQ-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00266] In some embodiments, the encoded Factor VIII polypeptides described herein include a porcine OBI-1-type B-domain linker, including the amino acids SFAQNSRPPSASAPKPPVLRHQR (SEQ ID NO: 31) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the porcine OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00267] In some embodiments, the encoded Factor VIII polypeptides described herein include a human OBI-1-type B-domain linker, including amino acids 760-772/1655-1667 of the

wild-type human Factor VIII B-domain (FVIII-FL-AA; SEQ ID NO: 19). In some embodiments, the human OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the human OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the human OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00268] In some embodiments, the encoded Factor VIII polypeptides described herein include an O8-type B-domain linker, including the amino acids SFSQNSRHQAYRYRRG (SEQ ID NO: 32) from the wild-type porcine Factor VIII B-domain (Toschi et al., *Curr. Opin. Mol. Ther.* 12:517 (2010)). In some embodiments, the porcine OBI-1-type B-domain linker has one amino acid substitution relative to the corresponding wild-type sequence. In some embodiments, the porcine OBI-1-type B-domain linker has two amino acid substitutions relative to the corresponding wild-type sequence. In some embodiments, a glycosylation peptide is inserted into the porcine OBI-1-type B-domain linker. In some embodiments, the glycosylation peptide is selected from those shown in Figure 13 (SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance).

[00269] Removal of the B-domain from Factor VIII constructs does not appear to affect the activity of the activated enzyme (e.g., FVIIIa), presumably because the B-domain is removed during activation. However, the B-domain of Factor VIII contains several residues that are post-translationally modified, e.g., by N- or O-linked glycosylation. In silico analysis (Prediction of N-glycosylation sites in human proteins, R. Gupta, E. Jung and S. Brunak, in preparation (2004)) of the wild-type Factor VIII B-domain predicts that at least four of these sites are glycosylated in vivo (Figure 14). It is thought that these modifications within the B-domain contribute to the post-translational regulation and/or half-life of Factor VIII in vivo.

[00270] While the Factor VIII B-domain is absent in mature Factor VIIIa protein, glycosylation within the B-domain of the precursor Factor VIII molecule may increase the circulating half-life of the protein prior to activation. Thus, in some embodiments, the

polypeptide linker of the encoded Factor VIII constructs described herein includes one or more glycosylation sequences, to allow for glycosylation in vivo. In some embodiments, the polypeptide linker includes at least one consensus glycosylation sequence (e.g., an N- or O-linked glycosylation consensus sequence). In some embodiments, the polypeptide linker includes at least two consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least three consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least four consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least five consensus glycosylation sequences. In some embodiments, the polypeptide linker includes at least 6, 7, 8, 9, 10, or more consensus glycosylation sequences.

[00271] In some embodiments, the polypeptide linker contains at least one N-linked glycosylation sequence N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least two N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least three N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least four N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least five N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T. In some embodiments, the polypeptide linker contains at least 6, 7, 8, 9, 10, or more N-linked glycosylation sequences N-X-S/T, where X is any amino acid other than P, S, or T.

[00272] In some embodiments, the polypeptide linker includes a glycosylation peptide with high sequence identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation polypeptide has at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than one amino acid substitution

relative to any of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has an amino acid sequence selected from any of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B.

[00273] In some embodiments, the glycosylation peptide has at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 90% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 95% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 98% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B.

[00274] In some embodiments, the glycosylation peptide has no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, and is encoded by a polynucleotide sequence having at least 90% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, and is encoded by a polynucleotide sequence having at least 95% identity to

a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, and is encoded by a polynucleotide sequence having at least 98% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B.

[00275] In some embodiments, the glycosylation peptide has no more than one amino acid substitution relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 90% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than one amino acid substitution relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 95% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has no more than one amino acid substitution relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 98% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B.

[00276] In some embodiments, the glycosylation peptide has a sequence selected from any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 90% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as

shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has a sequence selected from any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 95% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide has a sequence selected from any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B and is encoded by a polynucleotide sequence having at least 98% identity to a corresponding nucleotide sequence selected from SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance, as shown in Figures 13A-13B.

[00277] In some embodiments, the Factor VIII polypeptide encoded by a codon-altered polynucleotide described herein has a B-domain substituted linker in which a glycosylation peptide is inserted into the SQ linker sequence (amino acids 760-773 of CS04-FL-AA; SEQ ID NO: 2). In a specific embodiment, the glycosylation peptide is selected from any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, a glycosylation peptide having at least 92% identity to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, a glycosylation peptide having no more than two amino acid substitutions relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B, and a glycosylation peptide having no more than one amino acid substitution relative to any one of SEQ ID NOS 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, and 75, respectively, in order of appearance, as shown in Figures 13A-13B. In some embodiments, the glycosylation peptide is inserted in the SQ peptide between residues N768 and P769 (relative to CS04-FL-AA; SEQ ID NO: 2).

[00278] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to any one of those shown in Figure 6 (SEQ ID NOS 5-7 and 36-48, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 95% identity to any one of those shown

in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 96% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 97% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 98% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 99% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 99.5% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence has at least 99.9% identity to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance). In some embodiments, the third nucleotide sequence is identical to any one of those shown in Figure 13 (SEQ ID NOS 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, and 74, respectively, in order of appearance).

C. Codon-altered Polynucleotides Encoding a Factor VIII Variant with a Cleavable Linker

CS04 Codon Altered Polynucleotides

[00279] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to

CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00280] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[00281] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00282] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-FL-NA (SEQ ID NO: 1). In some embodiments, the nucleotide sequence is identical to CS04-FL-NA (SEQ ID NO: 1).

[00283] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS04-FL-AA (SEQ ID NO: 2).

[00284] In some embodiments, the Factor VIII variant encoded by the CS04 polynucleotide, having high sequence homology to CS04-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00285] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide

comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m5 amino acid substitution.

[00286] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m35 amino acid substitutions.

[00287] In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m125 amino acid substitutions.

CS01 Codon Altered Polynucleotides

[00288] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS01-LC-NA (SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site,

which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00289] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[00290] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00291] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments,

the nucleotide sequence has at least 95% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-FL-NA (SEQ ID NO: 13). In some embodiments, the nucleotide sequence is identical to CS01-FL-NA (SEQ ID NO: 13).

[00292] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 97% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 98% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-FL-AA (SEQ ID NO: 2). In some embodiments, the amino acid sequence is identical to CS01-FL-AA (SEQ ID NO: 2).

[00293] In some embodiments, the Factor VIII variant encoded by the CS01 polynucleotide, having high sequence homology to CS01-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00294] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m5 amino acid substitution.

[00295] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m35 amino acid substitutions.

[00296] In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m125 amino acid substitutions.

CS23 Codon Altered Polynucleotides

[00297] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a Factor VIII variant polypeptide with a linker that is cleavable in vivo. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The polypeptide linker includes a furin cleavage site, which allows for maturation in vivo (e.g., after expression in vivo or administration of the precursor polypeptide).

[00298] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[00299] In some embodiments, the polypeptide linker of the Factor VIII construct is encoded by a third nucleotide sequence having high sequence identity to BDLO04 (SEQ ID NO: 6), which encodes the 14-amino acid linker corresponding to amino acids 760-773 of CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the third nucleotide sequence has at least 95% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 96% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 97% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence has at least 98% identity to BDLO04 (SEQ ID NO: 6). In some embodiments, the third nucleotide sequence is identical to BDLO04 (SEQ ID NO: 6).

[00300] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-FL-NA

(SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-FL-NA (SEQ ID NO: 20). In some embodiments, the nucleotide sequence is identical to CS23-FL-NA (SEQ ID NO: 20).

[00301] In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 97% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 98% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-FL-AA (SEQ ID NO: 21). In some embodiments, the amino acid sequence is identical to CS23-FL-AA (SEQ ID NO: 21).

[00302] In some embodiments, the Factor VIII variant encoded by the CS23 polynucleotide, having high sequence homology to CS23-FL-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00303] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m5 amino acid substitution.

[00304] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the

Factor VIII variant encoded by the CS23 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m35 amino acid substitutions.

[00305] In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m125 amino acid substitutions.

D. Codon-altered Polynucleotides Encoding a Single-chain Factor VIII Protein

[00306] Factor VIII constructs in which the furin cleavage site located at the C-terminal end of the B-domain is removed retain activity as a single chain polypeptide, despite that normal maturation of the Factor VIII molecule cannot occur (Leyte et al. (1991)). Similarly, a B-domain deleted Factor VIII construct with an attenuated furin site (containing an R1664H amino acid substitution) is more biologically active than the corresponding Factor VIII construct with a wild-type furin cleavage site (Siner et al. (2013)). Accordingly, in some embodiments, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The single-chain Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The polypeptide linker does not include a furin cleavage site.

Single-chain CS04 Codon Altered Polynucleotides

[00307] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional

polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS04-HC-NA (SEQ ID NO: 3), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS04-LC-NA (SEQ ID NO: 4), which is the portion of CS04-FL-NA (SEQ ID NO: 1) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00308] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS04-HC-NA and CS04-LC-NA (SEQ ID NOS 3 and 4), respectively.

[00309] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence

has at least 99.5% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC1-NA (SEQ ID NO: 9). In some embodiments, the nucleotide sequence is identical to CS04-SC1-NA (SEQ ID NO: 9).

[00310] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 95% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 96% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 97% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 98% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS04-SC2-NA (SEQ ID NO: 11). In some embodiments, the nucleotide sequence is identical to CS04-SC2-NA (SEQ ID NO: 11).

[00311] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; *HsFVIII* Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS04-SC1-AA (SEQ ID NO: 10).

[00312] In some embodiments, the Factor VIII variant encoded by the CS04-SC1 polynucleotide, having high sequence homology to CS04-SC1-AA (e.g., at least 95%, 96%,

97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00313] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; *HsFVIII* Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS04-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS04-SC2-AA (SEQ ID NO: 12).

[00314] In some embodiments, the single-chain Factor VIII variant encoded by the CS04-SC2 polynucleotide, having high sequence homology to CS04-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00315] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises an m5 amino acid substitution.

[00316] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m23 amino

acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m35 amino acid substitutions.

[00317] In one embodiment, the single-chain Factor VIII variant encoded by the CS04 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS04 polynucleotide comprises m125 amino acid substitutions.

Single-chain CS01 Codon Altered Polynucleotides

[00318] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS01-HC-NA (SEQ ID NO: 24), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS01-LC-NA (SEQ ID NO: 25), which is the portion of CS01-FL-NA (SEQ ID NO: 13) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00319] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to

CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS01-HC-NA and CS01-LC-NA (SEQ ID NOS 24 and 25), respectively.

[00320] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS01-SC1-NA (SEQ ID NO: 26). In some embodiments, the nucleotide sequence is identical to CS01-SC1-NA (SEQ ID NO: 26).

[00321] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 95% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 96% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 97% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 98% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS01-SC2-NA (SEQ ID NO: 27). In some

embodiments, the nucleotide sequence has at least 99.9% identity to CS01-SC2-NA (SEQ ID NO: 27). In some embodiments, the nucleotide sequence is identical to CS01-SC2-NA (SEQ ID NO: 27).

[00322] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; *HsFVIII* Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS01-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS01-SC1-AA (SEQ ID NO: 10).

[00323] In some embodiments, the Factor VIII variant encoded by the CS01-SC1 polynucleotide, having high sequence homology to CS01-SC1-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00324] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; *HsFVIII* Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid

sequence has at least 99.9% identity to CS01-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS01-SC2-AA (SEQ ID NO: 12).

[00325] In some embodiments, the single-chain Factor VIII variant encoded by the CS01-SC2 polynucleotide, having high sequence homology to CS01-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00326] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises an m5 amino acid substitution.

[00327] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m35 amino acid substitutions.

[00328] In one embodiment, the single-chain Factor VIII variant encoded by the CS01 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS01 polynucleotide comprises m125 amino acid substitutions.

Single-chain CS23 Codon Altered Polynucleotides

[00329] In one embodiment, the codon-altered polynucleotides provided herein include a nucleotide sequence encoding a single-chain Factor VIII variant polypeptide. The Factor VIII polypeptide includes a Factor VIII light chain, a Factor VIII heavy chain, and an optional polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain. The heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having high sequence identity to CS23-HC-NA (SEQ ID NO: 22), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII heavy chain. The light chain of the Factor VIII polypeptide is encoded by a second nucleotide sequence with high sequence identity to CS23-LC-NA (SEQ ID NO: 23), which is the portion of CS23-FL-NA (SEQ ID NO: 20) encoding for a Factor VIII light chain. The optional polypeptide linker does not include a furin cleavage site.

[00330] In some embodiments, the first and second nucleotide sequences have at least 95% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 96% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 97% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 98% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.5% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences have at least 99.9% sequence identity to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively. In some embodiments, the first and second nucleotide sequences are identical to CS23-HC-NA and CS23-LC-NA (SEQ ID NOS 22 and 23), respectively.

[00331] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC1-NA (SEQ ID NO: 28). In some

embodiments, the nucleotide sequence has at least 95% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC1-NA (SEQ ID NO: 28). In some embodiments, the nucleotide sequence is identical to CS23-SC1-NA (SEQ ID NO: 28).

[00332] In some embodiments, the codon-altered polynucleotide has a nucleotide sequence with high sequence identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 95% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 96% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 97% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 98% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99.5% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence has at least 99.9% identity to CS23-SC2-NA (SEQ ID NO: 29). In some embodiments, the nucleotide sequence is identical to CS23-SC2-NA (SEQ ID NO: 29).

[00333] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10; human Factor VIII Δ (760-1667) (SPI; CS04 Δ (741-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99% identity to

CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC1-AA (SEQ ID NO: 10). In some embodiments, the amino acid sequence is identical to CS23-SC1-AA (SEQ ID NO: 10).

[00334] In some embodiments, the Factor VIII variant encoded by the CS23-SC1 polynucleotide, having high sequence homology to CS23-SC1-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00335] In some embodiments, the single-chain Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12; human Factor VIII Δ (772-1667) (SPI; *HsFVIII* Δ (753-1648), SPE)). In some embodiments, the Factor VIII variant encoded by the codon-altered polynucleotide has an amino acid sequence with high sequence identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 97% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 98% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.5% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence has at least 99.9% identity to CS23-SC2-AA (SEQ ID NO: 12). In some embodiments, the amino acid sequence is identical to CS23-SC2-AA (SEQ ID NO: 12).

[00336] In some embodiments, the single-chain Factor VIII variant encoded by the CS23-SC2 polynucleotide, having high sequence homology to CS23-SC2-AA (e.g., at least 95%, 96%, 97%, 98%, 99%, 99.5%, or 99.9% identity), comprises one or more amino acid substitutions selected from m1, m2, m3, m4, and m5.

[00337] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises an m1 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m2 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m3 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23

polynucleotide comprises an m4 amino acid substitution. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises an m5 amino acid substitution.

[00338] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises m12 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m13 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m23 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m24 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m25 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m34 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m35 amino acid substitutions.

[00339] In one embodiment, the single-chain Factor VIII variant encoded by the CS23 polynucleotide comprises m123 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m234 amino acid substitutions. In one embodiment, the Factor VIII variant encoded by the CS23 polynucleotide comprises m125 amino acid substitutions.

E. Factor VIII Expression Vectors

[00340] In some embodiments, the codon-altered polynucleotides described herein are integrated into expression vectors. Non-limiting examples of expression vectors include viral vectors (e.g., vectors suitable for gene therapy), plasmid vectors, bacteriophage vectors, cosmids, phagemids, artificial chromosomes, and the like.

[00341] Non-limiting examples of viral vectors include: retrovirus, e.g., Moloney murine leukemia virus (MMLV), Harvey murine sarcoma virus, murine mammary tumor virus, and Rous sarcoma virus; adenoviruses, adeno-associated viruses; SV40-type viruses; polyomaviruses; Epstein-Barr viruses; papilloma viruses; herpes viruses; vaccinia viruses; and polio viruses.

[00342] In some embodiments, the codon-altered polynucleotides described herein are integrated into a gene therapy vector. In some embodiments, the gene therapy vector is a retrovirus, and particularly a replication-deficient retrovirus. Protocols for the production of replication-deficient retroviruses are known in the art. For review, see Kriegler, M., *Gene Transfer and Expression, A Laboratory Manual*, W.H. Freeman Co., New York (1990) and Murry, E. J., *Methods in Molecular Biology*, Vol. 7, Humana Press, Inc., Clifton, N.J. (1991).

[00343] In one embodiment, the gene therapy vector is an adeno-associated virus (AAV) based gene therapy vector. AAV systems have been described previously and are generally well known in the art (Kelleher and Vos, *Biotechniques*, 17(6):1110-17 (1994); Cotten et al., *Proc Natl Acad Sci USA*, 89(13):6094-98 (1992); Curiel, *Nat Immun*, 13(2-3):141-64 (1994); Muzyczka, *Curr Top Microbiol Immunol*, 158:97-129 (1992); and Asokan A, et al., *Mol. Ther.*, 20(4):699-708 (2012), each incorporated herein by reference in their entireties for all purposes). Details concerning the generation and use of rAAV vectors are described, for example, in U.S. Patent Nos. 5,139,941 and 4,797,368, each incorporated herein by reference in their entireties for all purposes. In a particular embodiment, the AAV vector is an AAV-8 vector.

[00344] In some embodiments, the codon-altered polynucleotides described herein are integrated into a retroviral expression vector. These systems have been described previously, and are generally well known in the art (Mann *et al.*, *Cell*, 33:153-159, 1983; Nicolas and Rubinstein, In: *Vectors: A survey of molecular cloning vectors and their uses*, Rodriguez and Denhardt, eds., Stoneham: Butterworth, pp. 494-513, 1988; Temin, In: *Gene Transfer*, Kucherlapati (ed.), New York: Plenum Press, pp. 149-188, 1986). In a specific embodiment, the retroviral vector is a lentiviral vector (*see*, for example, Naldini *et al.*, *Science*, 272(5259):263-267, 1996; Zufferey *et al.*, *Nat Biotechnol*, 15(9):871-875, 1997; Blomer *et al.*, *J Virol.*, 71(9):6641-6649, 1997; U.S. Pat. Nos. 6,013,516 and 5,994,136).

[00345] A wide variety of vectors can be used for the expression of a Factor VIII polypeptide from a codon-altered polypeptide in cell culture, including eukaryotic and prokaryotic expression vectors. In certain embodiments, a plasmid vector is contemplated for use in expressing a Factor VIII polypeptide in cell culture. In general, plasmid vectors containing replicon and control sequences which are derived from species compatible with the

host cell are used in connection with these hosts. The vector can carry a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells. The plasmid will include the codon-altered polynucleotide encoding the Factor VIII polypeptide, operably linked to one or more control sequences, for example, a promoter.

[00346] Non-limiting examples of vectors for prokaryotic expression include plasmids such as pRSET, pET, pBAD, etc., wherein the promoters used in prokaryotic expression vectors include lac, trc, trp, recA, araBAD, etc. Examples of vectors for eukaryotic expression include: (i) for expression in yeast, vectors such as pAO, pPIC, pYES, pMET, using promoters such as AOX1, GAP, GAL1, AUG1, etc; (ii) for expression in insect cells, vectors such as pMT, pAc5, pIB, pMIB, pBAC, etc., using promoters such as PH, p10, MT, Ac5, OpIE2, gp64, polh, etc., and (iii) for expression in mammalian cells, vectors such as pSVL, pCMV, pRc/RSV, pcDNA3, pBPV, etc., and vectors derived from viral systems such as vaccinia virus, adeno-associated viruses, herpes viruses, retroviruses, etc., using promoters such as CMV, SV40, EF-1, UbC, RSV, ADV, BPV, and β -actin.

IV. Examples

Example 1 – Construction of a Codon-Altered Factor VIII Variant Expression Sequence

[00347] Two hurdles had to be overcome in order to create a Factor VIII coding sequence that is effective for gene therapy of hemophilia A. First, because of the genomic size limitations of conventional gene therapy delivery vectors (e.g., AAV virions), the encoded Factor VIII polypeptide had to be shortened considerably. Second, the coding sequence had to be altered to: (i) stabilize packaging interactions within the delivery vector, (ii) stabilize the mRNA intermediary, and (iii) improve the robustness of transcription/translation of the mRNA.

[00348] To achieve the first objective, Applicants started with a B-domain deleted Factor VIII variant construct, referred to herein as “FVIII-BDD-SQ.” In this construct, the B-domain is replaced with a fourteen amino acid sequence referred to as the “SQ” sequence. Recombinant FVIII-BDD-SQ is sold under the trade name REFACTO®, and has been shown to be effective for the management of hemophilia A. However, the native coding sequence for FVIII-BDD-SQ,

which includes human wild-type nucleic acid sequences for the Factor VIII heavy and light chains, is ineffectively expressed in gene therapy vectors.

[00349] To address the poor expression of the native FVIII-BDD-SQ, the codon optimization algorithm described in Fath et al. (PLoS ONE, 6:e17596 (2011)), modified as described in Ward et al. (Blood, 117:798 (2011)) and in McIntosh et al. (Blood, 121, 3335-3344 (2013)), was applied to the FVIII-BDD-SQ sequence to create first intermediate coding sequence CS04a. However, Applicants recognized that the CS04a sequence created using the modified algorithm could be improved by further modifying the sequence. Accordingly, Applicants re-introduced CpG dinucleotides, re-introduced the CGC codon for arginine, changed the leucine and serine codon distributions, re-introduced highly conserved codon pairs, and removed cryptic TATA box, CCAAT box, and splice site elements, while avoiding CpG islands and local overrepresentation of AT-rich and GC-rich stretches.

[00350] First, the modified algorithm systematically replaces codons containing CpG-dinucleotides (e.g., arginine codons) with non-CpG-dinucleotide codons, and eliminates/avoids CpG-dinucleotides created by neighboring codons. This strict avoidance of CpG dinucleotides is usually done to prevent TLR-induced immunity after intramuscular injection of DNA vaccines. However, doing so limits the codon optimization possibilities. For example, the modified algorithm excludes use of the complete set of CGX arginine codons. This is particularly disruptive in the coding of genes for expression in human cells, because CGC is the most frequently used arginine codon in highly expressed human genes. Additionally, avoiding the creation of CpGs by neighboring codons further limits the optimization possibilities (e.g., limits the number of codon pairs that may be used together).

[00351] Because TLR-induced immunity is not expected to be a problem associated with liver-directed, AAV-based gene therapy, codons including CpGs, and neighboring codons creating CpGs, were re-introduced into intermediate coding sequence CS04a, preferentially in the sequence coding for the Factor VIII light chain (e.g., at the 3' end of the FVIII-BDD-SQ coding sequence). This allowed for more frequent use of preferred human codons, particularly those for arginine. Care was taken, however, to avoid creation of CpG islands, which are regions of coding sequence having a high frequency of CpG sites. This is contrary to the teachings of

Krinner et al. (Nucleic Acids Res., 42(6):3551-64 (2014)), which suggests that CpG domains downstream of transcriptional start sites promote high levels of gene expression.

[00352] Second, the modified algorithm applies certain codons exclusively, such as CTG for leucine, GTG for valine, and CAG for glutamine. However, this offends the principles of balanced codon use, for example, as proposed in Haas et al. (Current Biology, 6(3):315-24 (1996)). To account for the overuse of preferred codons by the modified algorithm, alternate leucine codons were re-introduced where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[00353] Third, the modified algorithm replaces codon pairs without regard to how conserved they are in nature, when certain criteria (e.g., the presence of CG-dinucleotides) are met. To account for beneficial properties which may have been conserved by evolution, the most conserved codon pairs that were replaced by the algorithm and the most conserved preferred codon pairs, e.g., as described in Tats et al. (BMC Genomics 9:463 (2008)), were analyzed and adjusted where allowed by the other rules applied to the codon alteration (e.g., CpG frequency and GC content).

[00354] Fourth, serine codons used in the intermediate coding sequence were also re-engineered. Specifically, AGC, TCC, and TCT serine codons were introduced into the modified coding sequence with higher frequency, to better match overall for human codon usage (Haas et al., supra).

[00355] Fifth, TATA box, CCAAT box elements, and intron/exon splice sites were screened and removed from the modified coding sequence. When modifying the coding sequence, care was taken to avoid local overrepresentation of AT-rich or GC rich stretches.

[00356] Finally, in addition to optimizing the codon usage within the coding sequence, the structural requirements of the underlying AAV virion were considered when further refining the intermediate coding sequence CS04a. AAV vectors (e.g., the nucleic acid portion of an AAV virion) are packaged as single stranded DNA molecules into their capsids (for review, see, Daya and Berns, Clin. Microbiol Rev., 21(4):583-93 (2008)). The GC content of the vector is therefore likely to influence packaging of the genome and, thus, vector yields during production.

Like many algorithms, the modified algorithm used here creates an optimized gene sequence with a GC content of at least 60% (*see*, Fath et al., PLoS One, 6(3):e17596 (2011) (erratum in: PLoS One, (6)3 (2011)). However, the AAV8 capsid protein is encoded by a nucleotide sequence having a lower GC content of about 56%. Thus, to better mimic the native AAV8 capsid protein coding sequence, the GC content of the intermediate coding sequence CS04a was reduced to 56%.

[00357] The resulting CS04 coding sequence, shown in Figure 2, has an overall GC content of 56%. The CpG-dinucleotide content of the sequence is moderate. However, CpG dinucleotides are predominantly present in the downstream portion of the coding sequence, e.g., the portion coding for the Factor VIII light chain. The CS04 sequence has 79.77% nucleotide sequence identity to the corresponding coding sequences in wild-type Factor VIII (Genbank accession M14113).

[00358] For comparison purposes, several other codon-optimized, ReFacto constructs were prepared. CS01 was constructed by applying the codon-optimization algorithm of Fath et al., as modified by Ward et al., as done for CS04. However, unlike CS04, the CS01 construct does not contain any CpG islands. The CS08 ReFacto construct was codon-optimized as described in Radcliff P.M. et al., Gene Therapy, 15:289-97 (2008), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. The CS10 codon-optimized ReFacto construct was obtained from Eurofins Genomics (Ebersberg, Germany). The CS11 codon-optimized ReFacto construct was obtained from Integrated DNA Technologies, Inc. (Coralville, USA). The CH25 codon-optimized ReFacto construct was obtained from ThermoFischer Scientific's GeneArt services (Regensburg, Germany). The CS40 ReFacto construct consists of the wild type Factor VIII coding sequence. The algorithm used to construct CS23 is based on the JCAT tool (www.jcat.de), an on-line tool for codon-optimizations (Grote et al., 2005; Nucl. Acids Res. W526-31). The sequence was further modified to more reflect the codon usage of the albumin superfamily (Mirsafian et al. 2014: Sc. Word Journal 2014, ID 639682). The sequence identities shared between each of the ReFacto coding sequences is shown in **Table 2**, below.

Table 2 – Percent identity matrix for codon-altered Factor VIII constructs.

	CS01	CS04	CS08	CS10	CS11	CS40	CH25	CS23
CS01	100%							
CS04	93.0%	100%						
CS08	80.7%	82.2%	100%					
CS10	79.1%	79.4%	78.4%	100%				
CS11	78.3%	78.3%	78.1%	77.5%	100%			
CS40	79.6%	79.8%	76.7%	77.6%	75.4%	100%		
CH25	81.3%	85.1%	85.0%	79.9%	79.4%	75.8%	100%	
CS23	84.3%	89.2%	85.1%	80.3%	79.9	76.5%	93.2%	100%

[00359] Plasmids of each construct were constructed by cloning different synthetic DNA fragments into the same vector backbone plasmid (pCh-BB01). DNA synthesis of the Refacto-type BDD-FVIII fragments with flanking AscI and NotI enzyme restriction sites were done by ThermoFischer Scientific (Regensburg, Germany). The vector backbone contains two flanking AAV2-derived inverted terminal repeats (ITRs) that encompass a promoter/ enhancer sequence derived from the liver-specific murine transthyretin gene, AscI and NotI enzyme restriction sites for insertion of the respective Refacto-type BDD-FVIII and a synthetic polyA site. After ligation of the prepared vector backbone and inserts via the AscI and NotI sites, the resulting plasmids were amplified in milligram scale. The Refacto-type BDD-FVIII sequences of the constructs were verified by direct sequencing (Microsynth, Balgach, Switzerland). The cloning resulted in seven different plasmid constructs named pCS40, pCS01, pCS04, pCS08, pCS10, pCS11, and pCh25 (Figure 23). The constructs have the same vector backbone and encode the same B-domain deleted FVIII protein (Refacto-type BDD-FVIII), but differ in their FVIII coding sequence.

[00360] AAV8-based vectors were prepared by the three plasmid transfection method, as described in Grieger JC, et al. (Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector, Mol Ther., Oct 6. (2015) doi: 10.1038/mt.2015.187. [Epub ahead of print]), the content of which is hereby expressly incorporated by reference herein, in its entirety, for all purposes. HEK293 suspensions cells were used for plasmid transfections using the corresponding FVIII vector plasmid, the

helper plasmid pXX6-80 (carrying adenoviral helper genes), and the packaging plasmid pGSK2/8 (contributing the rep2 and cap8 genes). To isolate the AAV8 constructs, the cell pellets of one liter cultures were processed using iodixanol gradients, as described in Grieger et al. (2015, Supra). The procedure resulted in vector preparations called vCS01, vCS04, vCS08, vCS10, vCS11, and vCH25. Vectors were quantified by qPCR using the universal qPCR procedure targeting the AAV2 inverted terminal repeats (Aurnhammer, Human Gene Therapy Methods: Part B 23:18–28 (2012)). A control vector plasmid carrying AAV2 inverted terminal repeats served for preparing the standard curve. The resulting vCS04 construct is presented as SEQ ID NO: 8 in Figures 7A-7C.

[00361] The integrity of the vector genomes was analyzed by AAV agarose gel electrophoresis. The electrophoresis was performed as described in Fagone et al., Human Gene Therapy Methods 23:1-7 (2012). Briefly, AAV vector preparations were incubated at 75 °C for 10 minutes in the presence of 0.5% SDS and then cooled down to room temperature. Approximately 1.5E10 vector genomes (vg) were loaded per lane on a 1% 1xTAE agarose gel and electrophoresed for 60 min at 7 V/cm of gel length. The gel was then stained in 2x GelRed (Biotium Cat# 41003) solution and imaged by ChemiDocTMMP (Biorad). The results shown in Figure 24 demonstrate that the vCS01, vCS04, and vCS40 viral vectors have the same-sized genome, indicated by a distinct band in the 5kb range (Figure 24, lanes 2-4). Despite a vector size of approx. 5.2 kb, the genome is a homogenous band confirming correct packaging of the somewhat oversized genome (relative to an AAV wild-type genome of 4.7 kb). All other vCS vector preparations show the same genomic size (data not shown).

[00362] In order to confirm the expected pattern of capsid proteins, SDS PAGE followed by silver staining was performed with the vectors vCS01, vCS04, and vCS40 (Figure 25). As shown in the figure, the downstream purification procedure resulted in highly purified material displaying the expected protein pattern of VP1, VP2 and VP3 (Figure 25, lanes 2-4). The same pattern was seen with all other viral preparations (not shown). The SDS-PAGE procedure of AAV preparations was done according to standard procedures. Each lane contained 1E10 vg of the respective viral construct, and were separated on a 4-12% Bis-Tris (NuPAGE® Novex, Life Technologies) gel as per manufacturer's instructions. Silver staining was performed with a SilverQuest™ kit (Novex, Life Technologies) according to the manufacturer's instructions.

[00363] Surprisingly, AAV vectors vCS01 and vCS04 had higher virion packaging, measured by higher yields in AAV virus production, as compared to the vCS40 wild-type coding construct and the other codon-optimized constructs. As shown in **Table 3**, the vCS01 and vCS04 vectors replicated substantially better than vCS40, providing a 5-7 fold yield increase in AAV titer.

Table 3 – Yields per liter cell culture obtained with AAV vector constructs vCS01, vCS04, and vCD40, as purified from cell pellets.

Construct	Vector concentration [vg/ml] x10E12	Yields [vg /liter] x10E12	Fold increase vs wt
vCS40	2.0	11.0	-
vCS01	9.2	51.4	4.7
vCS04 – Sample 1	17.6	79.2	7.2
vCS04 – Sample 2	15.9	58.8	5.4

Example 2 – In Vivo Expression of Codon-Altered Factor VIII Variant Expression Sequences

[00364] To test the biological potency of the codon-altered Factor VIII variant sequences, the ReFacto-type FVIII constructs described in Example 1 were administered to mice lacking Factor VIII. Briefly, the assays were performed in C57Bl/6 FVIII knock-out (ko) mice (with 6-8 animals per group) by tail vein injection of 4E12 vector genomes (vg) per kilogram body weight of mouse. Blood was drawn 14 days after injection by retroorbital puncture and plasma was prepared and frozen using standard procedures. Expression levels at day 14 were chosen because there is minimal influence of inhibitory antibodies at this time, which are seen in some animals of this mouse model at later times. FVIII activity in the mouse plasma was determined using the Technochrome FVIII assay performed, with only minor modifications, as suggested by the manufacture (Technoclone, Vienna, Austria). For the assay, the plasma samples were appropriately diluted and mixed with assay reagents, containing thrombin, activated factor IX (FIXa), phospholipids, factor X and calcium. Following FVIII activation by thrombin a complex with FIXa, phospholipids and calcium is formed. This complex activates FX to activated FX (FXa) which in turn cleaves para-nitroanilide (pNA) from the chromogenic substrate. The

kinetics of pNA formation is measured at 405 nm. The rate is directly proportional to the FVIII concentration in the sample. FVIII concentrations are read from a reference curve and results are given in IU FVIII/milliliter.

[00365] The results, presented in **Table 4** below, demonstrate that the codon-altered sequences designed using commercial algorithms (CS10, CS11, and CH25) provided only a modest increase in BDD-Factor VIII (3-4 fold) as compared to the wild-type BDD-Factor VIII construct (CS40). Similarly, the codon-altered BDD-Factor VIII construct prepared as described in Radcliffe et al. (CS08), only provided a 3-4 fold increase in BDD-FVIII expression. This result is consistent with the results reported in Radcliff et al. Surprisingly, the CS01, CS04, and CS23 constructs provided much higher BDD-FVIII expression in the in-vivo biopotency assays (18-, 74-, and -30-fold increases, respectively).

Table 4 – Expression of FVIII in the plasma of FVIII-knock-out mice induced by the different AAV vector constructs.

Construct	Codon Algorithm	Average FVIII Expression at Day 14 [IU/ml]	Standard deviation	Number of mice	Fold increase vs wt
vCS40	Human wild-type	0.03	0.03	12	-
vCS01	Applicants'	0.55	0.28	22	18.3
vCS04	Applicants'	2.21	1.20	55	73.7
vCS08	Radcliffe et al.	0.11	0.01	6	3.6
vCS10	Eurofins	0.09	0.01	7	3.0
vCS11	IDT	0.08	0.02	8	2.7
vCH25	GeneArt	0.13	0.12	18	4.3
vCS23	Applicants'	0.91	0.32	5	30.3

Example 3 – Design of Glycosylation Peptides for the B-domain Substituted Linker

[00366] Others have shown that inclusion of a small peptide (the “V3 peptide”) containing six putative N-linked glycosylation sites from the wild-type Factor VIII B-domain, into a B-domain deleted gene therapy construct, increased Factor VIII levels in the plasma of mice

(McIntosh et al., Blood 121(17):3335-44 (2013)). However, in order to maintain the small size of the B-domain substituted linker, the glycosylation sites were taken out of the context of the wild-type B-domain. In silico prediction (Gupta et al., Supra) of the linker containing the V3 peptide suggests that only two of these glycosylation sites in the V3 peptide will be modified in vivo (Figure 15).

[00367] Thus, Applicants attempted to identify alternative glycosylation peptides that would support higher levels of glycosylation in vivo, which matched wild type glycosylation more closely than the V3 peptide. Applicants designed and tested several alternative glycosylation peptides, in silico. Several of these peptides, shown in Figures 13A-13B, were predicted to have equal or greater glycosylation in vivo than the V3 peptide, when placed between amino acids N768 and P769 of the B-domain substituted linker in SEQ ID NO:2. The results of the in silico predictions are shown in **Table 5**, below. **Table 5** also reports the results of expression experiments performed for several constructs encoding a ReFacto-type Factor VIII protein with a glycosylation peptide incorporated into the B-domain substituted linker, in a CS01 codon-optimized background.

Table 5 – Prediction of N-glycosylation in B-domain substituted linker peptides and performance of AAV vector constructs in vivo.

Sequence	Number of Predicted N-glycosylation sites	Day 28 expression [IU/ml]	SD	Number of mice [n]	Fold expression
vCS01	0	0.74	0.52	5	21
vNG1/CS01	4	n.d.	-	-	-
vNG4/CS01	3	1.93	0.57	6	55
vNG5/CS01	2	n.d.	-	-	-
vNG6/CS01	1	0.80	0.67	5	23
vNG9/CS01	1	n.d.	-	-	-
vNG10/CS01	2	2.66	0.52	6	76
vNG16/CS01	2	1.59	0.57	6	45
vNG17/CS01	2	n.d.	-	-	-
vNG18/CS01	2	n.d.	-	-	-
vNG19/CS01	2	0.88	0.25	5	25
vNG20/CS01	2	n.d.	-	-	-

vNG21/CS01	2	n.d.	-	-	-
vCS40	0	0.035	0.030	12	1

[00368] AAV vectors containing the NG variants were constructed as described in Example 1 and tested in FVIII knock-out mice as described in Example 2. All virus vectors (except the control vector vCS40) shown in **Table 5** are based on the algorithm as used in vCS01. A parallel set of constructs using the algorithm of vCS04 was also prepared (vNG/CS04 series) and is tested in the mouse model. Results were compared to the expression levels achieved with the wild-type vCS40 construct. The day 28 expression levels were chosen in this example, because expression levels of the majority of construct reached the highest levels at this time point. Three AAV vectors achieved greater than 40-fold FVIII expression levels including vNG4/CS01, vNG10/CS01 and vNG16/CS01 (**Table 5**). The corresponding constructs vNG4/CS04, vNG10/CS04 and vNG16/CS04 are expected to show even higher expression because they are based on the superior vCS04 algorithm.

[00369] Surprisingly, the AAV vectors of the vNG/CS01 series had higher virion packaging, measured by higher yields in AAV virus production, as compared to the vCS40 wild-type coding construct. As shown in **Table 6**, the vNG/CS01-based vectors replicated substantially better than vCS40, providing an approximately 3- fold yield increase in AAV titer.

[00370] **Table 6** – Yields per liter cell culture obtained with AAV vector constructs as purified from cell pellets.

Sequence	Vector conc. [vg/ml] $\times 10^{12}$	Yields [vg/liter] $\times 10^{12}$	Fold increase vs wild-type
vCS01	9.17	51.35	4.7
vNG1/CS01	2.13	17.04	1.5
vNG4/CS01	5.74	33.01	3.0
vNG5/CS01	6.91	27.29	2.5
vNG6/CS01	7.01	40.66	3.7
vNG9/CS01	6.39	29.39	2.7
vNG10/CS01	8.57	37.71	3.4
vNG16/CS01	5.3	28.36	2.6

vNG17/CS01	4.24	32.22	2.9
vNG18/CS01	6.11	37.88	3.4
vNG19/CS01	9.42	39.56	3.6
vNG20/CS01	4.09	30.27	2.8
vNG21/CS01	n.d	-	-
vCS40	2.03	11	1.0

Example 4 – Construction of Mutant BDD-FVIII constructs

[00371] Numerous different mutated Refacto-type BDD-FVIII constructs, carrying amino acid mutations within the Factor VIII heavy chain and/or B-domain substituted linker, were cloned and screened. The corresponding vectors, as referred to herein as the “vCS” series of vectors, encode BDD-FVIII variants in the CS01, CS04, and CS23 codon-altered backgrounds. The method used to construct the CS01 and CS04 backgrounds is described in Example 1. The method used to construct CS23 was based on the JCAT tool (www.jcat.de), an on-line tool for codon-optimizations (Grote et al., 2005; Nucl. Acids Res. W526-31). The sequence was further modified to better reflect the codon usage of the albumin superfamily (Mirsafian et al., Sc. Word Journal, ID 639682 (2014)), the content of which is hereby expressly incorporated by reference, in its entirety, for all purposes.

[00372] Combinations of three types of mutations were included in the FVIII sequences of the vCS series of constructs. The first amino acid change introduced into the FVIII sequence is the X1 mutation (TTYVNRSL (SEQ ID NO: 33); X. Xiao), which introduces an additional glycosylation site near the B-domain substituted linker. The X1 mutation is also referred to herein as the “m3” mutation. The second amino acid change made in the FVIII sequence includes the F328S (SPI, F309S SPE) mutation, an amino acid change known to improve secretion of FVIII (Swaaroop, J. Biol. Chem., 272:24121–24 (1997)). This mutation is also referred to herein as the “m1” mutation. The third change is the so-called X5 mutation, which is a combination of five amino acid changes in the A1 domain of the heavy chain that improves specific activity and secretion of BDD-FVIII (Cao et al., 2014; ASGCT abstract #460; details of mutations disclosed in oral presentation). The X5 mutation is also referred to herein as the “m2” mutation. Next, combinations of X1 and F328S (SPI, F309S SPE) were made, followed by

combinations of X1 and X5, also referred to as “X6,” and yet other combinations of X5 and F328S (SPI, F309S SPE) were made (**Table 7**).

[00373] **Gene synthesis and cloning of the vector plasmids.** The plasmids were constructed by cloning different synthetic DNA fragments into the same vector backbone plasmid (pCh-BB01). DNA synthesis of the Refacto-type BDD-FVIII fragments with flanking *AscI* and *NotI* enzyme restriction sites were done by ThermoFischer Scientific (Regensburg, Germany). The vector backbone contains two flanking AAV2-derived inverted terminal repeats (ITRs) that encompass a promoter/enhancer sequence derived from the liver-specific murine transthyretin gene, *AscI* and *NotI* enzyme restriction sites for insertion of the respective Refacto-type BDD-FVIII, and a synthetic polyA site. After ligation of the prepared vector backbone and insertions via the *AscI* and *NotI* sites, the resulting plasmids were amplified in milligram scale. The Refacto-type BDD-FVIII sequences of the constructs were verified by direct sequencing (Microsynth, Balgach, Switzerland). The cloning resulted in different plasmid constructs, as shown in Figure 44.

[00374] **Small scale vector preparations and quantification by quantitative PCR (qPCR).** AAV8-based vectors were prepared by the three plasmid transfection method essentially as described in Grieger et al. (2015, Supra). HEK293 suspension cells were used for plasmid transfections using the corresponding FVIII vector plasmid, the helper plasmid pXX6X80 (carrying adenoviral helper genes) and the packaging plasmid pGSK2/8 (contributing the rep2 and cap8 genes). In the downstream process the cell pellet of a one liter culture was processed using iodixanol gradients as described above. The procedure resulted in vector preparations as outlined in **Table 8**. Vectors were quantified by qPCR using the universal qPCR procedure targeting the AAV2 inverted terminal repeats (Aurnhammer, HUMAN GENE THERAPY METHODS: Part B 23:18–28 (2012)). An accurately quantified vector plasmid carrying AAV2 Inverted terminal repeats served for preparing the standard curve.

[00375] **AAV vector characterizations.** The integrity of the vector genome was analyzed by AAV agarose gel electrophoresis. The electrophoresis was done similar as described in Fagone et al. (Human Gene Therapy Methods, 23:1-7 (2012)). AAV vector preparations were incubated at 75 °C for 10 minutes in the presence of 0.5% SDS and then cooled down to room

temperature. Approximately 1.5×10^{10} vector genomes (vg) were loaded per lane on a 1% 1xTAE agarose gel and electrophoresed for 60 min at 7 V/cm of gel length. The gel was then stained in 2x GelRed (Biotium Cat# 41003) solution and imaged by ChemiDocTMMP (Biorad). The results of a selection of vectors are shown in Figure 45. The viral vectors vCS04 (control), vCS17, vCS20, vCS24, vCS16 and vCS40 (control) show all the same-sized genome as a distinct band in the 5kb range (Figure 45, lanes 2-7; arrow right side). Despite a vector size of approx. 5.2 kb, the genome is a homogenous band confirming correct packaging of the somewhat oversized genome (relative to an AAV wild-type genome of 4.7 kb).

[00376] In order to confirm purity of the vector and the expected pattern of capsid proteins, SDS PAGE followed by silver staining was performed with the vectors, as shown in Figure 46. As shown in the figure, the downstream purification procedure resulted in highly purified material displaying the expected protein pattern of VP1, VP2 and VP3 (Figure 46 lanes 2-9; arrows right hand side). The SDS-PAGE procedure of AAV preparations was done according to standard procedures. The amounts of 1×10^{10} vg per lane were separated on a 4-12% Bis-Tris (NuPAGE® Novex, Life Technologies) gel as per manufacturer's instructions. Silver staining was performed with a SilverQuestTM kit (Novex, Life Technologies) according to the instructions of the manufacturer.

[00377] **In-vivo biopotency screening of vectors.** The different Refacto-type BDD-FVIII constructs were screened in mice. The assay was performed in C57Bl/6 FVIII knock-out (ko) mice (with 6-8 animals per group) by tail vein injection of 4×10^{12} vector genomes (vg) per kilogram body weight of mouse. Blood was drawn 14 days after injection by retroorbital puncture and plasma was prepared and frozen using standard procedures. FVIII activity in mouse plasma was determined with a chromogenic assay from Technoclone with minor modifications (Technochrome FVIII, Technoclone, Vienna, Austria). In brief, the plasma sample was appropriately diluted and mixed with assay reagents, containing thrombin, activated factor IX (FIXa), phospholipids, factor X and calcium. Following FVIII activation by thrombin a complex with FIXa, phospholipids and calcium is formed. This complex activates FX to activated FX (FXa) which in turn cleaves para-nitroanilide (pNA) from the chromogenic substrate. The kinetics of pNA formation is measured at 405 nm. The rate is directly

proportional to the FVIII concentration in the sample. FVIII concentrations are read from a reference curve and results are given in IU FVIII/milliliter.

[00378] The results of the mouse biopotency assay (day 14 expression data of FVIII in international units per milliliter [IU/ml] in mouse plasma and fold expression compared to the wild-type vCS40 control) are shown in **Table 7**. AAV vectors vCS19, vCS26 and vCS32 all contain the X1 glycosylation site in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. As seen in **Table 7**, surprisingly high expression levels were obtained, as compared to the wild-type construct vCS40 (level defined as 1). vCS26, for instance, expressed 202-fold higher levels compared to the wild-type vCS40 vector. Another control construct for the X1-series of vectors, vCH111, that contains the X1 mutation in the Geneart codon context, showed a more modest increase in expression (12-fold).

[00379] **Vectors** vCS16, vCS28, and vCS34 all contain the F328S (SPI, F309S SPE) mutation enhancing secretion in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. As seen in **Table 7**, high expression levels (45-93-fold higher than the wt vCS40 control) were obtained with vCS16 and vCS28.

[00380] **Vectors** vCS20, vCS24, and vCS33 contain the X5 mutation in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. The best performing variant in the X5 series was vCS20, achieving levels of >3 units/ml after day 14 and a 121-fold increase over the wt vCS40 control.

[00381] **Vectors** vCS17, vCS29, and vCS31 contain the combination of the X1 and F328S (SPI, F309S SPE) mutations in the CS01, CS04, and CS23 codon-altered backgrounds, respectively (Table 6). The vCS17 and vCS29 constructs achieved very high expression levels in the mouse studies (115 to 246-fold increase over the vCS40 control). Remarkably, in the FVIII KO mouse model used, the majority of mice treated with the vCS17 construct did not develop neutralizing antibodies over time, evidenced by increasing levels of FVIII at later time points (e.g., day 28 and day 42; data not shown). This is an unexpected finding, because in some other constructs the expression levels began to decrease with time due to the formation of neutralizing antibodies. The CS01 background combined with the secretion-enhancing mutations F328S (SPI, F309S SPE) and X1 resulted in low immunogenicity induction.

[00382] **Vectors** vCS18, vCS27, and vCS35 contain the combination of the X1 and X5 mutations in the CS01, CS04, and CS23 codon-altered backgrounds, respectively. The combination of these two mutations was also very efficient. A 145-fold increase over the vCS40 control could be achieved with vCS18, for example (**Table 7**).

[00383] Vectors vCS48 and vCS49 contain the combination of the X5 and F328S (SPI, F309S SPE) mutations in the CS01 and CS04 codon-altered backgrounds, respectively. The combination of these two mutations was also very efficient. One of the largest increases of all mutants, a 239-fold increase over the vCS40 control, could be achieved with vCS49 confirming the special value of the combinations including the F328S (SPI, F309S SPE) mutation.

[00384] A further surprising observation was that the mutant AAV vectors grew substantially better than the vCS40 construct harboring the wild-type BDD-FVIII codons. Sequence optimization resulted in a several-fold yield increase in vector production. In some of the best expressing constructs (e.g., vCS29, vCS17, vCS20, and vCS26) the increase in yields due to codon-alteration and/or mutant sequence was approximately 3-5-fold higher, as compared to the wild-type vector (**Table 8**).

[00385] Expression of BDD-FVIII in the plasma of FVIII-knock-out mice induced by the different AAV vector constructs is shown in **Table 7**. The constructs have the same vector backbone, however, encode different types of mutated FVIII, including different codon optimization backgrounds. Expression levels at day 14 were chosen because at this time point there is minimal influence of inhibitory antibodies usually seen in some animals in the mouse model at later times. N.d., not determined.

Table 7 – In vivo biopotency data of vCS constructs.

#	Vector	Algorithm, mutations	Day 14 expression [IU/ml]	SD	Number of mice [n]	Fold expression
1	vCS19	CS01, X1	2.34	1.10	13	78
2	vCS26	CS04, X1	6.07	2.72	12	202
3	vCS32	CS23, X1	n.d.	-	-	-
4	vCS16	CS01, F328S	1.35	0.88	6	45
5	vCS28	CS04, F328S	2.78	0.92	7	93
6	vCS34	CS23, F328S	n.d.	-	-	-
7	vCS20	CS01, X5	3.62	1.96	21	121
8	vCS24	CS04, X5	0.79	0.89	18	26
9	vCS33	CS23, X5	n.d.	-	-	n.d.
10	vCS17	CS01, X1, F328S	3.44	1.92	20	115
11	vCS29	CS04, X1, F328S	7.39	2.64	9	246
12	vCS31	CS23, X1, F328S	n.d.			n.d.
13	vCS18	CS01, X1+X5 (X6)	4.34	2.50	6	145
14	vCS27	CS04, X1+X5 (X6)	8.03	3.97-	6-	268-
15	vCS35	CS23, X1+X5 (X6)	n.d.	-	-	-
19	vCS48	CS01, X5, F328S	2.54	0.72	8	85
20	vCS49	CS04, X5, F328S	7.17	1.30	7	239
	controls					
16	vCS40	Human wild-type	0.03	0.03	12	1
17	vCh25	Geneart	0.13	0.12	18	4
18	vCh111	Geneart +X1	0.37	0.21	17	12

Table 8 – Yields per liter cell culture (packaging efficiency) obtained with the different AAV vector constructs. The vectors were purified out of the cell pellets; n.d., not determined.

	construct	Algorithm, mutations	Vector conc. [vg/ml] $\times 10^{12}$	Yields [vg /liter] $\times 10^{12}$	Fold increase vs wt
1	vCS19	CS01, X1	9.71	36	3.22
2	vCS26	CS04, X1	5.93	32	2.87
3	vCS32	CS23, X1	n.d.	n.d.	n.d.
4	vCS16	CS01, F328S	6.51	29	2.56
5	vCS28	CS04, F328S	5.85	32	2.88
6	vCS34	CS23, F328S	n.d.	n.d.	n.d.
7	vCS20	CS01, X5	9.90	50	4.48
8	vCS24	CS04, X5	3.00	16	1.46
9	vCS33	CS23, X5	n.d.	n.d.	n.d.
10	vCS17	CS01, X1, F328S	8.94	37	3.34
11	vCS29	CS04, X1, F328S	7.42	53	4.72
12	vCS31	CS23, X1, F328S	n.d.	n.d.	n.d.
13	vCS18	CS01, X1+X5 (X6)	21.20	53	4.75
14	vCS27	CS04, X1+X5 (X6)	4.15	19	1.67
15	vCS35	CS23, X1+X5 (X6)	n.d.	n.d.	n.d.
16	vCS48	CS01, X5, F328S	7.14	42.1	3.77
17	vCS49	CS04, X5, F328S	8.27	37.2	3.33
18	vCS40	Human wild-type	2.03	11	1.00

[00386] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1. A polynucleotide comprising a nucleotide sequence encoding a Factor VIII polypeptide, the Factor VIII polypeptide comprising a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain,
wherein the heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 3;
wherein the light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 4; and
wherein the polypeptide linker comprises a furin cleavage site and a glycosylation peptide having an amino acid sequence of SEQ ID NO:55.
2. The polynucleotide of claim 1, wherein the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity over the entire length of SEQ ID NO: 38.
3. The polynucleotide of claim 1 or 2, wherein the polypeptide linker is encoded by a third nucleotide sequence of SEQ ID NO:38.
4. The polynucleotide of any one of claims 1-3, wherein:
the first nucleotide sequence is SEQ ID NO: 3; and
the second nucleotide sequence is SEQ ID NO: 4.
5. The polynucleotide of claim 4, wherein the polypeptide linker is encoded by a third nucleotide sequence of SEQ ID NO:38.
6. The polynucleotide according to any one of claims 1 to 5, further comprising a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.
7. The polynucleotide of claim 6, wherein the promoter element is a liver-specific promoter sequence upstream of the nucleotide sequence encoding the Factor VIII polypeptide.

8. The polynucleotide of claim 7, further comprising an intron sequence positioned between the liver-specific promoter sequence and the nucleotide sequence encoding the Factor VIII polypeptide.

9. An adeno-associated virus (AAV) vector comprising a polynucleotide according to any one of claims 1 to 8.

10. An adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 1 to 8.

11. A host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 1 to 8.

12. A method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to any one of claims 1 to 8 into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

13. Use of an adeno-associated virus (AAV) particle according to claim 10 in the manufacture of a medicament for treating hemophilia A in a patient in need thereof.

14. An *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to claim 10.

15. A polynucleotide comprising a nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO:1, wherein the polynucleotide encodes a Factor VIII polypeptide having I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19.

16. The polynucleotide of claim 15, wherein:
the Factor VIII polypeptide comprises a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain; and

the polypeptide linker includes a glycosylation peptide having an amino acid sequence of SEQ ID NO:55 inserted between the heavy chain and the light chain.

17. The polynucleotide of claim 15 or 16, comprising the sequence of SEQ ID NO:94.

18. The polynucleotide of any one of claims 15-17, further comprising a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

19. The polynucleotide of claim 18, wherein the promoter element is a liver-specific promoter sequence upstream of the nucleotide sequence encoding the Factor VIII polypeptide.

20. The polynucleotide of claim 19, further comprising an intron sequence positioned between the liver-specific promoter sequence and the nucleotide sequence encoding the Factor VIII polypeptide.

21. An adeno-associated virus (AAV) vector comprising a polynucleotide according to any one of claims 15-20.

22. An adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 15-20.

23. A host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 15-20.

24. A method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to any one of claims 15-20 into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

25. Use of an adeno-associated virus (AAV) particle according to claim 22 in the manufacture of a medicament for treating hemophilia A in a patient in need thereof.

26. An *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to claim 22.

27. A polynucleotide comprising a nucleotide sequence encoding a Factor VIII polypeptide, the Factor VIII polypeptide comprising a light chain, a heavy chain, and a polypeptide linker joining the C-terminus of the heavy chain to the N-terminus of the light chain, the Factor VIII polypeptide having I105V, A127S, G151K, M166T, and L171P amino acid substitutions, relative to SEQ ID NO: 19,

wherein the heavy chain of the Factor VIII polypeptide is encoded by a first nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 3;

wherein the light chain of the Factor FVIII polypeptide is encoded by a second nucleotide sequence having at least 99% identity over the entire length of SEQ ID NO: 4; and

wherein the polypeptide linker comprises a furin cleavage site and a glycosylation peptide having an amino acid sequence of SEQ ID NO:55.

28. The polynucleotide of claim 27, wherein the polypeptide linker is encoded by a third nucleotide sequence having at least 95% identity over the entire length of SEQ ID NO:38.

29. The polynucleotide of claim 27 or 28, wherein the polypeptide linker is encoded by a third nucleotide sequence of SEQ ID NO:38.

30. The polynucleotide of any one of claims 27-29, further comprising a promoter element operably linked to the polynucleotide encoding the Factor VIII polypeptide.

31. The polynucleotide of claim 30, wherein the promoter element is a liver-specific promoter sequence upstream of the nucleotide sequence encoding the Factor VIII polypeptide.

32. The polynucleotide of claim 31, further comprising an intron sequence positioned between the liver-specific promoter sequence and the nucleotide sequence encoding the Factor VIII polypeptide.

33. An adeno-associated virus (AAV) vector comprising a polynucleotide according to any one of claims 27-32.

34. An adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 27-32.

35. A host cell infected with an adeno-associated virus (AAV) particle comprising a polynucleotide according to any one of claims 27-32.

36. A method for producing an adeno-associated virus (AAV) particle comprising introducing a polynucleotide according to any one of claims 27-32 into a mammalian host cell, wherein the polynucleotide is competent for replication in the mammalian host cell.

37. Use of an adeno-associated virus (AAV) particle according to claim 34 in the manufacture of a medicament for treating hemophilia A comprising in a patient in need thereof.

38. An *in vitro* method for transducing a host cell comprising contacting the host cell with an adeno-associated virus (AAV) particle according to claim 34.