(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 April 2020 (30.04.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/084162 A1

(51) International Patent Classification:
A61K 48/00 (2006.01)

(21) International Application Number:
PCT/EP2019/079358

(22) International Filing Date:
28 October 2019 (28.10.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
18202986.8 26 October 2018 (26.10.2018) EP

(71) Applicant: **VRIJE UNIVERSITEIT BRUSSEL** [BE/BE]; Pleinlaan 2, 1050 Brussel (BE).

(72) Inventors: **CHUAH, Lay Khim**; Tiensesteenweg 250, 3360 Bierbeek (BE). **VANDENDRIESSCHE, Thierry**; Tiensesteenweg 250, 3360 Bierbeek (BE).

(74) Agent: **DE CLERCQ & PARTNERS**; Edgard Gevaert-dreef 10a, 9830 Sint-Martens-Latem (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *of inventorship (Rule 4.17(iv))*

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** LIVER-SPECIFIC NUCLEIC ACID REGULATORY ELEMENTS AND METHODS AND USE THEREOF

(57) **Abstract:** The present invention relates to nucleic acid regulatory elements that are able to enhance liver-specific expression of genes, methods employing these regulatory elements and uses of these elements. Expression cassettes and vectors containing these nucleic acid regulatory elements are also disclosed. The present invention is particularly useful for applications using gene therapy, more particularly liver- directed gene therapy, and for vaccination purposes.

WO 2020/084162 A1

LIVER-SPECIFIC NUCLEIC ACID REGULATORY ELEMENTS AND METHODS AND USE THEREOF

FIELD

The present invention relates to nucleic acid regulatory elements that are able to enhance liver-specific expression of genes, methods employing these regulatory elements and uses thereof. The invention further encompasses expression cassettes, vectors and pharmaceutical compositions comprising these regulatory elements. The present invention is particularly useful for applications using gene therapy, more particularly liver-directed gene therapy, and for vaccination purposes.

BACKGROUND

Convincing evidence continues to emerge from clinical trials that gene therapy is yielding therapeutic effects in patients suffering from a wide range of diseases. In particular, liver-directed gene therapy is a promising modality to obtain sustained hepatocyte-specific expression of secreted factors into the circulation. Indeed, many acquired, complex and genetic diseases (hepatic diseases *sensu stricto* as well as some hereditary disorders that do not directly lead to liver disease but manifest themselves primarily elsewhere in the body, e.g. hemophilia A or B, familial hypercholesterolemia, ornithine transcarbamylase deficiency, or α-antitrypsin deficiency) are associated with altered gene expression in the liver. In addition, the liver often falls prey to infections with pathogens (such as hepatitis viruses). Finally, the liver can undergo malignant transformation and give rise to liver cancer (hepatocellular carcinoma) or functionally degenerate as a consequence of pharmaceutical treatments and chemotherapy, drug or alcohol abuse. Consequently, there has been substantial and increasing interest in the use of gene therapy to express a functional gene in the liver to replace a needed protein or to block the expression of an altered or undesired gene product, for instance by RNA interference or dominant-negative inhibitory proteins, or to restore hepatocyte function in a degenerating liver. Liver-directed gene therapy can also be used to express anti-viral compounds, such as VP22 protein fused to the C terminal of hepatitis B virus core protein (i.e. VP22 fusion protein-based dominant negative mutant), which can inhibit hepatitis B virus replication, ISG56 and IFITM1, which can inhibit hepatitis C virus replication, and interferon-inducible MyD88 protein, which can inhibit hepatitis B virus replication. Transduction of liver cells with immunomodulatory cytokines may be useful to induce immune responses against e.g. viral hepatitis or liver neoplasms, or to suppress immune responses towards the gene delivery vectors or the transgene. Another application of transgene delivery to the liver is for DNA vaccination. The optimisation of antigen expression is an important consideration in DNA vaccine vector design. It is clear that transgene expression may be increased through the use of optimised promoters and polyadenylation (polyA) sequences. However, in some circumstances, for example to express antigens that induce cell death upon overexpression, it may be necessary to optimise DNA vaccines to produce reduced transgene expression. Tissue-specificity is also considered important. Also other aspects of vector design may influence the efficacy of the vaccine. A rational approach to improve

the efficacy of DNA vaccination would optimise the: (i) vector backbone DNA sequence; (ii) transgene sequence; (iii) co-expression of stimulatory sequences; (iv) delivery system used for the vector; and (v) targeting of the vector for appropriate immune stimulation (as reviewed in Garmory et al. 2003). The backbone of a DNA vaccine vector could be further modified to enhance immunogenicity via the manipulation of the DNA to include certain sequences, so that the DNA itself will have an adjuvantising effect. DNA vaccine vectors contain many CpG motifs (consisting of unmethylated CpG dinucleotides flanked by two 5' purines and two 3' pyrimidines) that, overall, induce a Th1-like pattern of cytokine production, and are thought to account for strong CTL responses frequently seen following DNA vaccination. It is possible to augment responses to DNA vaccine vectors by incorporating CpG motifs into the DNA backbone of the plasmid. Alternatively, immune responses may be modulated or enhanced by the co-expression of stimulatory molecules or cytokines or through the use of localisation or secretory signals, or ligand to direct antigens to sites appropriate for immune modulation. Finally, a variety of routes of administration of DNA vaccines have been studied, including intramuscular, intradermal, subcutaneous, intravenous, intraperitoneal, oral, vaginal, intranasal and, more recently, non-invasive delivery to the skin.

Efforts to deliver transgenes to liver have focused on vectors derived from adenoviruses, retroviruses, lentiviruses, and adeno-associated viruses (AAV), and plasmids. Adeno-associated viral (AAV) vector is by far the most promising gene delivery vehicle for liver-directed gene therapy. AAV vectors have a favorable safety profile and are capable of achieving persistent transgene expression. Further, several immunologically distinct AAV serotypes have been isolated from human and non-human primates, and they induce minimal immune responses, which make AAV vectors well suited for liver-directed gene therapy. The major limitation of AAV vectors is the limited packaging capacity of the vector particles (i.e. approximately 4.7 kb), constraining the size of the transgene expression cassette to obtain functional vectors.

There remain however concerns regarding the efficacy and safety of some gene delivery approaches. The major limiting factors are: insufficient and/or transient transgene expression levels, and inappropriate expression of the transgene in unwanted cell types. Higher vector doses are typically used in gene therapy clinical trials to improve therapeutic efficacy. However, this often triggers T-cell–mediated immune responses against the vector capsid antigens displayed by transduced cells, particularly hepatocytes, in the context of MHC class I. This contributes to the elimination of the gene-modified cells and liver toxicity, resulting in short-term gene expression. Moreover, inadvertent transgene expression in antigen-presenting cells (APCs), increases the risk of untoward immune responses against the gene-modified hepatocytes and/or the therapeutic transgene product. Consequently, there is a need to generate improved gene therapy vectors allowing the use of lower and safer vector doses that enable sustained hepatocyte-specific expression of the therapeutic gene. The availability of more potent vectors would also ease manufacturing needs.

Conventional methods of vector design relied on haphazard trial-and-error approaches whereby transcriptional enhancers were combined with promoters to boost expression levels. Though this could sometimes be effective, it often resulted in non-productive combinations that resulted in either modest or no increased expression levels of the gene of interest and/or loss of tissue specificity. Moreover, these conventional approaches did not take into account the importance of including evolutionary conserved regulatory motifs into the expression modules, which is particularly relevant for clinical translation.

A computational approach depending upon a modified distance difference matrix (DDM) – multidimensional scaling (MDS) strategy (De Bleser et al. 2007. Genome Biol 8, R83) has proven to be useful for the *in silico* identification of clusters of evolutionary conserved transcription factor binding site (TFBS) motifs associated with robust tissue-specific expression in liver (WO 2009/130208) and heart (WO2011/051450). The combination of one of these liver-specific regulatory elements, in particular the Serpin enhancer, with a synthetic codon-optimized hyperfunctional FIX transgene (i.e., Padua R338L) has been shown to significantly increase FIX expression and activity (WO 2014/064277).

It is an object of the present invention to further increase the efficiency and safety of liver-directed gene therapy.

SUMMARY

The present inventors have relied on a computational approach to identify evolutionarily conserved transcription factor binding site (TFBS) motifs associated with highly expressed liver-specific genes defined herein as nucleic acid regulatory elements. This required several consecutive computational steps: (1) liver-specific genes were identified that are highly expressed based on RNAseq (RNA sequencing) expression data obtained with normal human tissues; (2) publicly available databases were used to extract the corresponding promoter sequences; (3) a computational approach was employed to identify clusters of transcription factor binding site motifs (TFBS); (4) the genomic context of the highly expressed genes was screened for evolutionary conserved clusters of TFBS (i.e. CREs). These regulatory elements were subsequently validated *in vivo* yielding efficient gene expression. The nucleic acid regulatory elements identified herein, allow for the use of lower and thus safer vector doses in gene therapy, while maximizing therapeutic efficacy.

The invention therefore provides the following aspects:

Aspect 1. A nucleic acid regulatory element for enhancing liver-specific gene expression, comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 20, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof; wherein said nucleic acid regulatory element comprises at least a transcription factor binding site (TFBS) for SP1 and for EP300.

Aspect 2. The nucleic acid regulatory element according to aspect 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4 and SEQ ID NO: 12, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof; wherein said nucleic acid regulatory element further comprises a TFBS for HNF 4G, a TFBS for CEBPB, a TFBS for P300, a TFBS for HDAC2, a TFBS for JUND, a TFBS for FOSL2, a TFBS for ZBTB7A, a TFBS for CEBPD and a TFBS for RXRA.

Aspect 3. The nucleic acid regulatory element according to aspect 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 13, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof; wherein said nucleic acid regulatory element further comprises a TFBS for POLR2A, a TFBS for MYBL2, a TFBS for FOXA1, a TFBS for FOXA2, a TFBS for ARID3A, a TFBS for POLR2A and a TFBS for HEY1.

Aspect 4. The nucleic acid regulatory element according to aspect 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 14, SEQ ID NO: 16 and SEQ ID NO: 20, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof; wherein said nucleic acid regulatory element further comprises a TFBS for HNF 4G, a TFBS for CEBPB, a TFBS for P300 and a TFBS for HNF 4A.

Aspect 5. The nucleic acid regulatory element according to any one of aspects 1 to 4, wherein said functional fragment comprises at least 20, preferably at least 25, contiguous nucleotides from the sequence from which it is derived, and wherein said functional fragment comprises at least 1, preferably at least 5, of the TFBS that are present in the sequence from which it is derived.

Aspect 6. A nucleic acid regulatory element for enhancing liver-specific gene expression hybridizing under stringent conditions to the nucleic acid regulatory element according to any one of aspect 1 to 5, or to its complement.

Aspect 7. The nucleic acid regulatory element according to any one of aspects 1 to 6, having a maximal length of 600 nucleotides, preferably of 500 nucleotides, more preferably of 400 nucleotides, even more preferably of 300 nucleotides.

Aspect 8. A nucleic acid expression cassette comprising at least one nucleic acid regulatory element according to any one of aspects 1 to 7, operably linked to a promoter and a transgene.

Aspect 9. The nucleic acid expression cassette according to aspect 8, further comprising at least one, preferably three, liver-specific regulatory element(s) different from the nucleic acid regulatory element according to any one of aspects 1 to 4, preferably wherein said at least one, preferably three, liver-specific regulatory element(s) comprise SEQ ID NO: 22 or a sequence having at least 95% identity to said sequence.

Aspect 10. The nucleic acid expression cassette according to aspect 9, comprising a nucleic acid regulatory element comprising SEQ ID NO: 6, SEQ ID NO: 4; SEQ ID NO: 12, or a sequence having at

least 95% identity to any one of said sequences, and at least one, preferably three, nucleic acid regulatory element(s) comprising SEQ ID NO: 22 or a sequence having at least 95% identity to said sequence.

Aspect 11. The nucleic acid expression cassette according to any one of aspects 8 to 10, wherein the transgene encodes for a therapeutic protein or an immunogenic protein.

Aspect 12. The nucleic acid expression cassette according to aspect 11, wherein the transgene encodes for coagulation factor IX (FIX), preferably wherein said transgene is codon-optimized coagulation factor FIX, or wherein said coagulation factor FIX contains a hyper-activating mutation, preferably wherein said hyper-activating mutation corresponds to an R338L amino acid substitution, more preferably wherein said transgene encoding for coagulation factor FIX has a nucleic acid sequence as defined by SEQ ID NO: 25.

Aspect 13. The nucleic acid expression cassette according to aspect 11, wherein the transgene encodes for coagulation factor VIII (FVIII), preferably wherein said transgene is codon-optimized coagulation factor FVIII, or wherein said coagulation factor VIII has a deletion of the B domain, preferably wherein said B domain of said FVIII is replaced by a linker defined by SEQ ID NO: 54, more preferably wherein said transgene encodes for coagulation factor VIII has a nucleic acid sequence defined by SEQ ID NO: 26.

Aspect 14. The nucleic acid expression cassette according to any one of aspects 8 to 13, wherein the promoter is a liver-specific promoter, preferably a promoter derived from the transthyretin (TTR) promoter, more preferably the minimal promotor of the transthyretin gene (TTRmin) defined by SEQ ID NO: 27.

Aspect 15. The nucleic acid expression cassette according to aspect 14, further comprising a liver-specific regulatory element comprising SEQ ID NO: 24, thereby comprising the combination of the TTRe and TTRm nucleic acids as defined by SEQ ID NO: 28.

Aspect 16. The nucleic acid expression cassette according to any one of aspects 8 to 15, further comprising a minute virus of mouse (MVM) intron, preferably the MVM intron defined by SEQ ID NO: 29.

Aspect 17. The nucleic acid expression cassette according to any one of aspects 8 to 16, further comprising a transcriptional termination signal derived from the bovine growth hormone polyadenylation signal (BGHpA), preferably the BGHpA defined by SEQ ID NO: 30.

Aspect 18. A vector comprising the nucleic acid regulatory element according to any one of aspects 1 to 7, or the nucleic acid expression cassette according to any one of aspects 8 to 17, preferably a viral vector, more preferably a vector derived from an adeno-associated virus (AAV), even more preferably a self-complementary AAV vector.

Aspect 19. The vector according to aspect 18, having SEQ ID NO: 38, SEQ ID NO: 36 or SEQ ID NO: 44, preferably SEQ ID NO: 38.

Aspect 20. A pharmaceutical composition comprising the nucleic acid expression cassette according to any one of aspects 8 to 17, or the vector according to aspect 18 or 19, and a pharmaceutically acceptable carrier.

Aspect 21. The nucleic acid regulatory element according to any one of aspects 1 to 7, the nucleic acid expression cassette according to any one of aspects 8 to 17, the vector according to aspect 18 or 19, or the pharmaceutical composition according to aspect 20 for use in gene therapy, preferably liver-directed gene therapy, or in vaccination therapy, preferably prophylactic vaccination.

Aspect 22. The nucleic acid regulatory element, the nucleic acid expression cassette, the vector, or the pharmaceutical composition for use according to aspect 21, wherein the gene therapy is for the treatment of hemophilia A or hemophilia B.

Aspect 23. An *in vitro* or *ex vivo* method for expressing a transgene product in liver cells comprising:

- introducing the nucleic acid expression cassette according to any one of aspects 8 to 17, or the vector according to aspect 18 or 19 into the liver cells;

- expressing the transgene product in the liver cells.

Aspect 24. The nucleic acid regulatory elements according to any one of aspects 1 to 7, the nucleic acid expression cassettes according to any one of aspects 8 to 17 or the vectors according to any one of aspects 18 to 19 for use in enhancing liver-specific gene expression of a transgene, preferably wherein said use is an *in vitro* use.

Aspect 25. A method for enhancing liver-specific gene expression of a transgene comprising the use of the nucleic acid regulatory elements according to any one of aspects 1 to 7, the nucleic acid expression cassettes according to any one of aspects 8 to 17 or the vectors according to any one of aspects 18 or 19.

BRIEF DESCRIPTION OF DRAWINGS

FIG 1: Schematic representation of the AAVsc constructs. Design of the AAVsc-HS-CRE-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 33- 53) (A), AAVsc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 31) (B) and AAVsc-3xSERP-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 32) (C). SERP: Serpin; TTRm: minimal transthyretin; TTRenh: TTR enhancer; MVM: minute virus of mouse intron; hFIXcoPadua: codon-optimized human FIX Padua R338L; bghpolyA: polyadenylation signal derived from bovine growth hormone. All constructs used are flanked with AAV inverted terminal repeat.

FIG 2: Comprehensive *in vivo* comparison of transgene expression levels obtained with HS-CRE-containing constructs versus controls without HS-CRE. Adult C57BL/6 male mice were

hydrodynamically transfected with AAVsc-HS-CRE1-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 33), AAVsc-HS-CRE2-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 34), AAVsc-HS-CRE3-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 35), AAVsc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 36), AAVsc-HS-CRE5-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 37), AAVsc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 38), AAVsc-HS-CRE7-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 39), AAVsc-HS-CRE8-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 40), AAVsc-HS-CRE9-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 41), AAVsc-HS-CRE10-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 42), AAVsc-HS-CRE11-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 43), AAVsc-HS-CRE12-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 44), AAVsc-HS-CRE13-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 45), AAVsc-HS-CRE14-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 46), AAVsc-HS-CRE15-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 47), AAVsc-HS-CRE16-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 48), AAVsc-HS-CRE17-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 49), AAVsc-HS-CRE18-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 50), AAVsc-HS-CRE19-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 51), AAVsc-HS-CRE20-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 52), AAVsc-HS-CRE21-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 53), abbreviated as HS-CRE1, HS-CRE2, HS-CRE3, HS-CRE4, HS-CRE5, HS-CRE6, HS-CRE7, HS-CRE8, HS-CRE9, HS-CRE10, HS-CRE11, HS-CRE12, HS-CRE13, HS-CRE14, HS-CRE15, HS-CRE16, HS-CRE17, HS-CRE18, HS-CRE19, HS-CRE20 and HS-CRE21, respectively. As controls, AAVsc-3XSERP-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 32) and AAVsc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 31) were used, indicated as 3XSERP-TTRenh-TTRmin and TTRenh-TTRmin, respectively. The production of the transgene product (i.e. human FIX-Padua) in the plasma was quantified using a human-FIX specific ELISA, 1 day (panel A) and 2 days (panel B) post-transfection.

FIG 3: Comparison of FIX levels (measured as ng/ml mouse plasma) in adult C57BL/6 mice which were injected via the tail vein with $1X10^9$ vector genomes (vg) (A) or $5x10^9$ vg (B) AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA or AAV9sc-SERP3X-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA vectors, indicated in abbreviated form as HS-CRE4, HS-CRE6, no HS-CRE and 3XSERP, respectively. Blood was collected on successive days post-injection into buffered citrate and the concentration of hFIX antigen in citrated plasma was measured by ELISA.

DESCRIPTION

As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

The terms "comprising", "comprises" and "comprised of" as used herein are synonymous with "including", "includes" or "containing", "contains", and are inclusive or open-ended and do not exclude additional, non-recited members, elements or method steps. The terms also encompass "consisting of" and "consisting essentially of", which enjoy well-established meanings in patent terminology.

The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, preferably +/-5% or less, more preferably +/-1% or less, and still more preferably +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" refers is itself also specifically, and preferably, disclosed.

Whereas the terms "one or more" or "at least one", such as one or more members or at least one member of a group of members, is clear per se, by means of further exemplification, the term encompasses inter alia a reference to any one of said members, or to any two or more of said members, such as, e.g., any ≥3, ≥4, ≥5, ≥6 or ≥7 etc. of said members, and up to all said members. In another example, "one or more" or "at least one" may refer to 1, 2, 3, 4, 5, 6, 7 or more.

The discussion of the background to the invention herein is included to explain the context of the invention. This is not to be taken as an admission that any of the material referred to was published, known, or part of the common general knowledge in any country as of the priority date of any of the claims.

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. All documents cited in the present specification are hereby incorporated by reference in their entirety. In particular, the teachings or sections of such documents herein specifically referred to are incorporated by reference.

Unless otherwise defined, all terms used in disclosing the invention, including technical and scientific terms, have the meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. By means of further guidance, term definitions are included to better appreciate the teaching of the invention. When specific terms are defined in connection with a particular aspect of the invention or a particular embodiment of the invention, such connotation is meant to apply throughout this specification, i.e., also in the context of other aspects or embodiments of the invention, unless otherwise defined.

In the following passages, different aspects or embodiments of the invention are defined in more detail. Each aspect or embodiment so defined may be combined with any other aspect(s) or embodiment(s) unless clearly indicated to the contrary. In particular, any feature indicated as being preferred or

advantageous may be combined with any other feature or features indicated as being preferred or advantageous.

Reference throughout this specification to "one embodiment", "an embodiment" means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention, and form different embodiments, as would be understood by those in the art. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

The terms or definitions provided herein are provided to aid in the understanding of the invention. Unless specifically defined herein, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Press, Plainsview, New York (1989); and Ausubel et al., Current Protocols in Molecular Biology (Supplement 47), John Wiley & Sons, New York (1999), for definitions and terms of the art.

In an aspect, the invention relates to a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of (i.e., the regulatory element may for instance additionally comprise sequences used for cloning purposes, but the indicated sequences make up the essential part of the regulatory element, e.g. they do not form part of a larger regulatory region such as a promoter), or consisting of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21 or a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof (i.e. a functional fragment of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21 or of a sequence having high percentage sequence identity to any of said sequences, and still performing the function of said regulatory element, i.e. increasing expression).

A "nucleic acid regulatory element" or "regulatory element", also called "CRE" (cis-regulatory element), "CRM" (cis-regulatory module), or "HS-CRE" (hepatocyte-specific cis-regulatory element) as used herein refers to a transcriptional control element, in particular a non-coding cis-acting transcriptional control element, capable of regulating and/or controlling transcription of a gene, in particular tissue-specific transcription of a gene. Regulatory elements comprise at least one transcription factor binding site (TFBS), more in particular at least one binding site for a tissue-specific transcription factor, most particularly at least one binding site for a liver-specific transcription factor. Typically, regulatory elements as used herein increase or enhance promoter-driven gene expression when compared to the transcription of the gene from the promoter alone, without the regulatory elements. Thus, regulatory elements particularly comprise enhancer sequences, although it is to be understood that the regulatory elements enhancing transcription are not limited to typical far upstream enhancer sequences, but may occur at any distance of the gene they regulate. Indeed, it is known in the art that sequences regulating transcription may be situated either upstream (e.g. in the promoter region) or downstream (e.g. in the 3'UTR) of the gene they regulate *in vivo*, and may be located in the immediate vicinity of the gene or further away. Of note, although regulatory elements as disclosed herein typically comprise naturally occurring sequences, combinations of (parts of) such regulatory elements or several copies of a regulatory element, i.e. regulatory elements comprising non-naturally occurring sequences, are themselves also envisaged as regulatory element. Regulatory elements as used herein may comprise part of a larger sequence involved in transcriptional control, e.g. part of a promoter sequence. However, regulatory elements alone are typically not sufficient to initiate transcription, but require a promoter to this end. The regulatory elements disclosed herein are provided as nucleic acid molecules, i.e. isolated nucleic acids, or isolated nucleic acid molecules. Said nucleic acid regulatory element hence have a sequence which is only a small part of the naturally occurring genomic sequence and hence is not naturally occurring as such, but is isolated therefrom.

The term "nucleic acid" as used herein typically refers to an oligomer or polymer (preferably a linear polymer) of any length composed essentially of nucleotides. A nucleotide unit commonly includes a heterocyclic base, a sugar group, and at least one, e.g. one, two, or three, phosphate groups, including modified or substituted phosphate groups. Heterocyclic bases may include *inter alia* purine and pyrimidine bases such as adenine (A), guanine (G), cytosine (C), thymine (T) and uracil (U) which are widespread in naturally-occurring nucleic acids, other naturally-occurring bases (*e.g.*, xanthine, inosine, hypoxanthine) as well as chemically or biochemically modified (*e.g.*, methylated), non-natural or derivatised bases. Sugar groups may include *inter alia* pentose (pentofuranose) groups such as preferably ribose and/or 2-deoxyribose common in naturally-occurring nucleic acids, or arabinose, 2-deoxyarabinose, threose or hexose sugar groups, as well as modified or substituted sugar groups. Nucleic acids as intended herein may include naturally occurring nucleotides, modified nucleotides or mixtures thereof. A modified nucleotide may include a modified heterocyclic base, a modified sugar moiety, a modified phosphate group or a combination thereof. Modifications of phosphate groups or

sugars may be introduced to improve stability, resistance to enzymatic degradation, or some other useful property. The term "nucleic acid" further preferably encompasses DNA, RNA and DNA/RNA hybrid molecules, specifically including hnRNA, pre-mRNA, mRNA, cDNA, genomic DNA, amplification products, oligonucleotides, and synthetic (e.g., chemically synthesised) DNA, RNA or DNA/RNA hybrids. A nucleic acid can be naturally occurring, e.g., present in or isolated from nature; or can be non-naturally occurring, e.g., recombinant, i.e., produced by recombinant DNA technology, and/or partly or entirely, chemically or biochemically synthesised. A "nucleic acid" can be double-stranded, partly double stranded, or single-stranded. Where single-stranded, the nucleic acid can be the sense strand or the antisense strand. In addition, nucleic acid can be circular or linear.

As used herein "transcription factor binding site", "transcription factor binding sequence" or "TFBS" refers to a sequence of a nucleic acid region to which transcription factors bind. Non-limiting examples of TFBS include binding sites for E-box binding proteins E47 and E12, CCAAT/enhancer-binding protein, also known as CEB/P, C/EPB or CEPB; hepatocyte nuclear factor 1 homeobox A, also known as HNF-1A, HNF1, IDDM20, LFB1, MODY3, TCF-1 or TCF; interferon regulatory factor 1, also known as IRF1, IRF-1 or MAR; lymphoid enhancer-binding factor 1, also known as LEF1, LEF-1, TCF10, TCF1ALPHA or TCF7L3; forkhead box protein O4, also known as FOXO4, AFX, AFX1 or MLLT7; forkhead box protein O1, also known as forkhead in rhabdomyosarcoma, FOXO1, FKH1, FKHR or FOXO1A. Transcription factor binding sites may be found in databases such as Transfac®.

Sequences disclosed herein may be part of sequences of regulatory elements capable of controlling transcription of liver-specific genes *in vivo*, in particular controlling the following genes: albumin also known as ALB, apolipoprotein A2 also known as ApoA2, apolipoprotein C2 also known as ApoC2, apolipoprotein C1 also known as ApoC1, apolipoprotein A1 also known as ApoA1, apolipoprotein C3 also known as APOC3, APOCIII, transthyretin, also known as TTR, retinal binding protein 4, also known as RBP4, Cytochrome P450 2E1 also known as CYP2E1, alpha-1-acid glycoprotein 1 also known as orosomucoid 1 or ORM1.

Accordingly, in embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from ALB regulatory elements, i.e. regulatory elements that control expression of the ALB gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 1 and SEQ ID NO 2 or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from APOA2 regulatory elements, i.e. regulatory elements that control expression of the APOA2 gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 3 AND SEQ ID NO: 4 or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from TTR regulatory elements, i.e. regulatory elements that control expression of the TTR gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 5 or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from APOC1 regulatory elements, i.e. regulatory elements that control expression of the APOC1 gene *in vivo*, e.g.

regulatory elements comprising SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from RBP4 regulatory elements, i.e. regulatory elements that control expression of the RBP4 gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 12, SEQ ID NO: 13 or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from APOC3 regulatory elements, i.e. regulatory elements that control expression of the APOC3 gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from CYP2E1 regulatory elements, i.e. regulatory elements that control expression of the CYP2E1 gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 19 or functional fragments thereof as described herein elsewhere. In embodiments, the nucleic acid regulatory elements disclosed herein comprise a sequence from ORM1 regulatory elements, i.e. regulatory elements that control expression of the ORM1 gene *in vivo*, e.g. regulatory elements comprising SEQ ID NO: 20, SEQ ID NO: 21, or functional fragments thereof as described herein elsewhere.

The nucleic acid regulatory elements comprising SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16 or SEQ ID NO: 20, comprise the TFBS for SP1 and the TFBS for EP300.

The nucleic acid regulatory element comprising SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12 comprise the TFBS for SP1, the TFBS for EP300, the TFBS for HNF 4G, the TFBS for CEBPB, the TFBS for P300, the TFBS for HDAC2, the TFBS for JUND, the TFBS for FOSL2, the TFBS for ZBTB7A, the TFBS for CEBPD and the TFBS for RXRA.

The nucleic acid regulatory element comprising SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 13 comprise the TFBS for SP1, the TFBS for EP300, the TFBS for POLR2A, the TFBS for MYBL2, the TFBS for FOXA1, the TFBS for FOXA2, the TFBS for ARID3A, the TFBS for POLR2A and the TFBS for HEY1.

The nucleic acid regulatory element comprising SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 14, SEQ ID NO: 16 and SEQ ID NO: 20 comprise the TFBS for SP1, the TFBS for EP300, the TFBS for HNF 4G, the TFBS for CEBPB, the TFBS for P300 and the TFBS for HNF 4A.

As used herein, the terms “sequence identity” and “identical sequence” and the like refer to the degree in sequence identity or similarity between two polymeric molecules, e.g., between two nucleic acid molecules, e.g., two DNA molecules. Sequence alignments and determination of sequence identity can be done, e.g., using the Basic Local Alignment Search Tool (BLAST) originally described by Altschul et al. 1990 (J Mol Biol 215: 403-10), such as the "Blast 2 sequences" algorithm described by Tatusova and Madden 1999 (FEMS Microbiol Lett 174: 247-250). Typically, the percentage sequence identity is

calculated over the entire length of the sequence. As used herein, the term "substantially identical" denotes at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98% or 99%, sequence identity.

The term "functional fragment" as used in the application refers to fragments of the regulatory element sequences disclosed herein that retain the capability of regulating liver-specific expression, i.e. they can still confer tissue specificity and they are capable of regulating expression of a (trans)gene in the same way (although possibly not to the same extent) as the sequence from which they are derived. Functional fragments may preferably comprise at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120, at least 150, at least 200, at least 250, at least 300, at least 350, or at least 400 contiguous nucleotides from the sequence from which they are derived. Also preferably, functional fragments may comprise at least 1, more preferably at least 2, at least 3, or at least 4, even more preferably at least 5 or all of the transcription factor binding sites (TFBS) that are present in the sequence from which they are derived.

The term "liver-specific expression" as used in the application, refers to the preferential or predominant expression of a (trans)gene (as RNA and/or polypeptide) in the liver, in liver tissue or in liver cells, as compared to other (i.e. non-liver) tissues or cells. According to particular embodiments, at least 50%, more particularly at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99% or 100% of the (trans)gene expression occurs within liver tissue or liver cells. According to a particular embodiment, liver-specific expression entails that there is no 'leakage' of expressed gene product to other organs or tissue than liver, such as lung, muscle, brain, kidney and/or spleen.

The same applies mutatis mutandis for hepatocyte-specific expression and hepatoblast-specific expression, which may be considered as particular forms of liver-specific expression. Throughout the application, where liver-specific is mentioned in the context of expression, hepatocyte-specific expression and hepatoblast-specific expression are also explicitly envisaged.

As used herein, the term "liver cells" encompasses the cells predominantly populating the liver and encompasses mainly hepatocytes, oval cells, liver sinusoidal endothelial cells (LSEC) and cholangiocytes (epithelial cells forming the bile ducts).

The term "hepatocyte," as used herein, refers to a cell that has been differentiated from a progenitor hepatoblast such that it is capable of expressing liver-specific phenotype under appropriate conditions. The term "hepatocyte" also refers to hepatocytes that are de-differentiated. The term includes cells *in vivo* and cells cultured *ex vivo* regardless of whether such cells are primary or passaged.

The term "hepatoblast" as used herein, refers to an embryonic cell in the mesoderm that differentiates to give rise to a hepatocyte, an oval cell, or a cholangiocyte. The term includes cells *in vivo* and cells cultured *ex vivo* regardless of whether such cells are primary or passaged.

In embodiments, the invention relates to a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a functional fragment of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences. Said "functional fragments" of the regulatory elements identified herein are defined as comprising at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, and/or as comprising at least 1, preferably at least 5, more preferably all of the transcription factor binding sites (TFBS) that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 20, a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof. Said "functional fragments" of the regulatory elements identified herein are defined as comprising the TFBS for SP1 and the TFBS for EP300, and/or comprising at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, and/or as comprising at least 1, preferably at least 5, more preferably all of the TFBS that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 13, a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof. Said "functional fragments" of the regulatory elements identified herein are defined as comprising at least the TFBS for SP1 and the TFBS for EP300, and/or as comprising at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, or and/or as comprising at least 1, preferably all of the TFBS that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of, a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12 or a sequence having at least

80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof, wherein said nucleic acid regulatory element comprises the TFBS for SP1, the TFBS for EP300, the TFBS for HNF 4G, the TFBS for CEBPB, the TFBS for P300, the TFBS for HDAC2, the TFBS for JUND, the TFBS for FOSL2, the TFBS for ZBTB7A, the TFBS for CEBPD and the TFBS for RXRA; and wherein said functional fragment comprises at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, and wherein said functional fragment comprises at least 1, preferably at least 5, more preferably at least 10 or at least 15, of the TFBS that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of, a sequence selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 13, or a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof, wherein said nucleic acid regulatory element comprises the TFBS for SP1, the TFBS for EP300, the TFBS for POLR2A, the TFBS for MYBL2, the TFBS for FOXA1, the TFBS for FOXA2, the TFBS for ARID3A, the TFBS for POLR2A and the TFBS for HEY1; and wherein said functional fragment comprises at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, and wherein said functional fragment comprises at least 1, preferably at least 5, more preferably at least 10 or at least 15, of the TFBS that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of, a sequence selected from the group consisting of: SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 14, SEQ ID NO: 16 and SEQ ID NO: 20, or a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof, wherein said nucleic acid regulatory element comprises the TFBS for SP1, the TFBS for EP300, the TFBS for HNF 4G, the TFBS for CEBPB, the TFBS for P300 and the TFBS for HNF 4A; and wherein said functional fragment comprises at least 20, preferably at least 25, more preferably at least 50, at least 100, at least 200 or at least 250, contiguous nucleotides from the sequence from which it is derived, and wherein said functional fragment comprises at least 1, preferably at least 5, more preferably at least 10 or at least 15, of the TFBS that are present in the sequence from which it is derived.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 6, or a sequence having at least 80%, preferably at least 85%,

more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to SEQ ID NO: 6, or a functional fragment thereof as described herein.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 4, or a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to SEQ ID NO: 4, or a functional fragment thereof as described herein.

In further embodiments, the invention provides a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 12, or a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to SEQ ID NO: 12, or a functional fragment thereof as described herein.

It is also possible to make nucleic acid regulatory elements that comprise an artificial sequence by combining two or more (e.g. two, three, four, five or more) identical or different sequences disclosed herein or functional fragments thereof as described herein.

Accordingly, in certain embodiments a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising at least two sequences selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof, preferably selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 13, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein, more preferably selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein elsewhere.

For example, disclosed herein is a nucleic acid regulatory element comprising, consisting essentially of, or consisting of SEQ ID NO:6, SEQ ID NO: 4 and SEQ ID NO: 12; a nucleic acid regulatory element comprising, consisting essentially of, or consisting of 2, 3, 4, or 5 repeats, e.g. tandem repeats, of SEQ ID NO:6; a nucleic acid regulatory element comprising, consisting essentially of, or consisting of 2, 3, 4, or 5 repeats, e.g. tandem repeats, of SEQ ID NO: 4; or a nucleic acid regulatory element comprising, consisting essentially of, or consisting of 2, 3, 4, or 5 repeats, e.g. tandem repeats, of SEQ ID NO: 12. It is also possible to make nucleic acid regulatory elements that comprise an artificial sequence by

combining one or more (e.g. one, two, three, four, five or more) sequences disclosed herein or functional fragments thereof as described herein with one or more liver-specific regulatory elements which are known in the art. Non-limiting examples of known liver-specific regulatory elements are disclosed in WO 2009/130208, WO 2016/146757 or WO 01/98482 which are incorporated by reference in their entirety herein.

A particularly preferred example of a liver-specific regulatory element for combining with a nucleic acid regulatory element disclosed herein is a nucleic acid regulatory element derived from the SERPINA1 gene promoter, more particularly a regulatory element comprising the sequence as defined in SEQ ID NO: 22, a sequence having at least 85%, preferably at least 90%, more preferably at least 95%, such as 96%, 97%, 98% or 99%, identity to said sequence, or a functional fragment thereof as described herein. The liver-specific regulatory element consisting of the sequence as defined in SEQ ID NO: 22 is herein referred to as “Serpin enhancer”, “SerpEnh”, or “Serp”.

Accordingly, in certain embodiments a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising the sequence SEQ ID NO: 22, a sequence having at least 85%, preferably at least 90%, more preferably at least 95%, such as 96%, 97%, 98% or 99%, identity to said sequence, or a functional fragment thereof as described herein, and at least one or two or more (e.g. two, three, four, five or more) identical or different sequences selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; preferably selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 20, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; more preferably selected from the group SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 13, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; more preferably SEQ ID NO: 6 or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; even more preferably selected from the group SEQ ID NO: 6, SEQ ID NO: 4 and SEQ ID NO: 12, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; more preferably SEQ ID NO: 6 or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein.

In a particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 22, more particularly 3 repeats, e.g. tandem repeats, of SEQ ID NO:22 (Serp), and SEQ ID NO: 6. In a more particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 23 (3xSerp) and SEQ ID NO: 6.

In a particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 22, more particularly 3 repeats, e.g. tandem repeats, of SEQ ID NO: 22 (Serp), and SEQ ID NO: 4. In a more particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 23 (3xSerp) and SEQ ID NO: 4.

In a particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 22, more particularly 3 repeats, e.g. tandem repeats, of SEQ ID NO: 22 (Serp), and SEQ ID NO: 12. In a more particular embodiment, a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising, consisting essentially of, or consisting of SEQ ID NO: 23 (3xSerp) and SEQ ID NO: 12.

A further particularly preferred example of a liver-specific regulatory element for combining with a nucleic acid regulatory element disclosed herein is the “TTR enhancer”, “TTRe” or “TTREnh”, more particularly a regulatory element comprising the sequence as defined in SEQ ID NO: 24, a sequence having at least 85%, preferably at least 90%, more preferably at least 95%, such as 96%, 97%, 98% or 99%, identity to said sequence, or a functional fragment thereof as described herein. Accordingly, in certain embodiments a nucleic acid regulatory element for enhancing liver-specific gene expression is provided comprising the sequence SEQ ID NO: 24, a sequence having at least 85%, preferably at least 90%, more preferably at least 95%, such as 96%, 97%, 98% or 99%, identity to said sequence, or a functional fragment thereof as described herein, and at least one or two or more (e.g. two, three, four, five or more) identical or different sequences selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; preferably selected from the group consisting of SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 20, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as

described herein; more preferably selected from the group SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 13, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; more preferably SEQ ID NO: 6 or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; even more preferably selected from the group SEQ ID NO: 6, SEQ ID NO: 4 and SEQ ID NO: 12, or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein; more preferably SEQ ID NO: 6 or a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein.

A further particularly preferred example of a liver-specific regulatory element for combining with a nucleic acid regulatory element disclosed herein is a regulatory element comprising the "hepatic locus control element" as described in WO 01/98482 (which is hereby incorporated by reference herein), more particularly a regulatory element comprising the sequence as defined in SEQ ID NO: 57 or SEQ ID NO: 58, a sequence having at least 85%, preferably at least 90%, more preferably at least 95%, such as 96%, 97%, 98% or 99%, identity to said sequence, or a functional fragment thereof as described herein.

A further particularly preferred example of a liver-specific regulatory element for combining with a nucleic acid regulatory element disclosed herein is a regulatory element comprising at least one, preferably three tandem repeats of the Serpin enhancer (e.g. SEQ ID NO: 22), such as a regulatory element comprising SEQ ID NO: 23 (3xSerp) and the transthyretin enhancer (TTRe) as defined by SEQ ID NO: 24. Examples of liver-specific regulatory elements comprising at least one, preferably three tandem repeats of the Serpin enhancer and the TTRe are described in WO 2016/146757 which is incorporated by reference in its entirety herein.

The one or more liver-specific regulatory elements which are known in the art may be located upstream or downstream of the one or more (e.g. one, two, three, four, five or more) nucleic acid regulatory elements disclosed herein Preferably, the one or more (e.g. one, two, three, four, five or more) nucleic acid regulatory elements disclosed herein or functional fragments thereof as described herein are located upstream of the one or more liver-specific regulatory elements which are known in the art.

Particular examples of nucleic acid regulatory elements that comprise an artificial sequence include the regulatory elements that are obtained by rearranging the transcription factor binding sites (TFBS) that are present in the sequences disclosed herein. Said rearrangement may encompass changing the order of the TFBSs and/or changing the position of one or more TFBSs relative to the other TFBSs and/or changing the copy number of one or more of the TFBSs. For example, also disclosed herein is a nucleic acid regulatory element for enhancing liver-specific gene expression comprising binding sites for ATF3, AR1D3A, 2BTB7A, BHLHE40, CEBPB, CEBPD, CHD2, CTCF, EP300, ELF1, EZH2, ESRRA,

FOXA1, FOXA2, FOSL2, HEY1, GATA2, GABPA, HEY1, HDAC2, HNF 4A, HNF 4G, HSF1, JUN, JUND, MAFF, MAFK, MAX, MYBL2, MBD4, MAZ, MXI1, MYC, NF1C, NR2C2, NR2F2, POLR2A, PO12, p300 , PPARGC1A, RAD21, RXR5, RXRA, RCOR1, REST, SMC3, SUZ12, SHC3, SP1, SRF, SREBP1, SIN3AK20, TAF1, TCF7L2, TBP, TEAD4, TBP, TCF12, TCF4, USF1, USF2, YY1, ZBTB7A and/or ZBTB33. In further examples, these nucleic acid regulatory elements comprise at least two, such as 2, 3, 4, or more copies of one or more of the recited TFBSs.

In case the regulatory element is provided as a single stranded nucleic acid, e.g. when using a single-stranded AAV vector, the complement strand is considered equivalent to the disclosed sequences. Hence, also disclosed herein is a nucleic acid regulatory element for enhancing liver-specific gene expression comprising, consisting essentially of, or consisting of the complement of a sequence described herein, in particular a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein.

Also disclosed herein is a nucleic acid regulatory element for enhancing liver-specific gene expression hybridizing under stringent conditions to a nucleic acid regulatory element described herein, in particular to the nucleic acid regulatory element comprising, consisting essentially of, or consisting of a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21 , a sequence having at least 90%, preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, a functional fragment thereof as described herein, or to its complement. Said nucleic acid regulatory elements do not need to be of equal length as the sequence they hybridize to. In preferred embodiments, the size of said hybridizing nucleic acid regulatory element does not differ more than 25% in length, in particular 20% in length, more in particular 15% in length, most in particular 10% in length from the sequence it hybridizes to.

The expression “hybridize under stringent conditions”, refers to the ability of a nucleic acid molecule to hybridize to a target nucleic acid molecule under defined conditions of temperature and salt concentration. Typically, stringent hybridization conditions are no more than 25°C to 30°C (for example, 20°C, 15°C, 10°C or 5°C) below the melting temperature (Tm) of the native duplex. Methods of calculating Tm are well known in the art. By way of non-limiting example, representative salt and temperature conditions for achieving stringent hybridization are: 1x SSC, 0.5% SDS at 65°C. The abbreviation SSC refers to a buffer used in nucleic acid hybridization solutions. One liter of the 20x

(twenty times concentrate) stock SSC buffer solution (pH 7.0) contains 175.3 g sodium chloride and 88.2 g sodium citrate. A representative time period for achieving hybridization is 12 hours.

Preferably the regulatory elements as described herein are fully functional while being only of limited length. This allows their use in vectors or nucleic acid expression cassettes without unduly restricting their payload capacity. Accordingly, in embodiments, the regulatory element disclosed herein is a nucleic acid of 600 nucleotides or less, 500 nucleotides or less, 400 nucleotides or less, 300 nucleotides or less, 250 nucleotides or less, 200 nucleotides or less, 190 nucleotides or less, or 180 nucleotides or less, preferably 500 nucleotides or less, more preferably 300 nucleotides or less.

However, it is to be understood that the disclosed nucleic acid regulatory elements retain regulatory activity (i.e. with regard to specificity and/or activity of transcription) and thus they particularly have a minimum length of 20 nucleotides, 25 nucleotides, 30 nucleotides, 35 nucleotides, 40 nucleotides, 45 nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, 160 nucleotides, 170 nucleotides, or all nucleotides of the respective sequences of the regulatory elements define by SEQ ID NO: 1 to 21.

In certain embodiments, the invention provides for a nucleic acid regulatory element of 600 nucleotides or less, 500 nucleotides or less, 400 nucleotides or less, 300 nucleotides or less, 250 nucleotides or less, 200 nucleotides or less, 190 nucleotides or less, or 180 nucleotides or less, preferably 500 nucleotides or less, more preferably 300 nucleotides or less, for enhancing liver-specific gene expression comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21a sequence having at least 80%, preferably at least 85%, more preferably at least 90%, even more preferably at least 95%, such as 95%, 96%, 97%, 98%, or 99%, identity to any of these sequences, or a functional fragment thereof as described herein. The nucleic acid regulatory elements disclosed herein may be used in a nucleic acid expression cassette. Accordingly, in an aspect the invention provides for the use of the nucleic acid regulatory elements as described herein in a nucleic acid expression cassette.

In an aspect the invention provides a nucleic acid expression cassette comprising a nucleic acid regulatory element as described herein, operably linked to a promoter. In embodiments, the nucleic acid expression cassette does not contain a transgene. Such nucleic acid expression cassette may be used to drive expression of an endogenous gene. In preferred embodiments, the nucleic acid expression cassette comprises a nucleic acid regulatory element as described herein, operably linked to a promoter and a transgene.

As used herein, the term "nucleic acid expression cassette" refers to nucleic acid molecules that include one or more transcriptional control elements (such as, but not limited to promoters, enhancers and/or regulatory elements, polyadenylation sequences, and introns) that direct (trans)gene expression in one or more desired cell types, tissues or organs. Typically, they will also contain a transgene, although it is

also envisaged that a nucleic acid expression cassette directs expression of an endogenous gene in a cell into which the nucleic acid cassette is inserted.

The term "operably linked" as used herein refers to the arrangement of various nucleic acid molecule elements relative to each such that the elements are functionally connected and are able to interact with each other. Such elements may include, without limitation, a promoter, an enhancer and/or a regulatory element, a polyadenylation sequence, one or more introns and/or exons, and a coding sequence of a gene of interest to be expressed (i.e., the transgene). The nucleic acid sequence elements, when properly oriented or operably linked, act together to modulate the activity of one another, and ultimately may affect the level of expression of the transgene. By modulate is meant increasing, decreasing, or maintaining the level of activity of a particular element. The position of each element relative to other elements may be expressed in terms of the 5' terminus and the 3' terminus of each element, and the distance between any particular elements may be referenced by the number of intervening nucleotides, or base pairs, between the elements. As understood by the skilled person, operably linked implies functional activity, and is not necessarily related to a natural positional link. Indeed, when used in nucleic acid expression cassettes, the regulatory elements will typically be located immediately upstream of the promoter (although this is generally the case, it should definitely not be interpreted as a limitation or exclusion of positions within the nucleic acid expression cassette), but this needs not be the case *in vivo*. E.g., a regulatory element sequence naturally occurring downstream of a gene whose transcription it affects is able to function in the same way when located upstream of the promoter. Hence, according to a specific embodiment, the regulatory or enhancing effect of the regulatory element is position-independent.

In particular embodiments, the nucleic acid expression cassette comprises one nucleic acid regulatory element as described herein or a functional fragment thereof as described herein. In alternative embodiments, the nucleic acid expression cassette comprises two or more, such as, e.g., 2, 3, 4, 5, 6, 7, 8, 9, or 10, nucleic acid regulatory elements as described herein, or functional fragments thereof as described herein, i.e. they are combined modularly to enhance their regulatory (and/or enhancing) effect. In yet alternative embodiments, the nucleic acid expression cassette comprises at least one nucleic acid regulatory element as described herein, or functional fragments thereof as described herein; and at least one liver-specific nucleic acid regulatory element known in the art, or functional fragments thereof as described herein. In further embodiments, at least two of the two or more nucleic acid regulatory elements, or functional fragments thereof as described herein, are identical or substantially identical. In yet further embodiments, all of the two or more regulatory elements, or functional fragments thereof as described herein, are identical or substantially identical. The copies of the identical or substantially identical nucleic acid regulatory elements, or functional fragments thereof as described herein, may be provided as tandem repeats in the nucleic acid expression cassette. In alternative further embodiments, at least two of the two or more nucleic acid regulatory elements, or functional fragments thereof as described herein, are different from each other. The nucleic acid expression cassette may also

comprise a combination of identical and substantially identical nucleic acid regulatory elements, or functional fragments thereof as described herein, and non-identical nucleic acid regulatory elements, or functional fragments thereof as described herein.

For example, the nucleic acid expression cassette may comprise a nucleic acid regulatory element comprising SEQ ID NO: 6, and a nucleic acid regulatory element comprising SEQ ID NO: 22 (SERP); the nucleic acid expression cassette may comprise a nucleic acid regulatory element comprising SEQ ID NO: 6 and 2, 3, or 4 nucleic acid regulatory elements comprising SEQ ID NO: 22 (SERP); the nucleic acid expression cassette may comprise 2, 3, or 4 nucleic acid regulatory element comprising SEQ ID NO: 6 and a nucleic acid regulatory elements comprising SEQ ID NO: 22 (SERP); the nucleic acid expression cassette may comprise 2, 3, or 4 nucleic acid regulatory element comprising SEQ ID NO: 6 and 2, 3, or 4 nucleic acid regulatory elements comprising SEQ ID NO: 22 (SERP). Similary, the nucleic acid expression cassette may comprise 2, 3, or 4 nucleic acid regulatory element comprising SEQ ID NO: 6 and a nucleic acid regulatory elements comprising SEQ ID NO: 23 (3xSERP).

As used in the application, the term "promoter" refers to nucleic acid sequences that regulate, either directly or indirectly, the transcription of corresponding nucleic acid coding sequences to which they are operably linked (e.g. a transgene or endogenous gene). A promoter may function alone to regulate transcription or may act in concert with one or more other regulatory sequences (e.g. enhancers or silencers, or regulatory elements). In the context of the present application, a promoter is typically operably linked to a regulatory element as disclosed herein to regulate transcription of a (trans)gene. When a regulatory element as described herein is operably linked to both a promoter and a transgene, the regulatory element can (1) confer a significant degree of liver-specific expression *in vivo* (and/or in hepatoblasts, hepatocytes, oval cells, cholangiocytes, liver stem cells or liver-derived cell lines *in vitro*) of the transgene and/or (2) can increase the level of expression of the transgene in the liver and/or in hepatoblasts, hepatocytes, oval cells, liver stem cells and cholangiocytes (*in vivo)* or in cell lines derived therefrom (*in vitro)*.

The promoter may be homologous (i.e. from the same species as the animal, in particular mammal, to be transfected with the nucleic acid expression cassette) or heterologous (i.e. from a source other than the species of the animal, in particular mammal, to be transfected with the expression cassette). As such, the source of the promoter may be any virus, any unicellular prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, or may even be a synthetic promoter (i.e. having a non-naturally occurring sequence), provided that the promoter is functional in combination with the regulatory elements described herein. In preferred embodiments, the promoter is a mammalian promoter, in particular a murine or human promoter.

The promoter may be an inducible or constitutive promoter.

The enrichment in liver-specific TFBS in the nucleic acid regulatory elements disclosed herein in principle allows the regulatory elements to direct liver-specific expression even from a promoter that

itself is not liver-specific. Hence, the regulatory elements disclosed herein can be used in nucleic acid expression cassettes in conjunction with their natural promoter, as well as with another promoter. Preferably, the nucleic acid expression cassettes disclosed herein comprise a liver-specific promoter. This is to increase liver specificity and/or avoid leakage of expression in other tissues. Non-limiting examples of liver-specific promoters are provided on the Liver Specific Gene Promoter Database (LSPD, http://rulai.cshl.edu/LSPD/), and include, for example, the albumin (ALB) promoter, the apolipoprotein A2 (APOA2) promoter, the transthyretin (TTR) promoter, the apolipoprotein C1 (APOC1) promoter, the retinol binding protein 4 (RBP4) promoter, the apolipoprotein C3 (APOC3) promoter, the cytochrome P450 2E1 (CYP2E1) promoter, the alpha-1-acid glycoprotein 1 (ORM1) promoter, the serpin A1 (SERPINA1) promoter, the apolipoprotein A1 (APOA1) promoter, the complement factor B (CFB) promoter, the ketohexokinase (KHK) promoter, the hemopexin (HPX) promoter, the nicotinamide N-methyltransferase (NNMT) promoter, the (liver) carboxylesterase 1 (CES1) promoter, the protein C (PROC) promoter, the apolipoprotein C3 (APOC3) promoter, the mannan-binding lectin serine protease 2 (MASP2) promoter, the hepcidin antimicrobial peptide (HAMP) promoter, the serpin peptidase inhibitor, clade C (antithrombin), member 1 (SERPINC1) promoter, the alpha 1-antitrypsin (AAT) promoter.

In particularly preferred embodiments, the promoter is a mammalian liver-specific promoter, in particular a murine or human liver-specific promoter.

In preferred embodiments, the promoter is from the transthyretin gene, in particular the murine or human transthyretin gene, such as the minimal transthyretin promoter (TTRm) as defined in SEQ ID NO: 27.

In particular embodiments, if the nucleic acid regulatory element for enhancing liver-specific gene expression as taught herein is combined with the TTRe enhancer (e.g. as defined in SEQ ID NO: 24) the promotor is preferably TTRm (e.g. as defined in SEQ ID NO: 27), and the combination of the TTRe and TTRm nucleic acid modules is preferably as set forth in SEQ ID NO: 28. Accordingly, in particular embodiments, the promoter is the transthyretin (TTR) promoter, thereby comprising the combination of the TTRe and TTRm nucleic acids as defined by SEQ ID NO: 28.

In particular embodiments, if the nucleic acid regulatory element for enhancing liver-specific gene expression as taught herein is combined with the “hepatic locus control element” as described in WO 01/98482 hereby incorporated by reference, more particularly a regulatory element comprising the sequence as defined in SEQ ID NO: 57 or SEQ ID NO: 58, the promotor is preferably the AAT promoter.

Furthermore, the promoter does not need to be the promoter of the transgene in the nucleic acid expression cassette, although it is possible that the transgene is transcribed from its own promoter.

To minimize the length of the nucleic acid expression cassette, the regulatory elements may be linked to minimal promoters, or shortened versions of the promoters described herein. A "minimal promoter”

(also referred to as basal promoter or core promoter) as used herein is part of a full-size promoter still capable of driving expression, but lacking at least part of the sequence that contributes to regulating (e.g. tissue-specific) expression. This definition covers both promoters from which (tissue-specific) regulatory elements have been deleted- that are capable of driving expression of a gene but have lost their ability to express that gene in a tissue-specific fashion and promoters from which (tissue-specific) regulatory elements have been deleted that are capable of driving (possibly decreased) expression of a gene but have not necessarily lost their ability to express that gene in a tissue-specific fashion. Preferably, the promoter contained in the nucleic acid expression cassette disclosed herein is 1000 nucleotides or less in length, 900 nucleotides or less, 800 nucleotides or less, 700 nucleotides or less, 600 nucleotides or less, 500 nucleotides or less, 400 nucleotides or less, 300 nucleotides or less, or 250 nucleotides or less.

The term “transgene” as used herein refers to particular nucleic acid sequences encoding a polypeptide or a portion of a polypeptide to be expressed in a cell into which the nucleic acid sequence is introduced. However, it is also possible that transgenes are expressed as RNA, typically to control (e.g. lower) the amount of a particular polypeptide in a cell into which the nucleic acid sequence is inserted. These RNA molecules include but are not limited to gRNA, shRNA, circRNA, siRNA, antisense-RNA, and the like. etc. How the nucleic acid sequence is introduced into a cell is not essential to the invention, it may for instance be through integration in the genome or as an episomal plasmid. Of note, expression of the transgene may be restricted to a subset of the cells into which the nucleic acid sequence is introduced. The term “transgene” is meant to include (1) a nucleic acid sequence that is not naturally found in the cell (i.e., a heterologous nucleic acid sequence); (2) a nucleic acid sequence that is a mutant form of a nucleic acid sequence naturally found in the cell into which it has been introduced; (3) a nucleic acid sequence that serves to add additional copies of the same (i.e., homologous) or a similar nucleic acid sequence naturally occurring in the cell into which it has been introduced ; or (4) a silent naturally occurring or homologous nucleic acid sequence whose expression is induced in the cell into which it has been introduced.

The transgene may be homologous or heterologous to the promoter (and/or to the animal, in particular mammal, in which it is introduced, e.g. in cases where the nucleic acid expression cassette is used for gene therapy).

The transgene may be codon-optimized.

The transgene may be a full length cDNA or genomic DNA sequence, or any fragment, subunit or mutant thereof that has at least some biological activity. In particular, the transgene may be a minigene, i.e. a gene sequence lacking part, most or all of its intronic sequences. The transgene thus optionally may contain intron sequences. Optionally, the transgene may be a hybrid nucleic acid sequence, i.e., one constructed from homologous and/or heterologous cDNA and/or genomic DNA fragments. By 'mutant form' is meant a nucleic acid sequence that contains one or more nucleotides that are different from the

wild-type or naturally occurring sequence, i.e., the mutant nucleic acid sequence contains one or more nucleotide substitutions, deletions, and/or insertions. The nucleotide substitution, deletion, and/or insertion can give rise to a gene product (i.e. e., protein or nucleic acid) that is different in its amino acid/nucleic acid sequence from the wild type amino acid/nucleic acid sequence. Preparation of such mutants is well known in the art. In some cases, the transgene may also include a sequence encoding a leader peptide or signal sequence such that the transgene product will be secreted from the cell.

The transgene that may be contained in the nucleic acid expression cassettes described herein typically encodes a gene product such as RNA or a polypeptide (protein).

In embodiments, the transgene encodes a therapeutic protein or an immunogenic protein.

In embodiments, the transgene encodes a therapeutic protein. Non-limiting examples of therapeutic proteins include factor VIII, factor IX, factor VII, factor VIIa, factor X, von Willebrand factor, phenylalaninehydroxylase (PAH), α-glucosidase (GAA), C1 esterase inhibitor (C1-INH), lysosomal enzymes, lysosomal enzyme iduronate-2-sulfatase (I2S), erythropoietin (EPO), interferon-α, interferon-β, interferon-γ, interleukin 1 (IL-1), interleukin 2 (IL-2), interleukin 3 (IL-3), interleukin 4 (IL-4), interleukin 5 (IL-5), interleukin 6 (IL-6), interleukin 7 (IL-7), interleukin 8 (IL-8), interleukin 9 (IL-9), interleukin 10 (IL-10), interleukin 11 (IL-11), interleukin 12 (IL-12), chemokine (C-X-C motif) ligand 5 (CXCL5), granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor (SCF), keratinocyte growth factor (KGF), monocyte chemoattractant protein-1 (MCP-1), tumor necrosis factor (TNF), afamin (AFM), α1-antitrypsin, α-galactosidase A, α-L-iduronidase, ATP7b, ornithine transcarbamoylase, phenylalanine hydroxylase, lipoprotein lipase, apoliproteins, low-density lipoprotein receptor (LDL-R), albumin, glucose-6-phosphatase, VP22 fusion protein-based dominant negative mutant, ISG56,IFITM1, interferon-inducible MyD88 protein transgenes encoding antibodies, nanobodies, anti-viral dominant-negative proteins, and fragments, subunits or mutants thereof, etc.

In embodiments, the transgene encodes an immunogenic protein. Non-limiting examples of immunogenic proteins include epitopes and antigens derived from a pathogen such as hepatitis A, B, or C surface antigen proteins.

As used herein, the term “immunogenic” refers to a substance or composition capable of eliciting an immune response.

In certain embodiments, the transgene encodes FIX. The term “coagulation factor IX” has the meaning as known in the art. Synonyms of coagulation factor IX are “FIX” or “Christmas factor” or “F9” and can be used interchangeably. In particular, the term “coagulation factor IX” encompasses the human protein encoded by the mRNA sequence as defined in Genbank accession number NM_000133. Preferably, said FIX is a mutated FIX, which is hyperactive or hyper-functional as compared to the wild type FIX. Modifying functional activity of human coagulation factor can be done by bioengineering e.g. by introduction of point mutations. By this approach a hyperactive R338A variant was reported, which

showed a 3-fold increased clotting activity compared to the wild-type human FIX in an *in vitro* activated partial thromboplastin time assay (APPT) (Chang et al., 1998) and a 2 to 6-fold higher specific activity in hemophilia B mice transduced with the mutant FIX gene (Schuettrumpf et al., 2005). Further exemplary FIX point-mutants or domain exchange mutants with even higher clotting activities have been described: FIX, with the EGF-1 domain replaced with the EGF-1 domain from FVII, alone or in combination with a R338A point mutation (Brunetti-Pierri et al., 2009); the V86A/E277A/R338A triple mutant (Lin et al., 2010); the Y259F, K265T, and/or Y345T single, double or triple mutants (Milanov, et al., 2012); and the G190V point mutant (Kao et al., 2010); all incorporated herein by reference. Further exemplary FIX variants include Factor IX dalcinonacog alfa ("CB 2679d/ISU304") (Catalyst Biosciences Inc.).

In a particularly preferred embodiment, the FIX mutant is the one described by Simioni et al., in 2009 and denominated as the "factor IX Padua" mutant, causing X-linked thrombophilia. Said mutant factor IX is hyperactive and carries an R338L amino acid substitution. In an embodiment of the present invention, the FIX transgene encodes the human FIX protein, preferably the FIX transgene encodes for the Padua mutant of the human FIX protein. In an embodiment, the FIX transgene encodes for the Padua mutant of the human FIX protein with a sequence as set forth in SEQ ID NO: 25 (co-FIX-R338L).

In certain embodiments, the transgene encodes FVIII. The term "coagulation factor VIII" has the meaning as known in the art. Synonyms of coagulation factor VIII are "FVIII" or "anti-hemophilic factor" or "AHF" and can be used interchangeably herein. The term "coagulation factor VIII" encompasses, for example, the human protein having the amino acid sequence as defined in Uniprot accession number P00451. Preferably, said FVIII is a FVIII wherein the B domain is deleted (i.e. B domain deleted FVIII, also referred to as BDD FVIII or FVIIIΔB herein). The term "B domain deleted FVIII" encompasses for example, but without limitation, FVIII mutants wherein whole or a part of the B domain is deleted and FVIII mutants wherein the B domain is replaced by a linker. Non-limiting examples of B domain deleted FVIII are described in Ward et al. (2011), WO 2011/005968 and McIntosh et al. (2013), which are specifically incorporated by reference herein. In certain embodiments, said FVIII is B domain deleted FVIII wherein the B domain is replaced by a linker having the following sequence: SFSQNPPVLTRHQR (SEQ ID NO: 54) (i.e. SQ FVIII as defined in Ward et al. (2011)). In certain embodiments, said FVIII has SEQ ID NO: 26 (i.e. codon-optimized B domain deleted human FVIII or hFVIIIcopt), as disclosed also in WO 2011/0059, which is specifically incorporated by reference herein.

Other sequences may be incorporated in the nucleic acid expression cassette disclosed herein as well, typically to further increase or stabilize the expression of the transgene product (e.g. introns and/or polyadenylation sequences).

Any intron can be utilized in the expression cassettes described herein. The term "intron" encompasses any portion of a whole intron that is large enough to be recognized and spliced by the nuclear splicing apparatus. Typically, short, functional, intron sequences are preferred in order to keep the size of the expression cassette as small as possible which facilitates the construction and manipulation of the expression cassette. In some embodiments, the intron is obtained from a gene that encodes the protein that is encoded by the coding sequence within the expression cassette. The intron can be located 5' to the coding sequence, 3' to the coding sequence, or within the coding sequence. An advantage of locating the intron 5' to the coding sequence is to minimize the chance of the intron interfering with the function of the polyadenylation signal. In embodiments, the nucleic acid expression cassette disclosed herein further comprises an intron. Non-limiting examples of suitable introns are Minute Virus of Mice (MVM) intron, beta-globin intron (betaIVS-II), factor IX (FIX) intron A, Simian virus 40 (SV40) small-t intron, and beta-actin intron. Preferably, the intron is a minute virus of mouse (MVM) intron, more preferably the MVM intron as defined by SEQ ID NO: 29.

Any polyadenylation signal that directs the synthesis of a polyA tail is useful in the expression cassettes described herein, examples of those are well known to one of skill in the art. Exemplary polyadenylation signals include, but are not limited to, polyA sequences derived from the Simian virus 40 (SV40) late gene, the bovine growth hormone (BGH) polyadenylation signal, the minimal rabbit β-globin (mRBG) gene, and the synthetic polyA s(SPA) site as described in Levitt et al. (1989, Genes Dev 3:1019-1025). Preferably, the polyadenylation signal is the bovine growth hormone (BGH) polyadenylation signal (SEQ ID NO: 30).

In a particular embodiment, a nucleic acid expression cassette is disclosed and comprises:

- a liver-specific nucleic acid regulatory element comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof as described herein;

- the transthyretin enhancer (TTRe) as defined by SEQ ID NO: 24;,

- the liver-specific TTRm promoter (e.g. defined by SEQ ID NO: 27), and

- a transgene,

preferably wherein the combination of the TTRe and TTRm nucleic acids is defined by SEQ ID NO: 28.

In a particular embodiment, a nucleic acid expression cassette is disclosed and comprises:

- a liver-specific nucleic acid regulatory element comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO:

11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof as described herein;

- the transthyretin enhancer (TTRe) as defined by SEQ ID NO: 24,

- the liver-specific TTRm promoter (e.g. defined by SEQ ID NO: 27), and

- an intron, preferably the MVM intron (e.g. as defined by SEQ ID NO: 29), and

- a transgene, preferably wherein said transgene encodes for coagulation factor IX (FIX), preferably wherein said transgene is codon-optimized coagulation factor FIX, or wherein said coagulation factor FIX contains a hyper-activating mutation, preferably wherein said hyper-activating mutation corresponds to an R338L amino acid substitution, more preferably wherein said transgene encoding for coagulation factor FIX has a nucleic acid sequence as defined by SEQ ID NO: 25; or wherein said transgene encodes for coagulation factor VIII (FVIII), preferably wherein said transgene is codon-optimized coagulation factor FVIII, or wherein said coagulation factor VIII has a deletion of the B domain, preferably wherein said B domain of said FVIII is replaced by a linker defined by SEQ ID NO: 54, more preferably wherein said transgene encodes for coagulation factor VIII has a nucleic acid sequence defined by SEQ ID NO: 26;

preferably wherein the combination of the TTRe and TTRm nucleic acids is defined by SEQ ID NO: 28.

The nucleic acid regulatory element and the nucleic acid expression cassette disclosed herein may be used as such, or typically, they may be part of a nucleic acid vector. Accordingly, a further aspect relates to the use of a nucleic acid regulatory element as described herein or a nucleic acid expression cassette as described herein in a vector, in particular a nucleic acid vector.

In an aspect, the invention also provides a vector comprising a nucleic acid regulatory element as disclosed herein. In further embodiments, the vector comprises a nucleic acid expression cassette as disclosed herein.

The term “vector” as used in the application refers to nucleic acid molecules, e.g. double-stranded DNA, which may have inserted into it another nucleic acid molecule (the insert nucleic acid molecule) such as, but not limited to, a cDNA molecule. The vector is used to transport the insert nucleic acid molecule into a suitable host cell. A vector may contain the necessary elements that permit transcribing the insert nucleic acid molecule, and, optionally, translating the transcript into a polypeptide. The insert nucleic acid molecule may be derived from the host cell, or may be derived from a different cell or organism. Once in the host cell, the vector can replicate independently of, or coincidental with, the host chromosomal DNA, and several copies of the vector and its inserted nucleic acid molecule may be generated. The vectors can be episomal vectors (i.e., that do not integrate into the genome of a host cell), or can be vectors that integrate into the host cell genome. The term 'vector' may thus also be

defined as a gene delivery vehicle that facilitates gene transfer into a target cell. This definition includes both non-viral and viral vectors. Non-viral vectors include but are not limited to cationic lipids, liposomes, nanoparticles, PEG, PEI, plasmid vectors (e.g. pUC vectors, bluescript vectors (pBS) and pBR322 or derivatives thereof that are devoid of bacterial sequences (minicircles)) transposons-based vectors (e.g. PiggyBac (PB) vectors or Sleeping Beauty (SB) vectors), etc. Viral vectors are derived from viruses and include but are not limited to retroviral, lentiviral, adeno-associated viral, adenoviral, herpes viral, hepatitis viral vectors or the like. Typically, but not necessarily, viral vectors are replication-deficient as they have lost the ability to propagate in a given cell since viral genes essential for replication have been eliminated from the viral vector. However, some viral vectors can also be adapted to replicate specifically in a given cell, such as e.g. a cancer cell, and are typically used to trigger the (cancer) cell-specific (onco)lysis. Virosomes are a non-limiting example of a vector that comprises both viral and non-viral elements, in particular they combine liposomes with an inactivated HIV or influenza virus (Yamada et al., 2003). Another example encompasses viral vectors mixed with cationic lipids.

In preferred embodiments, the vector is a viral vector, such as a retroviral, lentiviral, adenoviral, or adeno-associated viral (AAV) vector, more preferably an AAV vector.

AAV vectors are preferably used as self-complementary, double-stranded AAV vectors (scAAV) in order to overcome one of the limiting steps in AAV transduction (i.e. single-stranded to double-stranded AAV conversion) (McCarty, 2001, 2003; Nathwani et al, 2002, 2006, 2011; Wu et al., 2008), although the use of single-stranded AAV vectors (ssAAV) are also encompassed herein.

Different serotypes of AAVs have been isolated and characterized, such as, for example AAV serotype 2, AAV serotype 5, AAV serotype 8, and AAV serotype 9, and all AAV serotypes are contemplated herein. AAV serotype 9 (AAV9) is ideally suited to achieve efficient transduction in liver. Accordingly, in certain embodiments, the vector is an AAV9 vector, more particularly a self-complementary AAV9 vector (scAAV9). Alternatively, engineered AAV capsids can be employed that impact on the efficacy of liver transduction, including this obtained by molecular evolution and selection, as described by Grimm et al. (2008) or Lisowski et al (2014).

In other embodiments, the vector is a non-viral vector, preferably a plasmid, a minicircle, or a transposon-based vector. Preferably, said transposon-based vectors are derived from Sleeping Beauty (SB) or PiggyBac (PB). A preferred SB transposon has been described in Ivics et al. (1997) and its hyperactive versions, including SB100X, as described in Mates et al. (2009). PiggyBac-based transposons are safe vectors in that they do no enhance the tumorigenic risk. Furthermore, liver-directed gene therapy with these vectors was shown to induce immune tolerance towards the transgene, in particular the hFIX or hFVIII transgene, comprised in the vector.

The transposon-based vectors are preferably administered in combination with a vector encoding a transposase for gene therapy. For example, the PiggyBac-derived transposon-based vector can be

administered with wild-type PiggyBac transposase (Pbase) or mouse codon-optimized PiggyBac transposase (mPBase) Preferably, said transposases are hyperactive transposases, such as, for example, hyperactive PB (hyPB) transposase containing seven amino acid substitutions (I30V, S103P, G165S, M282V, S509G, N538K, N570S) as described in Yusa et al. (2011), which is specifically incorporated by reference herein.

Transposon/transposase constructs can be delivered by hydrodynamic injection or using non-viral nanoparticles to transfect hepatocytes.

In yet other embodiments, the vector comprises viral and non-viral elements.

In specific embodiments, the following vectors are provided: AAVsc-HS-CRE1-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 33), AAVsc-HS-CRE2-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 34), AAVsc-HS-CRE3-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 35), AAVsc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 36), AAVsc-HS-CRE5-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 37), AAVsc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 38), AAVsc-HS-CRE7-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 39), AAVsc-HS-CRE8-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 40), AAVsc-HS-CRE9-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 41), AAVsc-HS-CRE10-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 42), AAVsc-HS-CRE11-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 43), AAVsc-HS-CRE12-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 44), AAVsc-HS-CRE13-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 45), AAVsc-HS-CRE14-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 46), AAVsc-HS-CRE15-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 47), AAVsc-HS-CRE16-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 48), AAVsc-HS-CRE17-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 49), AAVsc-HS-CRE18-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 50), AAVsc-HS-CRE19-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 51), AAVsc-HS-CRE20-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 52), AAVsc-HS-CRE21-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 53).

Accordingly, in particular embodiments, the vector has a sequence as set forth in SEQ ID NO: 38, 36, 44, 33, 37, 41, 43, 45, 46, 48 or 52, preferably a sequence as set forth in SEQ ID NO: 38, 36,44, 33, 37 or 45, more preferably a sequence as set forth in SEQ ID NO: 38, 36 or 44, even more preferably a sequence as set forth in SEQ ID NO: 38.

In a further aspect, the invention relates to the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein for use in medicine.

More particularly, the nucleic acid expression cassettes and vectors disclosed herein may be used, for example, to express proteins that are normally expressed and utilized in liver (i.e. structural proteins), or

to express proteins that are expressed in liver and that are then exported to the blood stream for transport to other portions of the body (i.e. secretable proteins). For example, the expression cassettes and vectors disclosed herein may be used to express a therapeutic amount of a gene product (such as a polypeptide, in particular a therapeutic protein, or RNA) for therapeutic purposes, in particular for gene therapy. Typically, the gene product is encoded by the transgene within the expression cassette or vector, although in principle it is also possible to increase expression of an endogenous gene for therapeutic purposes.

Accordingly, in particular embodiments, the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein may be for use in gene therapy, in particular liver-directed gene therapy.

Also disclosed herein is the use of the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein for the manufacture of a medicament for gene therapy, in particular liver-directed gene therapy.

Gene therapy protocols, intended to achieve therapeutic gene product expression in target cells, *in vitro*, but also particularly *in vivo*, have been extensively described in the art. These include, but are not limited to, intramuscular injection of plasmid DNA (naked or in liposomes), interstitial injection, instillation in airways, application to endothelium, intra-hepatic parenchyme, and intravenous or intra-arterial administration (e.g. intra-hepatic artery, intra-hepatic vein). Various devices have been developed for enhancing the availability of DNA to the target cell. A simple approach is to contact the target cell physically with catheters or implantable materials containing DNA. Another approach is to utilize needle-free, jet injection devices which project a column of liquid directly into the target tissue under high pressure. These delivery paradigms can also be used to deliver viral vectors. Another approach to targeted gene delivery is the use of molecular conjugates, which consist of protein or synthetic ligands to which a nucleic acid-or DNA- binding agent has been attached for the specific targeting of nucleic acids to cells (Cristiano et al., 1993a &b).

According to particular embodiments, the use of the nucleic acid expression cassettes and vectors as described herein is envisaged for gene therapy of liver cells (i.e. liver-directed gene therapy). According to a further particular embodiment, the use of the regulatory elements, expression cassettes or vectors is for gene therapy, in particular liver-directed gene therapy, *in vivo*. According to yet a further particular embodiment, the use is for a method of gene therapy, in particular liver-directed gene therapy, to treat hemophilia, in particular to treat hemophilia B or hemophilia A.

Gene transfer into mammalian hepatocytes has been performed using both *ex vivo* and *in vivo* procedures. The *ex vivo* approach requires harvesting of the liver cells, *in vitro* transduction with long-term expression vectors, and reintroduction of the transduced hepatocytes into the portal circulation (Kay et al., 1992; Chowdhury et al., 1991). *In vivo* targeting has been done by injecting DNA or viral vectors into the liver parenchyma, hepatic artery, or portal vein, as well as via transcriptional targeting

(Kuriyama et al., 1991; Kistner et al., 1996). Recent methods also include intraportal delivery of naked DNA (Budker et al., 1996) and hydrodynamic tail vein transfection (Liu et al., 1999; Zhang et al., 1999).

Also disclosed herein is the use of the nucleic acid regulatory elements, the nucleic acid expression cassettes, or the vectors disclosed herein for *in vivo or in vitro* expressing a protein in liver cells, more particularly by in vitro or in vivo transfecting or transducing liver cells with the nucleic acid expression cassette or the vector as described herein.According to a further aspect, methods for expressing a protein in liver cells are provided, comprising the steps of introducing in liver cells the nucleic acid regulatory element, the nucleic acid expression cassette or the vector as described herein, for example by transfection or transduction, and expressing the transgene protein product in the liver cells. These methods may be performed in vitro, ex vitro and in vivo.

A further aspect provides the nucleic acid regulatory elements, the nucleic acid expression cassettes or the vectors as taught herein for use in enhancing liver-specific gene expression of a transgene, preferably wherein said use is an *in vitro* use. A further aspect provides the nucleic acid regulatory elements, the nucleic acid expression cassettes or the vectors as taught herein for use in enhancing liver-specific expression of a protein or polypeptide encoded by a transgene, preferably wherein said use is an *in vitro* use

A further aspect provides a method for enhancing liver-specific gene expression of a transgene using the nucleic acid regulatory elements, the nucleic acid expression cassettes or the vectors as taught herein.

A further aspect provides a method for enhancing liver-specific expression of a protein or polypeptide encoded by a transgene using the nucleic acid regulatory elements, the nucleic acid expression cassettes or the vectors as taught herein.

In vivo methods of gene therapy for a subject in need thereof are also provided, comprising the steps of introducing in the liver of the subject a nucleic acid expression cassette containing a transgene encoding a therapeutic protein, and expressing a therapeutic amount of the therapeutic protein in the liver. According to a further embodiment, the method comprises the steps of introducing in the liver of the subject a vector comprising the nucleic acid expression cassette containing a transgene encoding a therapeutic protein, and expressing a therapeutic amount of the therapeutic protein in the liver.

Non-viral transfection or viral vector-mediated transduction of liver cells may also be performed by *in vitro* or *ex vivo* procedures. The *in vitro* approach requires the *in vitro* transfection or transduction of liver cells, e.g. liver cells previously harvested from a subject, liver cell lines or liver cells differentiated from e.g. induced pluripotent stem cells or embryonic cells. The *ex vivo* approach requires harvesting of the liver cells from a subject, *in vitro* transfection or transduction, and optionally re-introduction of the transfected liver cells into the subject.

Also disclosed herein are *in vitro* or *ex vivo* methods for the production of therapeutic proteins, said method comprising:

- transfecting or transducing liver cells, preferably hepatic cell lines, with a nucleic acid expression cassette or a vector disclosed herein, wherein the nucleic acid expression cassette or the vector comprises a nucleic acid regulatory element disclosed herein operably linked to a promoter and a transgene, wherein the transgene encodes a therapeutic protein;
- culturing the liver cells under conditions suitable for expressing the transgene in the liver cells; and
- harvesting the therapeutic protein from the liver cell or the culture medium.

The herein disclosed *in vitro* uses and methods for expressing a transgene product in liver cells, more particularly in hepatic cell lines, may be particularly suitable for the recombinant production of therapeutic proteins such as therapeutics used in protein replacement therapy (e.g. factor VIII, factor IX, factor VIIa). The use of the nucleic acid regulatory elements disclosed herein allows to enhance liver-specific expression of transgenes, which advantageously results in increased production of the transgene product, more particularly the therapeutic protein.

It is understood by the skilled person that the use of the nucleic acid regulatory elements, the nucleic acid expression cassettes and vectors disclosed herein has implications beyond gene therapy, e.g. coaxed differentiation of stem cells into hepatogenic cells, transgenic models for over-expression of proteins in liver, etc.

The transgene product may be a polypeptide, in particular a therapeutic protein such as, e.g. factor VIII, factor IX, factor VII or factor VIIa, factor X, α--glucosidase (secretable GAA), von Willebrand factor, phenylalaninehydroxylase (PAH), C1 esterase inhibitor (C1-INH), lysosomal enzymes, lysosomal enzyme iduronate-2-sulfatase (I2S), erythropoietin (EPO), interferon-α, interferon-β, interferon-γ, interleukin 1 (IL-1), interleukin 2 (IL-2), interleukin 3 (IL-3), interleukin 4 (IL-4), interleukin 5 (IL-5), interleukin 6 (IL-6), interleukin 7 (IL-7), interleukin 8 (IL-8), interleukin 9 (IL-9), interleukin 10 (IL-10), interleukin 11 (IL-11), interleukin 12 (IL-12), chemokine (C-X-C motif) ligand 5 (CXCL5), granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor (SCF), keratinocyte growth factor (KGF), monocyte chemoattractant protein-1 (MCP-1), tumor necrosis factor (TNF), afamin (AFM), α1-antitrypsin, α-galactosidase A, α-L-iduronidase, ATP7b, ornithine transcarbamoylase, phenylalanine hydroxylase, lipoprotein lipase, apoliproteins, low-density lipoprotein receptor (LDL-R), albumin, glucose-6-phosphatase, transgenes encoding antibodies, nanobodies, anti-viral dominant-negative proteins, and fragments, subunits or mutants thereof. Alternatively, the transgene product may be RNA, such as siRNA.

Exemplary diseases and disorders that may benefit from gene therapy using the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein are liver diseases and disorders such as hepatitis infection, hemophilia, including hemophilia A and B, glycogen storage disorders (GSD) such as Pompe (e.g. GSD II, GSD III, and GSD IVall), (non-alcoholic) fatty liver disease (NAFLD), lysosomal storage diseases, Hunter syndrome, phenylketonuria (PKU), hereditary angioedema (HAE), liver cirrhosis, fascioliasis, alcoholic liver disease, liver cancer, including primary liver cancer (e.g., hepatocellular carcinoma and/or cholangiocarcinoma, angiosarcoma or hemangiosarcoma) and liver metastases, biliary cirrhosis, sclerosing cholangitis, centrilobular necrosis of liver, Budd–Chiari syndrome, hemochromatosis, Wilson's disease, transthyretin-related hereditary amyloidosis, Gilbert's syndrome, and pediatric liver diseases and disorders such as biliary atresia, alpha-1 antitrypsin deficiency, Alagille syndrome, and progressive familial intrahepatic cholestasis

"Fascioliasis" refers to a parasitic infection of the liver caused by a liver fluke of the *Fasciola* genus, mostly the *Fasciola hepatica*. "Hepatitis" refers to inflammation of the liver, and may be caused by various viruses (viral hepatitis) but also by some liver toxins (e.g. alcoholic hepatitis), autoimmunity (autoimmune hepatitis) or hereditary conditions. With "alcoholic liver disease" is meant herein any hepatic manifestation of alcohol overconsumption, including fatty liver disease, alcoholic hepatitis, and cirrhosis. Analogous terms such as "drug-induced" or "toxic" liver disease are also used herein to refer to the range of disorders caused by various drugs and environmental chemicals. "Fatty liver disease" (or hepatic steatosis) refers herein to a reversible condition where large vacuoles of triglyceride fat accumulate in liver cells. "Non-alcoholic fatty liver disease" denotes a spectrum of disease associated with obesity and metabolic syndrome, among other causes. Fatty liver may lead to inflammatory disease (i.e. steatohepatitis) and, eventually, cirrhosis. "Cirrhosis" denotes the formation of fibrous tissue (fibrosis) in the place of liver cells that have died due to a variety of causes, including e.g., viral hepatitis, alcohol overconsumption, and other forms of liver toxicity. Cirrhosis may cause chronic liver failure. "Primary liver cancer" most commonly manifests as "hepatocellular carcinoma" and/or "cholangiocarcinoma", rarer forms include angiosarcoma and "hemangiosarcoma" of the liver. Many liver malignancies are secondary lesions that have metastasized from primary cancers in the gastrointestinal tract or other organs, such as the kidneys, lungs, breast, or prostate. "Primary biliary cirrhosis" refers to a serious autoimmune disease of the bile capillaries. "Primary sclerosing cholangitis" refers to a serious chronic inflammatory disease of the bile duct, which is believed to be autoimmune in origin. "Centrilobular necrosis of liver" can be caused by leakage of enteric toxins into circulation. For example, *Salmonella toxins* in ileum have been shown to cause severe damage to liver hepatic cells. "Budd–Chiari syndrome" denotes the clinical picture caused by occlusion of the hepatic vein, which in some cases may lead to cirrhosis. Hereditary diseases that cause damage to the liver include "hemochromatosis", involving accumulation of iron in the body, and "Wilson's disease", which causes the body to retain copper. Liver damage is also a clinical feature of "alpha 1-antitrypsin deficiency" and

"glycogen storage disease type II". In "transthyretin-related hereditary amyloidosis", the liver produces a mutated transthyretin protein which has severe neurodegenerative and/or cardiopathic effects. "Gilbert's syndrome" refers to a genetic disorder of bilirubin metabolism, which can cause mild jaundice.

According to a very specific embodiment, the therapeutic protein encoded by the transgene in the nucleic acid expression cassette or the vector is factor IX, and the method is a method for treating hemophilia B. By expressing factor IX in the liver via gene therapy, hemophilia B can be treated (Snyder et al., 1999). According to another very specific embodiment, the therapeutic protein encoded by the transgene in the nucleic acid expression cassette or the vector is factor VIII, and the method is a method for treating hemophilia A.

Except when noted differently, the terms "subject" or "patient" are used interchangeably and refer to animals, preferably vertebrates, more preferably mammals, and specifically includes human patients and non-human mammals, such as e.g. mice. Preferred patients or subjects are human subjects.

As used herein, the terms "treat" or "treatment" refer to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) an undesired physiological change or disorder, such as the development or spread of proliferative disease, e.g., cancer. Beneficial or desired clinical results include, but are not limited to, alleviation of symptoms, diminishment of extent of disease, stabilised (i.e., not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, and remission (whether partial or total), whether detectable or undetectable. "Treatment" can also mean prolonging survival as compared to expected survival if not receiving treatment.

As used herein, a phrase such as "a subject in need of treatment" includes subjects, such as mammalian subjects, that would benefit from treatment of a given condition, such as, hemophilia B or hemophilia A. Such subjects will typically include, without limitation, those that have been diagnosed with the condition, those prone to have or develop the said condition and/or those in whom the condition is to be prevented.

The term "therapeutically effective amount" refers to an amount of a compound or pharmaceutical composition effective to treat a given condition in a subject, i.e., to obtain a desired local or systemic effect and performance. The term thus refers to the quantity of compound or pharmaceutical composition that elicits the biological or medicinal response in a tissue, system, animal, or human that is being sought by a researcher, veterinarian, medical doctor or other clinician, which includes alleviation of the symptoms of the disease or disorder being treated. In particular, these terms refer to the quantity of compound or pharmaceutical composition according to the invention which is necessary to prevent, cure, ameliorate, or at least minimize the clinical impairment, symptoms, or complications associated with a given condition, such as hemophilia if therapeutic protein encoded by the transgene is factor IX or VIII, in either a single or multiple dose.

In a particular embodiment, if the therapeutic protein encoded by the transgene is factor IX, the term implies that levels of factor IX in plasma are equal to or higher than the therapeutic concentration of at least about 1% of physiological activity, i.e. 10mU/ml (milli-units per milliliter) plasma, at least 5% of physiological activity or 50mU/ml plasma, at least 10% of physiological activity or 100mU/ml plasma, at least 15% of physiological activity or 150mU/ml, at least 20% of physiological activity or 200mU/ml plasma, at least 25% of physiological activity or 250mU/ml, at least 30% of physiological activity or 300mU/ml, at least 35% of physiological activity or 350mU/ml, at least 40% of physiological activity or 400mU/ml, at least 45% of physiological activity or 400mU/ml, at least 45% of physiological activity or 450mU/ml, at least 50% of physiological activity or 500mU/ml, at least 65% of physiological activity or 650mU/ml, at least 70% of physiological activity or 700mU/ml, at least 75% of physiological activity or 750mU/ml, at least 80% of physiological activity or 800mU/ml, at least 85% of physiological activity or 850mU/ml, at least 95% of physiological activity or 950mU/ml, or at least 100% of physiological activity or 1000mU/ml, in a subject can be obtained by transduction or transfection of the vector according to any one the embodiments described herein into a subject. Due to the very high efficiency of the nucleic acid expression cassettes and vectors of the present invention, this high therapeutic levels of factor IX in the subject can be obtained even by administering relatively low doses of vector.

In another particular embodiment, if the therapeutic protein encoded by the transgene is factor VIII, the term implies that through levels of factor VIII in plasma equal to or higher than the therapeutic concentration of 10mU/ml (milli-units per milliliter) plasma, 50mU/ml plasma, 100mU/ml plasma, 150mU/ml plasma, 200mU/ml plasma, 250mU/ml plasma, 300 mU/ml plasma, 350mU/ml plasma, 400mU/ml plasma, 450mU/ml plasma, 500mU/ml plasma, 550mU/ml plasma, 600mU/ml plasma, 650mU/ml plasma, 750mU/ml plasma, 800mU/ml plasma, 850mU/ml plasma, 900mU/ml plasma, 950mU/ml plasma, or higher can be obtained by transduction or transfection of any of the vectors disclosed herein into a subject. Due to the very high efficiency of the vectors and nucleic acid expression cassettes disclosed herein, these high therapeutic levels of factor VIII in the subject can be obtained even by administering relatively low doses of vector.

In particular embodiments, if the transgene encodes factor IX or factor VIII, the transduction of the vector according to any one of the embodiments defined herein into the subject can be done at a dose lower than 6×10^{13} vg/kg (viral genomes per kilogram) to obtain a therapeutic factor IX level of 100 mU/ml plasma or higher in a subject. For example, a level of factor IX of 300 mU/ml plasma or higher in a subject may be achieved at a dose lower than 5×10^{11} vg/kg. For hemophilia therapy, efficacy of the treatment can, for example, be measured by assessing the hemophilia-caused bleeding in the subject. In vitro tests such as, but not limited to the in vitro activated partial thromboplastin time assay (APPT), test factor IX chromogenic activity assays, blood clotting times, factor IX or human factor VIII-specific ELISAs are also available. Any other tests for assessing the efficacy of the treatment known in the art can of course be used.

The nucleic acid expression cassette, the vector or the pharmaceutical composition of the invention may be used alone or in combination with any of the know hemophilia therapies, such as the administration of recombinant or purified clotting factors. The nucleic acid expression cassette, the vector or the pharmaceutical composition of the invention can thus be administered alone or in combination with one or more active compounds. The latter can be administered before, after or simultaneously with the administration of the said agent(s).

A further object of the invention are pharmaceutical preparations which comprise a therapeutically effective amount of the nucleic acid expression cassette or the expression vector as defined herein, and a pharmaceutically acceptable carrier, i.e., one or more pharmaceutically acceptable carrier substances and/or additives, e.g., buffers, carriers, excipients, stabilisers, etc. The pharmaceutical composition may be provided in the form of a kit. The term “pharmaceutically acceptable” as used herein is consistent with the art and means compatible with the other ingredients of the pharmaceutical composition and not deleterious to the recipient thereof. The term "pharmaceutically acceptable salts" as used herein means an inorganic acid addition salt such as hydrochloride, sulfate, and phosphate, or an organic acid addition salt such as acetate, maleate, fumarate, tartrate, and citrate. Examples of pharmaceutically acceptable metal salts are alkali metal salts such as sodium salt and potassium salt, alkaline earth metal salts such as magnesium salt and calcium salt, aluminum salt, and zinc salt. Examples of pharmaceutically acceptable ammonium salts are ammonium salt and tetramethylammonium salt. Examples of pharmaceutically acceptable organic amine addition salts are salts with morpholine and piperidine. Examples of pharmaceutically acceptable amino acid addition salts are salts with lysine, glycine, and phenylalanine. The pharmaceutical composition according to the invention can be administered orally, for example in the form of pills, tablets, lacquered tablets, sugar-coated tablets, granules, hard and soft gelatin capsules, aqueous, alcoholic or oily solutions, syrups, emulsions or suspensions, or rectally, for example in the form of suppositories. Administration can also be carried out parenterally, for example subcutaneously, intramuscularly or intravenously in the form of solutions for injection or infusion. Other suitable administration forms are, for example, percutaneous or topical administration, for example in the form of ointments, tinctures, sprays or transdermal therapeutic systems, or the inhalative administration in the form of nasal sprays or aerosol mixtures, or, for example, microcapsules, implants or rods. The pharmaceutical composition can be prepared in a manner known per se to one of skill in the art. For this purpose, the nucleic acid expression cassette or the expression vector as defined herein, one or more solid or liquid pharmaceutically acceptable excipients and, if desired, in combination with other pharmaceutical active compounds, are brought into a suitable administration form or dosage form which can then be used as a pharmaceutical in human medicine or veterinary medicine.

According to another aspect, a pharmaceutical composition is provided comprising a nucleic acid expression cassette containing a transgene encoding a therapeutic protein, and a pharmaceutically acceptable carrier. According to another embodiment, the pharmaceutical composition comprises a vector containing the nucleic acid expression cassette containing a transgene encoding a therapeutic

protein, and a pharmaceutically acceptable carrier. According to further particular embodiments, the transgene encodes factor IX and the pharmaceutical composition is for treating hemophilia B or the transgene encodes factor VIII and the pharmaceutical composition is for treating hemophilia A.

The use of the nucleic acid expression cassette, its regulatory elements and the vector components as disclosed herein for the manufacture of these pharmaceutical compositions for use in treating hemophilia, preferably hemophilia B or hemophilia A, is also envisaged.

In an alternative example, the expression cassettes and vectors disclosed herein may be used to express an immunological amount of a gene product (such as a polypeptide, in particular an immunogenic protein, or RNA) for vaccination purposes.

In embodiments, the pharmaceutical composition may be a vaccine. The vaccine may further comprise one or more adjuvants for enhancing the immune response. Suitable adjuvants include, for example, but without limitation, saponin, mineral gels such as aluminium hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil or hydrocarbon emulsions, bacilli Calmette-Guerin (BCG), Corynebacterium parvum, and the synthetic adjuvant QS-21. Optionally, the vaccine may further comprise one or more immunostimulatory molecules. Non-limiting examples of immunostimulatory molecules include various cytokines, lymphokines and chemokines with immunostimulatory, immunopotentiating, and pro-inflammatory activities, such as interleukins (e.g., IL-1, IL-2, IL-3, IL-4, IL-12, IL-13); growth factors (e.g., granulocyte-macrophage (GM)-colony stimulating factor (CSF)); and other immunostimulatory molecules, such as macrophage inflammatory factor, Flt3 ligand, B7.1; B7.2, etc.

In embodiments, the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein may be for use as a vaccine, more particularly for use as a prophylactic vaccine.

Also disclosed herein is the use of the nucleic acid regulatory elements, the nucleic acid expression cassettes, the vectors, or the pharmaceutical compositions described herein for the manufacture of a vaccine, in particular for the manufacture of a prophylactic vaccine.

Also disclosed herein is a method of vaccination, in particular prophylactic vaccination, of a subject in need of said vaccination comprising:

- introducing in the subject, in particular in liver of the subject, a nucleic acid expression cassette, a vector or a pharmaceutical composition described herein, wherein the nucleic acid expression cassette, the vector or the pharmaceutical composition comprises a nucleic acid regulatory element described herein operably linked to a promoter and a transgene; and
- expressing an immunologically effective amount of the transgene product in the subject, in particular in liver of the subject.

An "immunologically effective amount" as used herein refers to the amount of (trans)gene product effective to enhance the immune response of a subject against a subsequent exposure to the immunogen encoded by the (trans)gene. Levels of induced immunity can be determined, e.g. by measuring amounts of neutralizing secretory and/or serum antibodies, e.g., by plaque neutralization, complement fixation, enzyme-linked immunosorbent, or microneutralization assay.

Typically, the amount of (trans)gene product expressed when using an expression cassette or vector as taught herein (i.e., with at least one liver-specific nucleic acid regulatory element comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof as described herein elsewhere) are higher than when an identical expression cassette or vector is used but without at least one liver-specific nucleic acid regulatory element as taught herein therein (e.g., comprising TTRe and/or Serp). More particularly, the expression is from about 1.1-fold to about 1.5-fold the expression obtained by the same nucleic acid expression cassette or vector without at least one liver-specific nucleic acid regulatory element as taught herein but comprising 3xSerp; or is from about 2-fold to about 6-fold the expression obtained by the same nucleic acid expression cassette or vector without at least one liver-specific nucleic acid regulatory element as taught herein (e.g., comprising TTRe and/or Serp). Moreover, the higher expression remains specific to liver. Furthermore, the expression cassettes and vectors described herein direct the expression of a therapeutic amount of the gene product for an extended period. Typically, therapeutic expression is envisaged to last at least 20 days, at least 50 days, at least 100 days, at least 200 days, and in some instances 300 days or more. Expression of the gene product (e.g. polypeptide) can be measured by any art-recognized means, such as by antibody-based assays, e.g. a Western Blot or an ELISA assay, for instance to evaluate whether therapeutic expression of the gene product is achieved. Expression of the gene product may also be measured in a bioassay that detects an enzymatic or biological activity of the gene product.

It is to be understood that although particular embodiments, specific constructions and configurations, as well as materials, have been discussed herein for methods and applications according to the present invention, various changes or modifications in form and detail may be made without departing from the scope and spirit of this invention.

The following examples are provided to better illustrate particular embodiments, and they should not be considered limiting the application. The application is limited only by the claims.

EXAMPLES

Example 1: Identification of liver-specific regulatory elements

Materials and methods

A computational approach to identify evolutionarily conserved transcription factor binding site (TFBS) motifs associated with highly expressed liver- specific genes defined herein as nucleic acid regulatory elements. This required several consecutive computational steps: (1) liver-specific genes were identified that are highly expressed based on RNAseq (RNA sequencing) expression data obtained with normal human tissues; The mRNA expression levels based on RNAseq of the top liver-expressed genes are as follows (in arbitrary units): ALB (46017,49); APOA2 (21859,94); TTR (9551,38); APOC1 (9505,84); APOA1 (9277,21); RBP4 (8322,91); APOC3 (8105,76); CYP2E1 (5932,55); ORM1 (5860,23); (2) publicly available databases were used to extract the corresponding promoter sequences; (3) a computational approach was employed to identify clusters of transcription factor binding site motifs (TFBS); (4) the genomic context of the highly expressed genes was screened for evolutionary conserved clusters of TFBS (i.e. CREs).

Results

RNA sequence (RNAseq) analysis led to a comprehensive list of highly expressed hepatocyte-specific genes. Based on these expression data, the promoters of these highly expressed genes were subjected to the aforementioned computational approach that led to the identification of 21 liver-specific cis-regulatory sequences (SEQ ID NO: 1 to 21) and Table 1.

Table 1

HS-CRE	Gene	Size (bp)	Position	SEQ ID NO
HS-CRE1	**ALB**	**336**	**Chr4:74269710-74270045**	**1**
HS-CRE2	**ALB**	**347**	**Chr4:74267148-74267494**	**2**
HS-CRE3	**APOA2**	**386**	**Chr1:161193242-161193627**	**3**
HS-CRE4	**APOA2**	**426**	**Chr1:161193972-161194397**	**4**
HS-CRE5	**TTR**	**260**	**Chr18:29171606-29171865**	**5**
HS-CRE6	**APOC1**	**267**	**Chr19:45416160-45416426**	**6**
HS-CRE7	**APOC1**	**313**	**Chr19:45418262-**	**7**

			45418574	
HS-CRE8	APOC1	513	Chr19:45417617-45418129	8
HS-CRE9	APOA1	306	Chr11:116708313-116708618	9
HS-CRE10	APOA1	406	Chr11:116709482-116709887	10
HS-CRE11	APOA1	504	Chr11:116711498-116712001	11
HS-CRE12	RBP4	431	Chr10:95360922-95361352	12
HS-CRE13	RBP4	448	Chr10:95361438-95361885	13
HS-CRE14	APOC3	335	Chr11:116700409-116700743	14
HS-CRE15	APOC3	431	Chr11:116699651-116700081	15
HS-CRE16	APOC3	559	Chr11:116697474-116698032	16
HS-CRE17	APOC3	370	Chr11:116696862-116697231	17
HS-CRE18	APOC3	431	Chr11:116696149-116696579	18
HS-CRE19	CYP2E1	481	Chr10:135342646-135343126	19
HS-CRE20	ORM1	345	Chr9:117084721-117085065	20
HS-CRE21	ORM1	386	Chr9:117080121-117080506	21

Example 2: *In vivo* validation of the identified hepatocyte-specific cis-regulatory elements (CREs) by hydrodynamic plasmid transfection

Materials and methods

Generation of the AAV plasmid constructs (pAAV-HS-CRE-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA)

The hepatocyte-specific (HS) cis-regulatory elements (CRE) HS-CRE1, HS-CRE2, HS-CRE3, HS-CRE4, HS-CRE5, HS-CRE6, HS-CRE7, HS-CRE8, HS-CRE9, HS-CRE10, HS-CRE11, HS-CRE12, HS-CRE13, HS-CRE14, HS-CRE15, HS-CRE16, HS-CRE17, HS-CRE18, HS-CRE19, HS-CRE20, HS-CRE21, identified in Example 1 were synthesized by conventional oligonucleotide synthesis and cloned into an AAVsc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA construct (SEQ ID NO: 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52 and 53, respectively). This construct was based on a self-complementary adeno-associated viral vector (scAAV) design and contained a minimal transthyretin promoter (TTRm) (SEQ ID NO: 27) coupled to a TTR enhancer (TTRe or TTRenh) (SEQ ID NO: 24) driving expression of a codon-optimized hyperactive human factor IX Padua (R338L) (SEQ ID NO: 25) cDNA. The different HS-CREs were cloned upstream of the TTRm/TTRe promoter/enhancer. The construct also contained an intron from minute virus of mouse (MVM) (SEQ ID NO: 29) and a bovine growth hormone polyadenylation site (bghpolyA) (SEQ ID NO: 30).

An additional plasmid construct containing the previously identified Serpin (SERP) cis-regulatory element as triplicate (SEQ ID NO: 22) (equivalent to HS-CRM8 as described in Nair et al., Blood, 2014; Chuah et al.; Mol Ther, 2014), designated as pAAVsc-3xSERP-TTRe-TTRm-MVM-FIXcoPadua-bghpA (SEQ ID NO: 32), was included for comparison (WO 2009/130208 and WO 2016/146757 which are specifically incorporated by reference herein). The different constructs are schematically represented in Figure 1.

Hydrodynamic gene delivery

The AAV plasmid constructs AAVsc-HS-CRE-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA *(containing the following HS-CRE:* HS-CRE1, HS-CRE2, HS-CRE3, HS-CRE4, HS-CRE5, HS-CRE6, HS-CRE7, HS-CRE8, HS-CRE9, HS-CRE10, HS-CRE11, HS-CRE12, HS-CRE13, HS-CRE14, HS-CRE15, HS-CRE16, HS-CRE17, HS-CRE18, HS-CRE19, HS-CRE20 or HS-CRE21, identified in Example 1) were injected hydrodynamically via the tail vein in adult C57BL/6 mice (18 to 19 g) at a dose of 750 ng per mouse in a total volume of 2 ml phosphate buffered saline (PBS). The AAVsc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA plasmid (SEQ ID NO: 31) without any HS-CRE element upstream of TTRenh-TTRm was used as a reference construct to assess the impact of the different cis-regulatory elements on the FIX expression levels. Two days after hydrodynamic injection, blood was collected into buffered citrate for all mice by phlebotomy of the retro-orbital plexus. The concentration of hFIX antigens in citrated plasma was measured by enzyme-linked immunosorbent assay (ELISA) based on the manufacturer's protocol (Diagnostica Stago, France). Each cohort included 3 mice per vector.

Results

The effect of the cis-regulatory elements identified in Example 1 on *in vivo* expression was assessed by measuring the FIX protein levels in mice that were injected with plasmid constructs expressing human clotting factor IX (hFIX) from a chimeric promoter composed of a potent liver-specific minimal transthyretin promoter in conjunction with the transthyrethin enhancer (TTRenh/TTRmin) in which the different CREs were cloned upstream of the TTRenh/TTRmin. FIX expression was assessed 1 or 2 days post-transfection demonstrating a significant increase in FIX by some of the new HS-CREs. In particular, about 52 % (11 of 21 HS-CREs) of the identified cis-regulatory elements induced a significant increase in hFIX expression on day 2 post-transfection (Figure 2) as compared to the control plasmid construct without any cis-regulatory element (i.e. AAVsc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA (SEQ ID NO: 31)). The newly identified regulatory elements were further compared to the previously identified liver-specific regulatory element designated as Serpin enhancer (WO 2009/130208, which is specifically incorporated by reference herein) present as a triplet repeat in the vector (i.e. pAAVsc-3xSERP-TTRe-TTRm-MVM-FIXcoPadua-bghpA (SEQ ID NO: 32)). We had previously demonstrated that repeating the SERP element 3 times boosts the expression of the gene of interest (e.g. FIX). Unexpectedly, we found that 3 of the new CREs, namely HS-CRE4, HS-CRE6 and HS-CRE12 outperformed all the other CREs (including SERP3X). In particular, HS-CRE4, HS-CRE6 and HS-CRE12 expressed FIX levels comparable or even higher than what could be achieved by incorporating a triplet SERP repeat in the AAV vector. Taking into consideration, the known effect of multiplying the SERP element on transgene expression levels, this indicates that, unexpectedly, HS-CRE4 (ApoA2; SEQ ID NO: 4), HS-CRE6 (ApoC1; SEQ ID NO: 6), and HS-CRE12 (RBP4; SEQ ID NO: 12) therefore significantly outperformed a single SERP element in terms of boosting transgene expression levels. Compared to all CREs, including SERP, HS-CRE6 was the most potent HS-CRE.

Example 3: *In vivo* validation of the identified hepatocyte-specific cis-regulatory elements (CREs) by AAV viral vector transduction

To confirm the *in vivo* results obtained by hydrodynamic (i.e. non-viral) plasmid transfection and broaden the scope of the invention towards also including viral vectors, AAV serotype 9 (designated as AAV9) vector particles were produced that contained the most potent CREs identified in Example 2 (namely HS-CRE4 and HS-CRE6), based on the non-viral, semi-high throughput *in vivo* screening and validation. The corresponding AAV9 vectors were based on the self-complementary (sc) configuration and designated as AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA and AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA. As controls, AAV9 vectors were produced that did not contain any CRE (designated as AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA)

or that contained a triplet repeat of the SERP element as reference (designated as AAV9sc-3XSERP-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA).

Vector production and purification and titration:

The aforementioned AAV serotype 9 vectors, namely AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA and AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-3XSERP-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, were produced by cotransfecting AAV-293 cells using calcium phosphate (Thermo Fisher Scientific, Waltham, MA, USA) with the AAV plasmid, a chimeric packaging construct and an adenoviral helper plasmid, as described previously. (VandenDriessche T, Thorrez L, Acosta-Sanchez A, et al. Efficacy and safety of adeno-associated viral vectors based on serotype 8 and 9 vs. lentiviral vectors for hemophilia B gene therapy. J Thromb Haemost. 2007; 5(1):16-24). Two days post-transfection cells were harvested and lysed by successive freeze/thaw cycles and sonication, followed by treatment with benzonase (Novagen, Madison, WI, USA) and deoxycholic acid (Sigma Aldrich (MERCK, check), St Louis, MO, USA) and 3 rounds of cesium chloride (Thermo Fisher Scientific, Waltham, MA, USA) density gradient ultracentrifugation. The fractions containing the AAV particles were collected and dialyzed in Dulbecco's phosphate buffered saline (D-PBS) (Gibco, IRL) containing 1 mM $MgCl_2$.

To determine vector titers, quantitative real-time PCR (ABI Prism 7900HT, Applied Biosystems, Foster City/CA, USA) with SYBR® Green (Thermo Fisher Scientific, Waltham, MA, USA) and primers for the FIX gene were used. The forward primer sequence was 5'-CACGAGAACGCCAACAAGAT-3' (SEQ ID NO: 55), the reverse primer sequence was 5'-CACTTCTCCTCCATGCACTC-3' (SEQ ID NO: 56). Standard curves were generated using known copy numbers of the corresponding vector plasmids.

AAV9 vector injection

The AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA, AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA and AAV9sc-3XSERP-TTRenh-TTRm-MVM-hFIXcoPadua-bghpoly vectors were injected at a dose of $1x10^9$ vg/mouse (Figure 3A) and a dose of $5x10^9$ vg/mouse (Figure 3B) into the tail vein of C57BL/6 mice (4-5 weeks old). Each cohort consists of 3 mice. Blood was collected at several time points after injection and plasma was analyzed for FIX protein levels by means of ELISA.

Mice injected with $5x10^9$ vg AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA or AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA yield significantly higher FIX levels than the control mice injected with the AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA vector that is devoid of any HS-CRE. This indicates that the HS-CRE4 and HS-CRE6 elements result in a significant 5 to 6-fold increase in FIX levels, that are relatively stable over at least 3 months. Similarly, mice injected with a lower dose (i.e. 10^9 vg) AAV9sc-HS-CRE4-TTRenh-TTRm-MVM-

hFIXcoPadua-bghpolyA or AAV9sc-HS-CRE6-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA vectors yield significantly higher FIX levels than the control mice injected with AAV9sc-TTRenh-TTRm-MVM-hFIXcoPadua-bghpolyA that is devoid of any HS-CRE. A dose-response was apparent, consistent with higher stable FIX levels following injection of $5x10^9$ vg compared to10^9 vg for all vectors tested.

The newly identified regulatory elements were further compared to the previously identified liver-specific regulatory element designated as Serpin enhancer (WO 2009/130208, which is specifically incorporated by reference herein) present as a triplet repeat in the vector (i.e. AAV9sc-3xSERP-TTRe-TTRm-MVM-FIXcoPadua-bghpA). We had previously demonstrated that repeating the SERP element 3 times boosts the expression of the gene of interest (e.g. FIX). Consistent with the hydrodynamic transfection data shown in Example 2, we found HS-CRE4 and HS-CRE6 expressed FIX levels comparable or even higher than what could be achieved by incorporating a triplet SERP repeat in the AAV vector. Taking into consideration, the known effect of multiplying the SERP element on transgene expression levels, this confirms once again, that, HS-CRE4 (ApoA2 ; SEQ ID NO: 4) and HS-CRE6 (ApoC1; SEQ ID NO: 6), outperformed a single SERP element in terms of boosting transgene expression levels. Compared to all CREs, including SERP, HS-CRE6 was again the most potent HS-CRE within the context of AAV9-based liver transduction.

Present inventors tested which transcription factors bind on HS-CREs 1-21 through experimental research.

It was found through experimental research that (1) all of HS-CRE6, HS-CRE4 and HS-CRE12 (which are the HS-CREs resulting in the highest liver-specific increase of the protein levels) have transcription factor binding sites for SP1, EP300, HNF 4G, CEBPB, P300, HDAC2, JUND, FOSL2, ZBTB7A, CEBPD and RXRA; (2) all of HS-CRE1, HS-CRE5 and HS-CRE13 (which are the HS-CREs showing a medium high liver-specific increase of the protein levels) have transcription factor binding sites for SP1, EP300, POLR2A, MYBL2, FOXA1, FOXA2, ARID3A, POLR2A and HEY1; and (3) all of HS-CRE9, HS-CRE11, HS-CRE14, HS-CRE16 and HS-CRE20 (which are the HS-CREs showing the lowest liver-specific increase of the protein levels) have transcription factor binding sites for SP1, EP300, HNF 4G, CEBPB, P300 and HNF 4A. Accordingly, all of HS-CRE6, HS-CRE4, HS-CRE12, HS-CRE1, HS-CRE5, HS-CRE13, HS-CRE9, HS-CRE11, HS-CRE14, HS-CRE16 and HS-CRE20 have transcription factor binding sites for SP1 and EP300.

References

BUDKER V, ZHANG G, KNECHTLE S, WOLFF JA. Naked DNA delivered intraportally expresses efficiently in hepatocytes. (1996) Gene Ther. Jul;3(7):593-8.

BRUNETTI-PIERRI N, GROVE NC, ZUO Y, EDWARDS R, PALMER D, CERULLO V, TERUYA J, NG P. Bioengineered factor IX molecules with increased catalytic activity improve the therapeutic index of gene therapy vectors for hemophilia B. Hum Gene Ther. 2009 May;20(5):479-85.

CHANG, J., JIN, J., LOLLAR, P., BODE, W., BRANDSTETTER, H., HAMAGUCHI, N., STRAIGHT, D. L. &STAFFORD, D. W. (1998). Changing residue 338 in human factor IX from arginine to alanine causes an increase in catalytic activity. J Biol Chem 273(20): 12089-12094.

CRISTIANO RJ, Smith LC, Kay MA, Brinkley BR, Woo SL. (1993a) Hepatic gene therapy: efficient gene delivery and expression in primary hepatocytes utilizing a conjugated adenovirus-DNA complex. Proc Natl Acad Sci USA. 90(24): 11548-11552.

CRISTIANO RJ, Smith LC, Woo SL. (1993b) Hepatic gene therapy: adenovirus enhancement of receptor-mediated gene delivery and expression in primary hepatocytes. 90(6): 2122-2126.

GRIMM D, LEE JS, WANG L, DESAI T, AKACHE B, STORM TA, KAY MA (2008). J Virol. 82(12):5887-5991.

IVICS Z, HACKETT PB, PLASTERK RH, IZSVÁK Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell. 91(4):501-510.

KAY MA, BALEY P, ROTHENBERG S, LELAND F, FLEMING L, PONDER KP, LIU T, FINEGOLD M, DARLINGTON G, POKORNY W, WOO SLC. (1992) Expression of human alpha 1 - antitrypsin in dogs after autologous transplantation of retroviral transduced hepatocytes. Proc Natl Acad Sci U S A. Jan 1 ;89(1):89-93.

KAO, C. Y., LIN, C. N., YU, I. S., TAO, M. H., WU, H. L., SHI, G. Y., YANG, Y. L., KAO, J. T. &LIN, S. W. (2010). FIX-Triple, a gain-of-function factor IX variant, improves haemostasis in mouse models without increased risk of thrombosis. Thromb Haemost 104(2): 355-365.

KISTNER A, GOSSEN M, ZIMMERMANN F, JERECIC J, ULLMER C, LYBBERT H, BUJARD H. (1996) Doxycycline- mediated quantitative and tissue-specific control of gene expression in transgenic mice. Proc Natl Acad Sci U S A. Oct 1;93(20): 10933-8.

LIN, C. N., KAO, C. Y., MIAO, C. H., HAMAGUCHI, N., WU, H. L., SHI, G. Y., LIU, Y. L., HIGH, K. A. &LIN, S. W. (2010). Generation of a novel factor IX with augmented clotting activities in vitro and in vivo. J Thromb Haemost 8(8): 1773-1783.

LISOWSKI L, DANE AP, CHU K, ZHANG Y, CUNNINGHAM SC, WILSON EM, NYGAARD S, GROMPE M, ALEXANDER IE, KAY MA.(2014) Nature. 506(7488):382-386.

LIU F, SONG Y, LIU D. (1999) Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. Gene Ther. Jul;6(7):1258-66.

McCARTY DM, MONAHAN PE, and SAMULSKI RJ. (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. 8, 1248-54.

McCARTY DM, FU H, MONAHAN PE, TOULSON CE, NAIK P, and SAMULSKI RJ. (2003). Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. Gene Ther. 10, 2112-8.

MÁTÉS L, CHUAH MK, BELAY E, JERCHOW B, MANOJ N, ACOSTA-SANCHEZ A, GRZELA DP, SCHMITT A, BECKER K, MATRAI J, MA L, SAMARA-KUKO E, GYSEMANS C, PRYPUTNIEWICZ D, MISKEY C, FLETCHER B, VANDENDRIESSCHE T,, IVICS Z, and IZSVAK Z. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat Genet. 41, 753-61.

MILANOV, ET AL., 2012 Engineered factor IX variants bypass FVIII and correct hemophilia A phenotype in mice Blood 119:602-611.

NAIR N, RINCON MY, EVENS H, SARCAR S, DASTIDAR S, SAMARA-KUKO E, GHANDEHARIAN O, MAN VIECELLI H, THÖNY B, DE BLESER P, VANDENDRIESSCHE T, CHUAH MK. (2014). Computationally designed liver-specific transcriptional modules and hyperactive factor IX improve hepatic gene therapy. Blood 123,3195-9.

NATHWANI AC, DAVIDOFF AM, HANAWA H, YUNYU HU, HOFFER FA, NIKANOROV A, SLAUGHTER C, NG CYC, ZHOU J, LOZIER J, MANDRELL TD, VANIN EF, and NIENHUIS AW. (2002). Sustained high- level expression of human factor IX (hFIX) after liver-targeted delivery of recombinant adeno-associated virus encoding the hFIX gene in rhesus macaques. Blood 100, 1662–1669.

NATHWANI AC, GRAY JT, NG CY, ZHOU J, SPENCE Y, WADDINGTON SN, TUDDENHAM EG, KEMBALL COOK G, McINTOSH J, BOON-SPIJKER M, MERTENS K, DAVIDOFF AM. (2006).Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. Blood 107, 2653-61.

NATHWANI AC, TUDDENHAM EG, RANGARAJAN S, ROSALES C, MCINTOSH J, LINCH DC, CHOWDARY P, RIDDELL A, PIE AJ, HARRINGTON C, O'BEIRNE J, SMITH K, PASI J, GLADER B, RUSTAGI P, NG CY, KAY MA, ZHOU J, SPENCE Y, MORTON CL, ALLAY J, COLEMAN J, SLEEP S, CUNNINGHAM JM, SRIVASTAVA D, BASNER-TSCHAKARJAN E, MINGOZZI F, HIGH KA, GRAY JT, REISS UM, NIENHUIS AW, and DAVIDOFF AM. (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med. 365, 2357-2365.

WU Z, SUN J, ZHANG T, YIN C, YIN F, VAN DYKE T, SAMULSKI RJ, and MONAHAN PE. (2008). Optimization of self-complementary AAV vectors for liver-directed expression results in sustained correction of hemophilia B at low vector dose. Mol Ther. 16, 280-9.

SCHUETTRUMPF, J., HERZOG, R. W., SCHLACHTERMAN, A., KAUFHOLD, A., STAFFORD, D. W. &ARRUDA, V. R. (2005). Factor IX variants improve gene therapy efficacy for hemophilia B. Blood 105(6): 2316-2323.

SIMIONI, P., TORMENE, D., TOGNIN, G., GAVASSO, S., BULATO, C., IACOBELLI, N. P., FINN, J. D., SPIEZIA, L., RADU, C. &ARRUDA, V. R. (2009). X-linked thrombophilia with a mutant factor IX (factor IX Padua). N Engl J Med 361(17): 1671-1675.

SNYDER RO, MIAO C, MEUSE L, TUBB J, DONAHUE BA, HUI-FENG LIN, STAFFORD DW, PATEL S, THOMPSON AR, NICHOLS T, READ MS , BELLINGER DA, BRINKHOUS KM, and KAY MA. (1999). Correction of hemophilia B in canine and murine models using recombinant adeno-associated viral vectors. Nat Med. 5, 64–70.

WARD, N.J. *ET AL.* Codon optimization of human factor VIII cDNAs leads to high-level expression. *Blood* **117**, 798-807 (2011).

YAMADA T, IWASAKI Y, TADA H, IWABUKI H, CHUAH MK, VANDENDRIESSCHE T, FUKUDA H, KONDO A, UEDA M, SENO M, TANIZAWA K, KURODA S. (2003) Nanoparticles for the delivery of genes and drugs to human hepatocytes. Nat Biotechnol. Aug;21 (8):885-90.

YUSA ET AL. A hyperactive piggyBac transposase for mammalian applications. Proc Natl Acad Sci U S A. 2011;108(4):1531-6.

ZHANG G, BUDKER V, WOLFF JA. (1999) High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA. Hum Gene Ther. JuI 1 ;10(10):1735-7.

DE BLESER P, HOOGHE B, VLIEGHE D, VAN ROY F. (2007) A distance difference matrix approach to identifying transcription factors that regulate differential gene expression. Genome Biol. 8(5): R83.

CHUAH, M.K., PETRUS, I., DE BLESER, P., LE GUINER, C., GERNOUX, G., ADJALI, O., NAIR, N., WILLEMS, J., EVENS, H., RINCÓN, M.Y., MATRAI, J., DI MATTEO, M., SAMARA-KUKO, E., YAN, B., ACOSTA-SANCHEZ, A., MELIANI, A., CHEREL, G., BLOUIN, V., CHRISTOPHE, O., MOULLIER, P., MINGOZZI, F., VANDENDRIESSCHE, T. (2014) Liver-specific transcriptional modules identified by genome-wide in silico analysis enable efficient gene therapy in mice and non-humans primates. Mol. Ther. 22(9):1605-13

GARMORY HS1, BROWN KA, TITBALL RW. (2003) DNA vaccines: improving expression of antigens. Genet Vaccines Ther. 1(1):2Wu C, Macleod I, Su AI. (2013) BioGPS and MyGene.info: organizing online, gene-centric information. Nucleic Acids Res. 41(Database issue):D561-5.

MCINTOSH J, LENTING PJ, ROSALES C, LEE D, RABBANIAN S, RAJ D, PATEL N, TUDDENHAM EG, CHRISTOPHE OD, MCVEY JH, WADDINGTON S, NIENHUIS AW, GRAY JT, FAGONE P, MINGOZZI F, ZHOU SZ, HIGH KA, CANCIO M, NG CY, ZHOU J, MORTON CL, DAVIDOFF AM, NATHWANI AC. (2003) Therapeutic levels of FVIII following a single peripheral vein administration of rAAV vector encoding a novel human factor VIII variant. Blood. 121(17):3335-44XIAO SJ, ZHANG C, ZOU Q, JI ZL. (2010) TiSGeD: database for tissue-specific genes. Bioinformatics 26(9):1273-5.

CLAIMS

1. A nucleic acid regulatory element for enhancing liver-specific gene expression, comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 20, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof; wherein said nucleic acid regulatory element comprises at least a transcription factor binding site (TFBS) for SP1 and for EP300.

2. The nucleic acid regulatory element according to claim 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 6, SEQ ID NO: 4 and SEQ ID NO: 12, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof, wherein said nucleic acid regulatory element further comprises a TFBS for HNF 4G, a TFBS for CEBPB, a TFBS for P300, a TFBS for HDAC2, a TFBS for JUND, a TFBS for FOSL2, a TFBS for ZBTB7A, a TFBS for CEBPD and a TFBS for RXRA.

3. The nucleic acid regulatory element according to claim 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 5 and SEQ ID NO: 13, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof wherein said nucleic acid regulatory element further comprises a TFBS for POLR2A, a TFBS for MYBL2, a TFBS for FOXA1, a TFBS for FOXA2, a TFBS for ARID3A, a TFBS for POLR2A and a TFBS for HEY1.

4. The nucleic acid regulatory element according to claim 1, comprising a sequence selected from the group consisting of: SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 14, SEQ ID NO: 16 and SEQ ID NO: 20, a sequence having at least 95% identity to any of said sequences, or a functional fragment thereof, wherein said nucleic acid regulatory element further comprises a TFBS for HNF 4G, a TFBS for CEBPB, a TFBS for P300 and a TFBS for HNF 4A.

5. A nucleic acid regulatory element for enhancing liver-specific gene expression hybridizing under stringent conditions to the nucleic acid regulatory element according to any one of claims 1 to 4, or to its complement.

6. The nucleic acid regulatory element according to any one of claims 1 to 5, having a maximal length of 600 nucleotides, preferably of 500 nucleotides, more preferably of 400 nucleotides, even more preferably of 300 nucleotides.

7. A nucleic acid expression cassette comprising at least one nucleic acid regulatory element according to any one of claims 1 to 6, operably linked to a promoter and a transgene.

8. The nucleic acid expression cassette according to claim 7, further comprising at least one, preferably three, liver-specific regulatory element(s) different from the nucleic acid regulatory element according to any one of claims 1 to 4, preferably wherein said at least one, preferably three, liver-specific

regulatory element(s) comprise SEQ ID NO: 22 or a sequence having at least 95% identity to said sequence.

9. The nucleic acid expression cassette according to claim 8, comprising a nucleic acid regulatory element comprising SEQ ID NO: 6, SEQ ID NO: 4; SEQ ID NO: 12, or a sequence having at least 95% identity to any one of said sequences, and at least one, preferably three, nucleic acid regulatory element(s) comprising SEQ ID NO: 22 or a sequence having at least 95% identity to said sequence.

10. The nucleic acid expression cassette according to any one of claims 7 to 9, wherein the transgene encodes a therapeutic protein or an immunogenic protein, preferably wherein said transgene encodes for

- coagulation factor IX (FIX), preferably wherein said transgene is codon-optimized coagulation factor FIX, or wherein said coagulation factor FIX contains a hyper-activating mutation, preferably wherein said hyper-activating mutation corresponds to an R338L amino acid substitution, more preferably wherein said transgene encoding for coagulation factor FIX has a nucleic acid sequence as defined by SEQ ID NO: 25; or

- coagulation factor VIII (FVIII), preferably wherein said transgene is codon-optimized coagulation factor FVIII, or wherein said coagulation factor VIII has a deletion of the B domain, preferably wherein said B domain of said FVIII is replaced by a linker defined by SEQ ID NO: 54, more preferably wherein said transgene encodes for coagulation factor VIII has a nucleic acid sequence defined by SEQ ID NO: 26.

11. The nucleic acid expression cassette according to any one of claims 7 to 10, wherein the promoter is a liver-specific promoter, preferably a promoter derived from the transthyretin (TTR) promoter, more preferably the minimal promotor of the transthyretin gene (TTRmin) defined by SEQ ID NO: 27.

12. The nucleic acid expression cassette according to claim 11, further comprising a liver-specific regulatory element comprising SEQ ID NO: 24, thereby comprising the combination of the TTRe and TTRm nucleic acids as defined by SEQ ID NO: 28.

13. The nucleic acid expression cassette according to any one of claims 7 to 12, further comprising a minute virus of mouse (MVM) intron, preferably the MVM intron defined by SEQ ID NO: 29.

14. The nucleic acid expression cassette according to any one of claims 7 to 13, further comprising a transcriptional termination signal derived from the bovine growth hormone polyadenylation signal (BGHpA), preferably the BGHpA defined by SEQ ID NO: 30.

15. A vector comprising the nucleic acid regulatory element according to any one of claims 1 to 6, or the nucleic acid expression cassette according to any one of claims 7 to 14, preferably a viral vector, more preferably a vector derived from an adeno-associated virus (AAV), even more preferably a self-complementary AAV vector.

16. The vector according to claim 15, having SEQ ID NO: 38, SEQ ID NO: 36 or SEQ ID NO: 44, preferably SEQ ID NO: 38.

17. A pharmaceutical composition comprising the nucleic acid expression cassette according to any one of claims 7 to 14, or the vector according to claim 15 or 16, and a pharmaceutically acceptable carrier.

18. The nucleic acid regulatory element according to any one of claims 1 to 6, the nucleic acid expression cassette according to any one of claims 7 to 14, the vector according to claim 15 or 16, or the pharmaceutical composition according to claim 17 for use in gene therapy, preferably liver-directed gene therapy, or in vaccination therapy, preferably prophylactic vaccination.

19. The nucleic acid regulatory element, the nucleic acid expression cassette, the vector, or the pharmaceutical composition for use according to claim 18, wherein the gene therapy is for the treatment of hemophilia A or hemophilia B.

20. An *in vitro* or *ex vivo* method for expressing a transgene product in liver cells comprising:

- introducing the nucleic acid expression cassette according to any one of claims 7 to 14, or the vector according to claim 15 or 16 into the liver cells;

- expressing the transgene product in the liver cells.

Figure 1

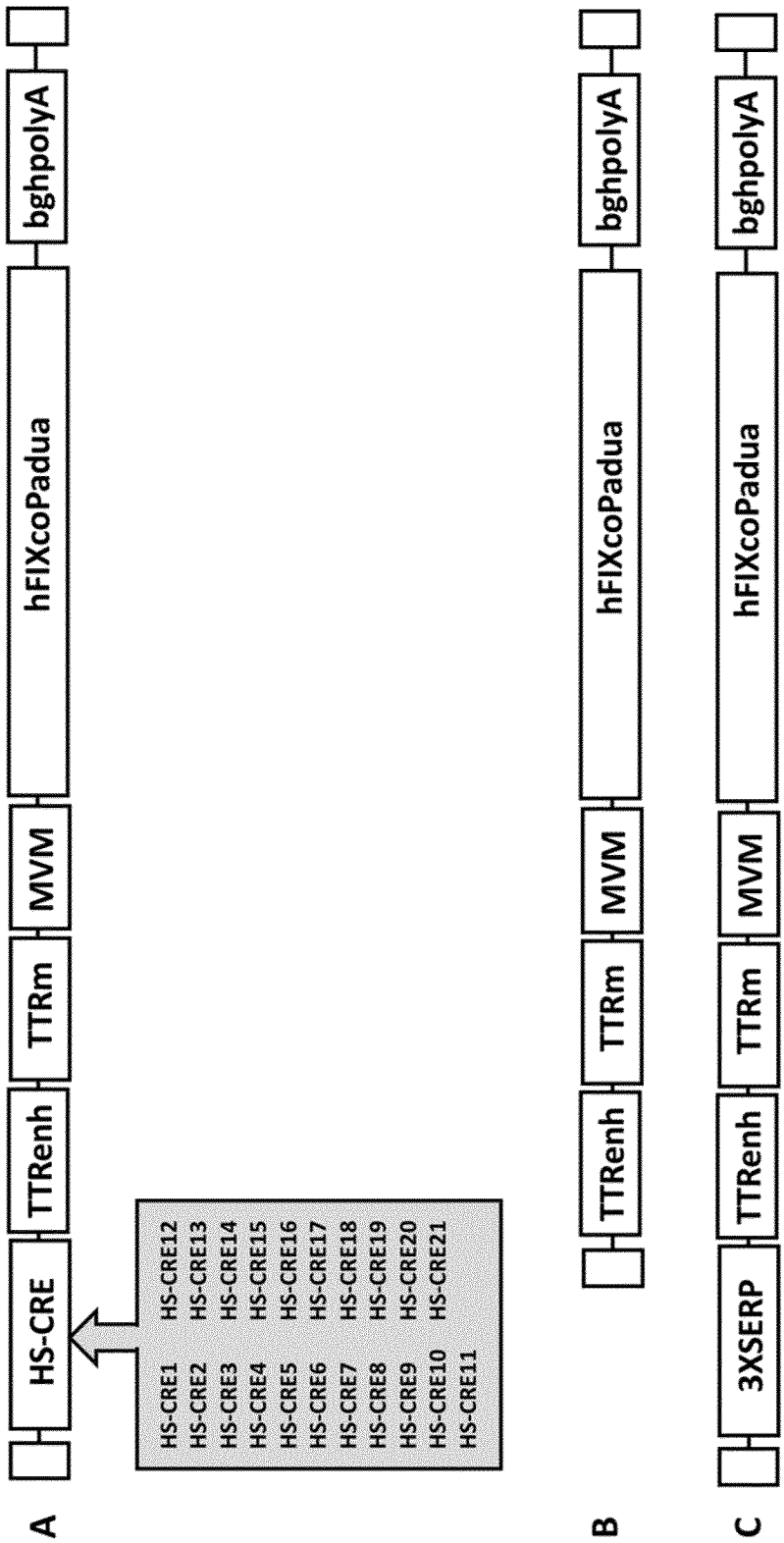

A
HS-CRE
TTRenh
TTRm
MVM
hFIXcoPadua
bghpolyA
HS-CRE1
HS-CRE2
HS-CRE3
HS-CRE4
HS-CRE5
HS-CRE6
HS-CRE7
HS-CRE8
HS-CRE9
HS-CRE10
HS-CRE11
HS-CRE12
HS-CRE13
HS-CRE14
HS-CRE15
HS-CRE16
HS-CRE17
HS-CRE18
HS-CRE19
HS-CRE20
HS-CRE21
B
TTRenh
TTRm
MVM
hFIXcoPadua
bghpolyA
C
3XSERP
TTRenh
TTRm
MVM
hFIXcoPadua
bghpolyA

Figure 2

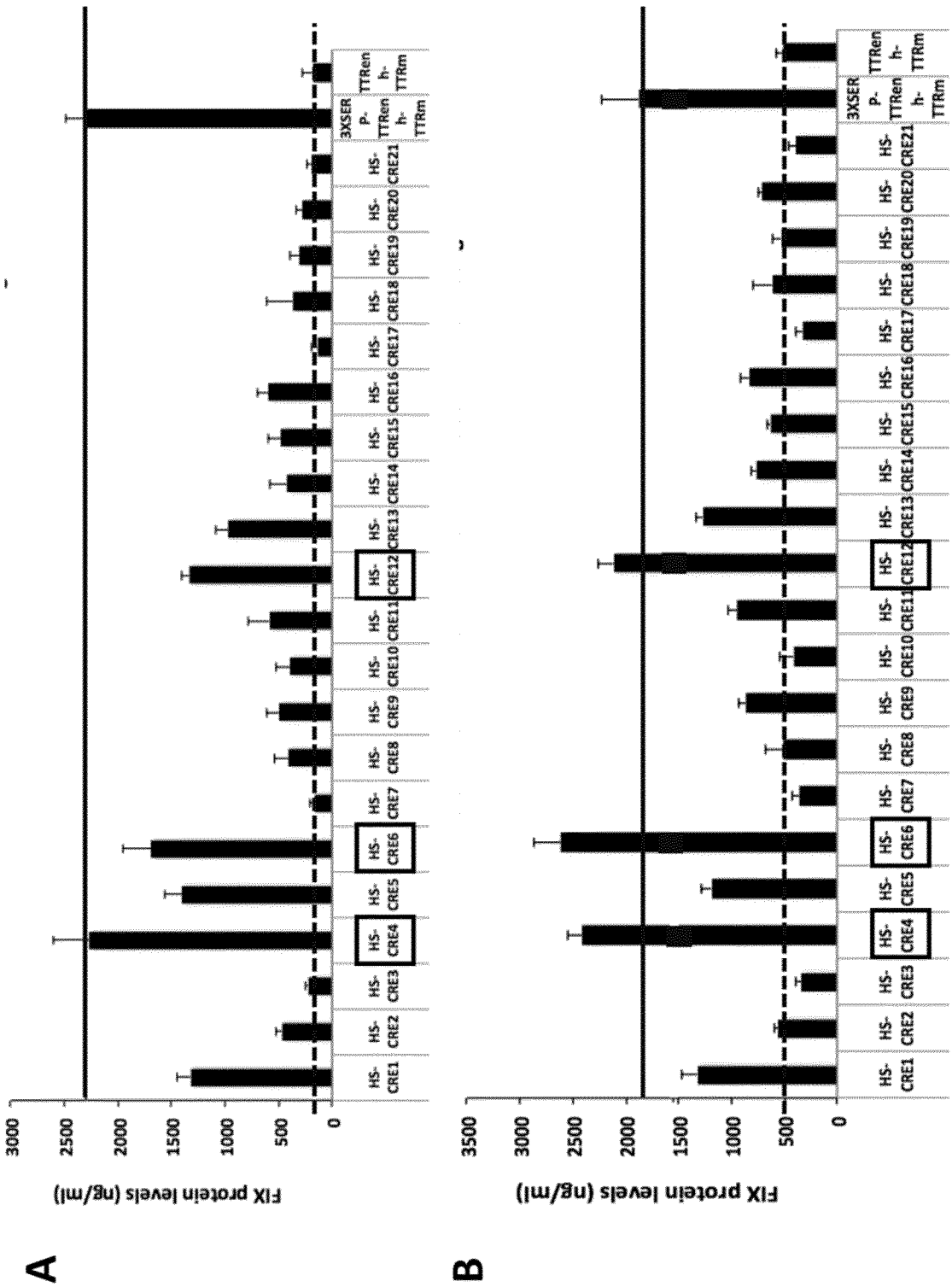
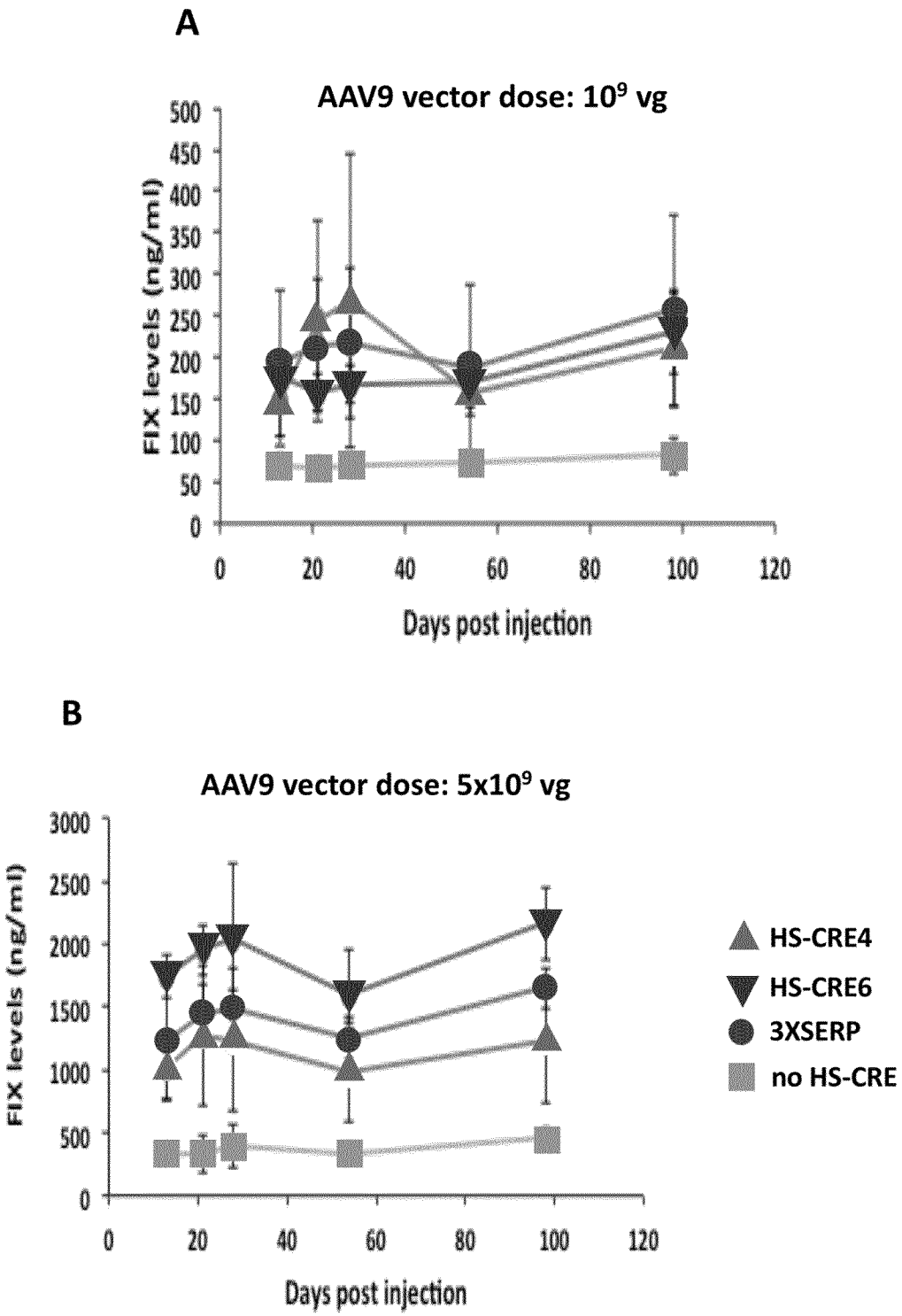

A
FIX protein levels (ng/ml)
3000
2500
2000
1500
1000
500
0
HS-CRE1
HS-CRE2
HS-CRE3
HS-CRE4
HS-CRE5
HS-CRE6
HS-CRE7
HS-CRE8
HS-CRE9
HS-CRE10
HS-CRE11
HS-CRE12
HS-CRE13
HS-CRE14
HS-CRE15
HS-CRE16
HS-CRE17
HS-CRE18
HS-CRE19
HS-CRE20
HS-CRE21
3XSER TTRen
P-TTRen
h-TTRm
B
FIX protein levels (ng/ml)
3500
3000
2500
2000
1500
1000
500
0
HS-CRE1
HS-CRE2
HS-CRE3
HS-CRE4
HS-CRE5
HS-CRE6
HS-CRE7
HS-CRE8
HS-CRE9
HS-CRE10
HS-CRE11
HS-CRE12
HS-CRE13
HS-CRE14
HS-CRE15
HS-CRE16
HS-CRE17
HS-CRE18
HS-CRE19
HS-CRE20
HS-CRE21
3XSER TTRen
P-TTRen
h-TTRm

Figure 3


A
AAV9 vector dose: 10^9 vg
FIX levels (ng/ml)
500
450
400
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
Days post injection
B
AAV9 vector dose: 5x10^9 vg
FIX levels (ng/ml)
3000
2500
2000
1500
1000
500
0
0
20
40
60
80
100
120
Days post injection
HS-CRE4
HS-CRE6
3XSERP
no HS-CRE

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2019/079358

A. CLASSIFICATION OF SUBJECT MATTER
INV. A61K48/00
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A61K

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2008/073303 A2 (SWITCHGEAR GENOMICS [US]; ALDRED SHELLEY FORCE [US] ET AL.) 19 June 2008 (2008-06-19) whole document esp. seq id nos 13943, 1095; claims 1, 10	1,2,5-20
X	WO 2013/186398 A1 (FOND TELETHON [IT]) 19 December 2013 (2013-12-19)	1,2, 5-11,15, 17-19
Y	whole document esp. seq id no 3	1,2,5-20
Y	WO 2011/034935 A2 (SWITCHGEAR GENOMICS [US]; ALDRED SHELLEY FORCE [US] ET AL.) 24 March 2011 (2011-03-24) whole document esp. seq id no 1149; [183]	1,2,5-20

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 March 2020	03/04/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Brück, Marianne

3

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2016/146757 A1 (UNIV BRUSSEL VRIJE [BE]) 22 September 2016 (2016-09-22) cited in the application whole document esp. seq id nos 4, 68,69 -----	1,2,5-20

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2019/079358

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

1, 2, 5-20(all partially)

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1, 2, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 6

   ---

2. claims: 1, 2, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 4

   ---

3. claims: 1, 2, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 12.

   ---

4. claims: 1, 3, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 1.

   ---

5. claims: 1, 3, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 5.

   ---

6. claims: 1, 4-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 9.

   ---

7. claims: 1, 4-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 11.

   ---

8. claims: 1, 3, 5-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 13.

   ---

9. claims: 1, 4-20(all partially)

   a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 14

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

---

10. claims: 1, 4-20(all partially)

a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 16

---

11. claims: 1, 4-20(all partially)

a nucleic acid regulatory element for enhancing liver-specific gene expression comprising SEQ ID NO: 20.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2019/079358

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008073303 A2	19-06-2008	EP 2097538 A2	09-09-2009
		US 2009018031 A1	15-01-2009
		WO 2008073303 A2	19-06-2008
WO 2013186398 A1	19-12-2013	NONE	
WO 2011034935 A2	24-03-2011	EP 2483408 A2	08-08-2012
		US 2011065100 A1	17-03-2011
		US 2014364336 A1	11-12-2014
		US 2017191987 A1	06-07-2017
		WO 2011034935 A2	24-03-2011
WO 2016146757 A1	22-09-2016	AU 2016232146 A1	12-10-2017
		CA 2979495 A1	22-09-2016
		DK 3270944 T3	27-01-2020
		EP 3270944 A1	24-01-2018
		JP 2018513678 A	31-05-2018
		US 2018071406 A1	15-03-2018
		WO 2016146757 A1	22-09-2016