METHODS AND COMPOSITIONS FOR REGULATION OF TRANSGENE EXPRESSION

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Application Nos. 61/537,349 filed September 21, 2011; U.S. Provisional Application 61/560,506 filed November 16, 2011; and U.S. Provisional Application 61/670,490 filed July 11, 2012, the disclosures of which are hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure is in the field of genome editing.

BACKGROUND

[0003] Gene therapy holds enormous potential for a new era of human therapeutics. These methodologies will allow treatment for conditions that have not been addressable by standard medical practice. Gene therapy can include the many variations of genome editing techniques such as disruption or correction of a gene locus, and insertion of an expressible transgene that can be controlled either by a specific exogenous promoter fused to the transgene, or by the endogenous promoter found at the site of insertion into the genome.

[0004] Delivery and insertion of the transgene are examples of hurdles that must be solved for any real implementation of this technology. For example, although a variety of gene delivery methods are potentially available for therapeutic use, all involve substantial tradeoffs between safety, durability and level of expression. Methods that provide the transgene as an episome (*e.g.* basic adenovirus, AAV and plasmid-based systems) are generally safe and can yield high initial expression levels, however, these methods lack robust episome replication, which may limit the duration of expression in mitotically active tissues. In contrast, delivery methods that result in the random integration of the desired transgene (*e.g.* integrating lentivirus) provide more durable expression but, due to the untargeted nature of the random insertion, may provoke unregulated growth in the recipient cells, potentially

leading to malignancy via activation of oncogenes in the vicinity of the randomly integrated transgene cassette. Moreover, although transgene integration avoids replication-driven loss, it does not prevent eventual silencing of the exogenous promoter fused to the transgene. Over time, such silencing results in reduced transgene expression for the majority of random insertion events. In addition, integration of a transgene rarely occurs in every target cell, which can make it difficult to achieve a high enough expression level of the transgene of interest to achieve the desired therapeutic effect.

[0005] In recent years, a new strategy for transgene integration has been developed that uses cleavage with site-specific nucleases to bias insertion into a chosen genomic locus (*see, e.g.,* co-owned US Patent 7,888,121). This approach offers the prospect of improved transgene expression, increased safety and expressional durability, as compared to classic integration approaches, since it allows exact transgene positioning for a minimal risk of gene silencing or activation of nearby oncogenes.

[0006] One approach involves the integration of a transgene into its cognate locus, for example, insertion of a wild type transgene into the endogenous locus to correct a mutant gene. Alternatively, the transgene may be inserted into a non-cognate locus chosen specifically for its beneficial properties. *See, e.g.,* U.S. Patent Publication No. 20120128635 relating to targeted insertion of a factor IX (FIX) transgene. Targeting the cognate locus can be useful if one wishes to replace expression of the endogenous gene with the transgene while still maintaining the expressional control exerted by the endogenous regulatory elements. Specific nucleases can be used that cleave within or near the endogenous locus and the transgene can be integrated at the site of cleavage through homology directed repair (HDR) or by end capture during non-homologous end joining (NHEJ). The integration process is determined by the use or non-use of regions of homology in the transgene donors between the donor and the endogenous locus.

[0007] Alternatively, the transgene may be inserted into a specific "safe harbor" location in the genome that may either utilize the promoter found at that safe harbor locus, or allow the expressional regulation of the transgene by an exogenous promoter that is fused to the transgene prior to insertion. Several such "safe harbor" loci have been described, including the AAVS1 and CCR5 genes in human cells, and

Rosa26 in murine cells (*see, e.g.,* co-owned United States patent applications nos. 20080299580; 20080159996 and 201000218264). As described above, nucleases specific for the safe harbor can be utilized such that the transgene construct is inserted by either HDR- or NHEJ- driven processes.

[0008] An especially attractive application of gene therapy involves the treatment of disorders that are either caused by an insufficiency of a secreted gene product or that are treatable by secretion of a therapeutic protein. Such disorders are potentially addressable via delivery of a therapeutic transgene to a modest number of cells, provided that each recipient cell expresses a high level of the therapeutic gene product. In such a scenario, relief from the need for gene delivery to a large number of cells can enable the successful development of gene therapies for otherwise intractable indications. Such applications would require permanent, safe, and very high levels of transgene expression. Thus the development of a safe harbor which exhibits these properties would provide substantial utility in the field of gene therapy.

[0009] A considerable number of disorders are either caused by an insufficiency of a secreted gene product or are treatable by secretion of a therapeutic protein. Clotting disorders, for example, are fairly common genetic disorders where factors in the clotting cascade are aberrant in some manner, *i.e.*, lack of expression or production of a mutant protein. Most clotting disorders result in hemophilias such as hemophilia A (factor VIII deficiency), hemophilia B (factor IX deficiency), or hemophilia C (factor XI deficiency). Treatment for these disorders is often related to the severity. For mild hemophilias, treatments can involve therapeutics designed to increase expression of the under-expressed factor, while for more severe hemophilias, therapy involves regular infusion of the missing clotting factor (often 2-3 times a week) to prevent bleeding episodes. Patients with severe hemophilia are often discouraged from participating in many types of sports and must take extra precautions to avoid everyday injuries.

[0010] Alpha-1 antitrypsin (A1AT) deficiency is an autosomal recessive disease caused by defective production of alpha 1-antitrypsin which leads to inadequate A1AT levels in the blood and lungs. It can be associated with the development of chronic obstructive pulmonary disease (COPD) and liver disorders. Currently, treatment of the diseases associated with this deficiency can involve infusion of exogenous A1AT and lung or liver transplant.

[0011] Lysosomal storage diseases (LSDs) are a group of rare metabolic monogenic diseases characterized by the lack of functional individual lysosomal proteins normally involved in the breakdown of waste lipids, glycoproteins and mucopolysaccharides. These diseases are characterized by a buildup of these compounds in the cell since it is unable to process them for recycling due to the mis-functioning of a specific enzyme. Common examples include Gaucher's (glucocerebrosidase deficiency- gene name: GBA), Fabry's (α galactosidase deficiency- GLA), Hunter's (iduronate-2-sulfatase deficiency-IDS), Hurler's (alpha-L iduronidase deficiency- IDUA), and Niemann-Pick's (sphingomyelin phosphodiesterase 1deficiency- SMPD1) diseases. When grouped together, LSDs have an incidence in the population of about 1 in 7000 births. These diseases have devastating effects on those afflicted with them. They are usually first diagnosed in babies who may have characteristic facial and body growth patterns and may have moderate to severe mental retardation. Treatment options include enzyme replacement therapy (ERT) where the missing enzyme is given to the patient, usually through intravenous injection in large doses. Such treatment is only to treat the symptoms and is not curative, thus the patient must be given repeated dosing of these proteins for the rest of their lives, and potentially may develop neutralizing antibodies to the injected protein. Often these proteins have a short serum half-life, and so the patient must also endure frequent infusions of the protein. For example, Gaucher's disease patients receiving the Cerezyme® product (imiglucerase) must have infusions three times per week. Production and purification of the enzymes is also problematic, and so the treatments are very costly (>$100,000 per year per patient).

[0012] Type I diabetes is a disorder in which immune-mediated destruction of pancreatic beta cells results in a profound deficiency of insulin, which is the primary secreted product of these cells. Restoration of baseline insulin levels provide substantial relief from many of the more serious complications of this disorder which can include "macrovascular" complications involving the large vessels: ischemic heart disease (angina and myocardial infarction), stroke and peripheral vascular disease, as well as "microvascular" complications from damage to the small blood vessels. Microvascular complications may include diabetic retinopathy, which affects blood vessel formation in the retina of the eye, and can lead to visual symptoms, reduced vision, and potentially blindness, and diabetic nephropathy, which may

involve scarring changes in the kidney tissue, loss of small or progressively larger amounts of protein in the urine, and eventually chronic kidney disease requiring dialysis. Diabetic neuropathy can cause numbness, tingling and pain in the feet and, together with vascular disease in the legs, contributes to the risk of diabetes-related foot problems (such as diabetic foot ulcers) that can be difficult to treat and occasionally require amputation as a result of associated infections.

[0013] Antibodies are secreted protein products whose binding plasticity has been exploited for development of a diverse range of therapies. Therapeutic antibodies can be used for neutralization of target proteins that directly cause disease (e.g. VEGF in macular degeneration) as well as highly selective killing of cells whose persistence and replication endanger the hose (e.g. cancer cells, as well as certain immune cells in autoimmune diseases). In such applications, therapeutic antibodies take advantage of the body's normal response to its own antibodies to achieve selective killing, neutralization, or clearance of target proteins or cells bearing the antibody's target antigen. Thus antibody therapy has been widely applied to many human conditions including oncology, rheumatology, transplant, and ocular disease Examples of antibody therapeutics include Lucentis® (Genentech) for the treatment of macular degeneration, Rituxan® (Biogen Idec) for the treatment of Non-Hodgkin lymphoma, and Herceptin® (Genentech) for the treatment of breast cancer. Albumin is a protein that is produced in the liver and secreted into the blood. In humans, serum albumin comprises 60% of the protein found in blood, and its function seems to be to regulate blood volume by regulating the colloid osmotic pressure. It also serves as a carrier for molecules with low solubility, for example lipid soluble hormones, bile salts, free fatty acids, calcium and transferrin. In addition, serum albumin carries therapeutics, including warfarin, phenobutazone, clofibrate and phenytoin. In humans, the albumin locus is highly expressed, resulting in the production of approximately 15 g of albumin protein each day. Albumin has no autocrine function, and there does not appear to be any phenotype associated with monoallelic knockouts and only mild phenotypic observations are found for biallelic knockouts (see Watkins *et al* (1994) *Proc Natl Acad Sci USA* 91:9417).

[0014] Albumin has also been used when coupled to therapeutic reagents to increase the serum half-life of the therapeutic. For example, Osborn *et al* (*J Pharm Exp Thera* (2002) 303(2):540) disclose the pharmacokinetics of a serum albumin-

interferon alpha fusion protein and demonstrate that the fusion protein had an approximate 140-fold slower clearance such that the half-life of the fusion was 18-fold longer than for the interferon alpha protein alone. Other examples of therapeutic proteins recently under development that are albumin fusions include Albulin-G™, Cardeva™ and Albugranin™ (Teva Pharmaceutical Industries, fused to Insulin, b-type natriuretic, or GCSF, respectively), Syncria® (GlaxoSmithKline, fused to Glucagon-like peptide-1) and Albuferon α-2B, fused to IFN-alpha (see *Current Opinion in Drug Discovery and Development*, (2009), vol 12, No. 2. p. 288). In these cases, Albulin-G™, Cardeva™ and Syncria® are all fusion proteins where the albumin is found on the N-terminus of the fusion, while Albugranin™ and Albuferon alpha 2G are fusions where the albumin is on the C-terminus of the fusion.

[0015] Thus, there remains a need for additional methods and compositions that can be used to express a desired transgene at a therapeutically relevant level, while avoiding any associated toxicity, and which may limit expression of the transgene to the desired tissue type, for example to treat genetic diseases such as hemophilias, diabetes, lysosomal storage diseases and A1AT deficiency. Additionally, there remains a need for additional methods and compositions to express a desired transgene at a therapeutically relevant level for the treatment of other diseases such as cancers.

SUMMARY

[0016] Disclosed herein are methods and compositions for creating a safe harbor in the genome of cells, for targeted insertion and subsequence expression of a transgene, for example expression of the transgene from a secretory tissue such as liver. In one aspect, the present invention refers to a non-naturally occurring fusion protein comprising a zinc finger protein or a TAL-effector domain protein that binds to an endogenous albumin gene and a cleavage domain, wherein the fusion protein modifies the endogenous albumin gene, and wherein the zinc finger protein comprises 4, 5 or 6 zinc finger domains comprising a recognition helix region, wherein the zinc finger proteins comprise the recognition helix regions shown in a single row of Table 1, Table 3, Table 5 or Table 8 or the TAL-effector domain protein that binds to a target site as shown in Table 12. Also described herein is a non-naturally occurring zinc-finger protein (ZFP) that binds to target site in a region of interest (*e.g.*, an albumin gene) in a genome, wherein the ZFP comprises one or more engineered zinc-

finger binding domains. In one example, the ZFP is a zinc-finger nuclease (ZFN) that cleaves a target genomic region of interest, wherein the ZFN comprises one or more engineered zinc-finger binding domains and a nuclease cleavage domain or cleavage half-domain. Cleavage domains and cleavage half domains can be obtained, for example, from various restriction endonucleases and/or homing endonucleases. In one example, the cleavage half-domains are derived from a Type IIS restriction

endonuclease *(e.g., Fok* I). In certain examples, the zinc finger domain recognizes a target site in an albumin gene, for example a zinc finger protein with the recognition helix domains ordered as shown in a single row of Tables 1, 3, 5 or 8. .

[0017] In another example, described herein is a Transcription Activator Like Effector (TALE) protein that binds to target site in a region of interest (*e.g.,* an albumin gene) in a genome, wherein the TALE comprises one or more engineered TALE binding domains. In one example, the TALE is a nuclease (TALEN) that cleaves a target genomic region of interest, wherein the TALEN comprises one or more engineered TALE DNA binding domains and a nuclease cleavage domain or cleavage half-domain. Cleavage domains and cleavage half domains can be obtained, for example, from various restriction endonucleases and/or homing endonucleases. In one example, the cleavage half-domains are derived from a Type IIS restriction endonuclease *(e.g., Fok* I). In certain examples, the TALE DNA binding domain recognizes a target site in an albumin gene, for example TALE DNA binding domain having the target sequence shown in a single row of Table 12.

[0018] The ZFN and/or TALEN as described herein may bind to and/or cleave the region of interest in a coding or non-coding region within or adjacent to the gene, such as, for example, a leader sequence, trailer sequence or intron, or within a non-transcribed region, either upstream or downstream of the coding region. In certain examples, the ZFN binds to and/or cleaves an albumin gene. In other examples, the ZFN and/or TALEN binds to and/or cleaves a safe-harbor gene, for example a CCR5 gene, a PPP1R12C (also known as AAVS1) gene or a *Rosa* gene. *See, e.g.,* U.S. Patent Publication Nos. 20080299580; 20080159996 and 201000218264. In another example, described herein are compositions comprising one or more of the zinc-finger and/or TALE nucleases described herein. In certain examples, the composition comprises one or more zinc-finger and/or TALE nucleases in combination with a pharmaceutically acceptable excipient.

[0019] In another aspect, the present invention refers to a polynucleotide encoding one or more fusion proteins of the present invention. Also disclosed herein is a polynucleotide encoding one or more ZFNs and/or TALENs described herein. The polynucleotide may be, for example, mRNA. In some examples, the mRNA may be chemically modified (See *e.g.* Kormann *et al,* (2011) *Nature Biotechnology* 29(2):154-157).

[0020] In another example, described herein is a ZFN and/or TALEN expression vector comprising a polynucleotide, encoding one or more ZFNs and/or TALENs

described herein, operably linked to a promoter. In one example, the expression vector is a viral vector. In one example, the viral vector exhibits tissue specific tropism.

[0021] In another aspect, the present invention refers to an isolated cell comprising one or more fusion proteins according to the present invention or one or more polynucleotides according to the present invention. Also, described herein is a host cell comprising one or more ZFN and/or TALEN expression vectors. The host cell may be stably transformed or transiently transfected or a combination thereof with one or more ZFP or TALEN expression vectors. In one example, the host cell is an embryonic stem cell. In other examples, the one or more ZFP and/or TALEN expression vectors express one or more ZFNs and/or TALENs in the host cell. In another example, the host cell may further comprise an exogenous polynucleotide donor sequence. Non-limiting examples of suitable host cells include eukaryotic cells or cell lines such as secretory cells (*e.g.*, liver cells, mucosal cells, salivary gland cells, pituitary cells, etc.), blood cells (red blood cells), stem cells, etc. In any of the examples described herein the host cell can comprise an embryo cell, for example, of a mouse, rat, rabbit or other mammal cell embryo.

[0022] In another aspect, the present invention refers to *in vitro* method of cleaving an endogenous albumin gene in a cell, the method comprising introducing, into the cell, one or more expression vectors comprising at least one fusion protein according to the present invention or at least one polynucleotide according to the present invention, under conditions such that the one or more fusion proteins are expressed and the albumin gene is cleaved. Also, described herein is a method for cleaving an albumin gene in a cell, the method comprising: introducing, into the cell, one or more polynucleotides encoding one or more ZFNs and/or TALENs that bind to a target site in the one or more albumin genes under conditions such that the ZFN(s) is (are) or TALENs is (are) expressed and the one or more albumin genes are cleaved.

[0023] In other examples, a genomic sequence in any target gene is replaced, for example using a ZFN or TALEN (or vector encoding said ZFN or TALEN) as described herein and a "donor" sequence (*e.g.*, transgene) that is inserted into the gene following targeted cleavage with the ZFN and/or TALEN. The donor sequence may be present in the ZFN or TALEN vector, present in a separate vector (*e.g.*, Ad or LV vector) or, alternatively, may be introduced into the cell using a

different nucleic acid delivery mechanism. Such insertion of a donor nucleotide sequence into the target locus (*e.g.*, albumin gene, other safe-harbor gene, etc.) results in the expression of the transgene carried by the donor under control of the target locus's (*e.g.* albumin) genetic control elements. In some examples, insertion of the transgene of interest, for example into an albumin gene results in expression of an intact exogenous protein sequence and lacks any albumin encoded amino acids. In other examples, the expressed exogenous protein is a fusion protein and comprises amino acids encoded by the transgene and by an albumin gene (*e.g.*, from the

endogenous target locus or, alternatively from albumin-encoding sequences on the transgene). In some instances, the albumin sequences will be present on the amino (N)-terminal portion of the exogenous protein, while in others, the albumin sequences will be present on the carboxy (C)- terminal portion of the exogenous protein. In other instances, albumin sequences will be present on both the N- and C-terminal portions of the exogenous protein. The albumin sequences may include full-length wild-type or mutant albumin sequences or, alternatively, may include partial albumin amino acid sequences. In certain examples, the albumin sequences (full-length or partial) serve to increase the serum half-life of the polypeptide expressed by the transgene to which it is fused and/or as a carrier. In some examples, the albumin-transgene fusion is located at the endogenous locus within the cell while in other embodiments, the albumin-transgene coding sequence is inserted into a safe harbor within a genome. In some examples, the safe harbor is selected from the AAVS1, Rosa, HPRT or CCR5 locus (see co-owned US patent publications Nos. 20080299580; 20080159996 and 201000218264, and US Provisional patent application No. 61/556,691).

[0023a] In one aspect, the present invention refers to a non-naturally occurring fusion protein of the present invention or a polynucleotide of the present invention for use in therapy. In another aspect, the present invention refers to a non-naturally occurring fusion protein of the present invention or a polynucleotide of the present invention, and a donor polynucleotide comprising a transgene encoding a therapeutic protein for use in therapy. In yet another aspect, the present invention refers to use of one or more expression vectors comprising at least one fusion protein according to the present invention or at least one polynucleotide according to the present invention in the manufacture of a medicament for treating a lysosomal storage disease, epidermolysis bullosa, AAT deficient emphysema or a blood disorder, wherein the vector or the polynucleotide is introduced into the cell under conditions such that the one or more fusion proteins are expressed and an endogenous albumin gene is cleaved.

In another example, the invention disclosure methods and compositions that can be used to express a transgene under the control of an albumin promoter *in vivo* (*e.g.,* endogenous or exogenous albumin promoter). In some examples, the transgene may encode a therapeutic protein of interest. The transgene may encode a protein such that the methods disclosed herein can be used for

production of protein that is deficient or lacking (*e.g.*, "protein replacement"). In some instances, the protein may be involved treatment for a lysosomal storage disease. Other therapeutic proteins may be expressed, including protein therapeutics for conditions as diverse as epidermolysis bullosa or AAT deficient emphysema. In other examples, the transgene may comprise sequences (*e.g.*, engineered sequences) such that the expressed protein has characteristics which give it novel and desirable features (increased half-life, changed plasma clearance characteristics etc.). Engineered sequences can also include amino acids derived from the albumin sequence. In some examples, the transgenes encode therapeutic proteins, therapeutic hormones, plasma proteins, antibodies and the like. In some examples, the transgenes may encode proteins involved in blood disorders such as clotting disorders. In some examples, the transgenes encode structural nucleic acids (shRNAs, miRNAs and the like).

[0025] In some examples, the methods disclosed herein may be used *in vivo* in transgenic animal systems. In some examples, the transgenic animal may be used in model development where the transgene encodes a human gene. In some instances, the transgenic animal may be knocked out at the corresponding endogenous locus, allowing the development of an *in vivo* system where the human protein may be studied in isolation. Such transgenic models may be used for screening purposes to identify small molecule, large biomolecules or other entities which may interact or modify the human protein of interest. In other examples, the transgenic animals may be used for production purposes, for example, to produce antibodies or other biomolecules of interest. In certain examples, the animal is a small mammal, for example a dog, rabbit or a rodent such as rat, a mouse or a guinea pig. In other examples, the animal is a non-human primate. In yet further embodiments, the animal is a farm animal such as a cow, goat or pig. In some examples, the transgene is integrated into the selected locus (*e.g.,* albumin or safe-harbor) into a stem cell (*e.g.,* an embryonic stem cell, an induced pluripotent stem cell, a hepatic stem cell, etc.) or animal embryo obtained by any of the methods described herein, and then the embryo is implanted such that a live animal is born. The animal is then raised to sexual maturity and allowed to produce offspring wherein at least some of the offspring comprise the integrated transgene.

[0026] In a still further example, provided herein is a method for site specific integration of a nucleic acid sequence into an endogenous locus (*e.g.,* albumin gene) of a chromosome, for example into the chromosome of an embryo. In certain examples, the method comprises: (a) injecting an embryo with (i) at least one DNA vector, wherein the DNA vector comprises an upstream sequence and a downstream sequence flanking the nucleic acid sequence to be integrated, and (ii) at least one RNA molecule encoding a zinc finger and/or TALE nuclease that recognizes the site of integration in the target locus (*e.g.,* albumin locus), and (b) culturing the embryo to allow expression of the zinc finger and/or TALE nuclease, wherein a double stranded break introduced into the site of integration by the zinc finger nuclease or TALEN is repaired, via homologous recombination with the DNA vector, so as to integrate the nucleic acid sequence into the chromosome.

[0027] Suitable embryos may be derived from several different vertebrate species, including mammalian, bird, reptile, amphibian, and fish species. Generally

speaking, a suitable embryo is an embryo that may be collected, injected, and cultured to allow the expression of a zinc finger or TALE nuclease. In some embodiments, suitable embryos may include embryos from small mammals (*e.g.,* rodents, rabbits, etc.), companion animals, livestock, and primates. Non-limiting examples of rodents may include mice, rats, hamsters, gerbils, and guinea pigs. Non-limiting examples of companion animals may include cats, dogs, rabbits, hedgehogs, and ferrets. Non-limiting examples of livestock may include horses, goats, sheep, swine, llamas, alpacas, and cattle. Non-limiting examples of primates may include capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. In other embodiments, suitable embryos may include embryos from fish, reptiles, amphibians, or birds. Alternatively, suitable embryos may be insect embryos, for instance, a *Drosophila* embryo or a mosquito embryo.

[0028] In any of the methods or compositions described herein, the cell containing the engineered locus (*e.g.,* albumin locus) can be a stem cell. Specific stem cell types that may be used with the methods and compositions disclosed herein include embryonic stem cells (ESC), induced pluripotent stem cells (iPSC) and hepatic or liver stem cells. The iPSCs can be derived from patient samples and from normal controls wherein the patient derived iPSC can be mutated to normal gene sequence at the gene of interest, or normal cells can be altered to the known disease allele at the gene of interest. Similarly, the hepatic stem cells can be isolated from a patient. These cells are then engineered to express the transgene of interest, expanded and then reintroduced into the patient.

[0029] In any of the methods described herein, the polynucleotide encoding the zinc finger nuclease(s) and/or TALEN(s) can comprise DNA, RNA or combinations thereof. In certain examples, the polynucleotide comprises a plasmid. In other examples, the polynucleotide encoding the nuclease comprises mRNA.

[0030] Also described is an embryo comprising at least one DNA vector, wherein the DNA vector comprises an upstream sequence and a downstream sequence flanking the nucleic acid sequence to be integrated, and at least one RNA molecule encoding a zinc finger nuclease that recognizes the chromosomal site of integration. Organisms derived from any of the embryos as described herein are also

provided (*e.g.,* embryos that are allowed to develop to sexual maturity and produce progeny).

[0031] In another example provided by the methods and compositions disclosed herein is the use of cells, cell lines and animals (*e.g.,* transgenic animals) in the screening of drug libraries and/or other therapeutic compositions (*i.e.*, antibodies, structural RNAs, etc.) for use in treatment of an animal afflicted with a clotting factor disorder. Such screens can begin at the cellular level with manipulated cell lines or primary cells, and can progress up to the level of treatment of a whole animal (*e.g.,* human).

[0032] A kit, comprising the ZFPs and/or TALENs of the invention, is also disclosed. In one aspect, the present invention refers to a kit comprising a fusion protein according to the present invention of a polynucleotide according to the present invention. The kit may comprise nucleic acids encoding the ZFPs or TALENs, (*e.g.* RNA molecules or ZFP or TALEN encoding genes contained in a suitable expression vector), donor molecules, suitable host cell lines, instructions for performing the methods of the invention, and the like.

[0033] Thus, the disclosure herein includes, but is not limited to, the following examples:

[0034] 1. A non-naturally occurring fusion protein comprising a DNA-binding protein that binds to an endogenous albumin gene and a cleavage domain, wherein the fusion protein modifies the endogenous albumin gene.

[0035] 2. The fusion protein of example 1, wherein the DNA-binding protein comprises a zinc finger protein.

[0036] 3. The fusion protein of example 2, wherein the zinc finger protein comprises 4, 5 or 6 zinc finger domains comprising a recognition helix region, wherein the zinc finger proteins comprise the recognition helix regions shown in a single row of Table 1, Table 3, Table 5 or Table 8.

[0037] 4. The fusion protein of example 1, wherein the DNA-binding protein comprises a TALE DNA-binding domain.

[0038] 5. The fusion protein of example 4, wherein the TALE DNA-binding domain binds to a target sequence shown in a single row of Table 12.

[0039] 6. A polynucleotide encoding one or more fusion proteins of examples 1 to 5.

[0040] 7. An isolated cell comprising one or more fusion proteins according to examples 1 to 5 or one or more polynucleotides according to example 6.

[0041] 8. The cell of example 7, wherein the cell is a stem cell or an embryo cell.

[0042] 9. The cell of example 8, wherein the stem cell is selected from the group consisting of an embryonic stem cell (ESC), an induced pluripotent stem cell (iPSC), a hepatic stem cell and a liver stem cell.

[0043] 10. A kit comprising a fusion protein according to examples 1 to 5 or a polynucleotide according to example 6 or a cell according to example 7-9.

[0044] 11. A method of cleaving an endogenous albumin gene in a cell, the method comprising:

introducing, into the cell, one or more expression vectors comprising at least one polynucleotide according to example 6, under conditions such that the one or more fusion proteins are expressed and the albumin gene is cleaved.

[0045] 12. The method of example 11, wherein the polynucleotide comprises an AAV vector.

[0046] 13. The method of example 11, wherein the cell is a liver cell.

[0047] 14. A method of introducing a transgene into an endogenous albumin gene, the method comprising:

cleaving the endogenous albumin gene according to the method of any of examples 15-17 in the presence of an exogenous polynucleotide comprising the transgene such that the transgene is integrated into the endogenous albumin gene.

[0048] 15. The method of example 14, wherein the transgene expresses a therapeutic protein.

[0049] 16. The method of example 15, wherein the therapeutic protein is involved in treating a lysosomal storage disease, epidermolysis bullosa, AAT deficient emphysema or blood disorders such as clotting disorders.

[0050] 17. The method of examples 15 or 16, wherein expression of the transgene is driven by the endogenous albumin control sequences.

[0051] 18. The method of any of examples 15-17, wherein the transgene further comprises albumin sequences.

[0052] 19. The method of example 18, wherein the albumin sequences are present on the amino (N)-terminal and/or carboxy (C)-terminal portion of the protein.

[0053] 20. A method of increasing the serum half-life of a polypeptide expressed from a transgene integrated into an endogenous albumin gene, the method comprising introducing the transgene into the endogenous albumin gene according to the method of example 18 or example 19, wherein the transgene expresses the polypeptide and albumin sequences such that the serum half-life of the polypeptide in increased.

[0054] 21. A method of treating a subject having a disease caused by a deficiency of a polypeptide, the method comprising,

introducing a transgene encoding the polypeptide into an isolated cell according to the method of examples 14-19 such that the transgene is expressed in the isolated cell; and introducing the isolated cell into the subject, thereby treating the disease.

[0055] 22. The method of example 21, wherein the cell is a liver cell or a stem cell.

[0056] 23. The cell of example22, wherein the stem cell is selected from the group consisting of an embryonic stem cell (ESC), an induced pluripotent stem cell (iPSC), a hepatic stem cell and a liver stem cell.

[0057] These and other aspects will be readily apparent to the skilled artisan in light of disclosure as a whole.

BRIEF DESCRIPTION OF THE DRAWINGS

[0058] **Figure 1, panels A and B,** are gels depicting the results of a Cel-I mismatch assay (Surveyor™, Transgenomic) that quantifies the degree to which ZFN cleavage of an endogenous chromosomal target, followed by imperfect repair via NHEJ, has yielded small insertions or deletions ("indels") of the targeted locus. For a description of the assay see Horton *et al. Methods Mol Biol*. (2010) 649:247-56. Figure 1A shows results using expression constructs for ZFNs targeted to the mouse albumin gene which were transfected into Neuro2A cells, where the cells were treated for 3 days at 37°C following transfection, and then analyzed for the fraction of modified target sites via Cel-I analysis. Figure 1B shows results for the same ZFNs and cells as Figure 1A except cells were subjected to hypothermic shock (30°C) during their 3 days of growth following transfection. The percent mismatch, or % indels shown at the bottom on the lanes, is a measure of the ZFN activity and

demonstrates that the mouse albumin specific ZFNs are able to induce up to 53% indels following cleavage of their endogenous chromosomal target in Neuro 2A cells.

[0059] **Figure 2, panels A and B,** are gels depicting the results of a Cel-I mismatch assay carried out on canine D17 cells transfected with constructs expressing the canine albumin specific ZFN pair SBS 33115/SBS34077 at two concentrations of plasmid DNA, 20 or 40 ng. Figure 2A depicts the results after 3 days while Figure 2B depicts the results after 10 days. This ZFN pair was able to induce indels in ~25-30% of target site sequences at day 3.

[0060] **Figure 3, panels A and B,** show alignments of the albumin genes from a variety of species of interest. Figure 3A shows an alignment of exon 1 and the 5' region of intron 1 of human (*H. sapiens*, SEQ ID NO:160), rhesus macaque monkeys (*M. mulatta,* SEQ ID NO:73), marmoset (*C. jacchus*, SEQ ID NO:74), dog (*C. familiaris,* SEQ ID NO:75), rat (*R. norvegicus,* SEQ ID NO:75) and mouse (*M. musculus,* SEQ ID NO:76). 3B shows an alignment of the remainder of intron 1 and a small fragment of exon 2. This region includes the Locus 1 to Locus 5 of human (SEQ ID NO:161), rhesus macaque monkeys (SEQ ID NO:77), marmoset (SEQ ID NO:78), dog (SEQ ID NO:79), rat (SEQ ID NO:80) and mouse (*M. musculus,* SEQ ID NO:81) which are loci in the albumin gene chosen for ZFN targeting. The sequences depicted show the starting codon ATG (large box in Figure 3A) and the boundaries of exon1 and intron 1 (Figure 3A) and intron 1 and exon 2 (Figure 3B).

[0061] **Figure 4, panels A and B,** depict the results of a Cel-I mismatch assay carried out on genomic DNA from liver tissue biopsied from mice injected with albumin-specific ZFNs expressed from a hepatotrophic AAV8 vector. The results are from 10 wild type mice (numbers 273- 282) injected intravenously via tail vein injection with two sets of ZFN pairs (pair 1: SBS30724 and SBS30725 and pair 2: SBS30872 and SBS30873). Figure 4A is a gel that quantifies the indels present in the amplicon encompassing the pair 1 site and Figure 4B is another gel that quantifies the indels present in the amplicon that encompasses the pair 2 site. The percent of albumin genes bearing ZFN-induced modifications in the liver biopsies is indicated at the bottom of the lanes, and demonstrates that the albumin ZFN pairs are capable of modifying up to 17% of targets when the nucleases are delivered *in vivo*.

[0062] **Figure 5, panels A and B,** show the results of a Cel-I mismatch assay carried out on genomic DNA from liver tissue biopsied from mice injected with

albumin-specific ZFNs expressed from different chimeric AAV vectors. Experimental details are provided in Example 5. Figure 5A demonstrates that the ZFNs are able to cleave the albumin target in the liver *in vivo* when introduced into the animal via AAV-mediated gene delivery. The percent of albumin genes bearing ZFN-induced modifications in the liver biopsies ranged up to 16 percent. Figure 5B shows a Western blot of liver tissue using either anti-Flag antibodies or anti-p65. The open reading frames encoding the ZFNs were fused to a sequence encoding a polypeptide FLAG-tag. Thus, the anti-Flag antibody detected the ZFNs and demonstrated ZFN expression in the mice receiving ZFNs. The anti-p65 antibody served as a loading control in these experiments and indicated that comparable amounts of protein were loaded in each lane.

[0063] **Figure 6** shows results from a mouse study in which groups of mice were treated with the mouse albumin specific ZFN pair 30724/30725 via delivery of differing doses of different AAV serotypes, and then assessed for gene modification using the Cel-I assay. The AAV serotypes tested in this study were AAV2/5, AAV2/6, AAV2/8.2 and AAV2/8 (see text for details). The dose levels ranged from 5e10 to 1e12 viral genomes, and three mice were injected per group. Viral genomes present per diploid cell were also calculated and are indicated at the bottom of each lane. The percent indels induced by each treatment is also indicated below each lane and demonstrates that this ZFN pair is capable of cleaving the albumin locus. Control mice were injected with phosphate buffered saline. A non-specific band is also indicated in the figure.

[0064] **Figure 7** is a graph depicting the expression of human factor IX (F.IX) from a transgene inserted into the mouse albumin locus *in vivo*. A human F.IX donor transgene was inserted into either the mouse albumin locus at intron 1 or intron 12 following cleavage with mouse albumin-specific ZFNs in wild type mice. The graph shows expression levels of F.IX over a period of 8 weeks following injection of the vectors. ZFN pairs targeting either intron 1 or intron 12 of mouse albumin were used in this experiment, as well as ZFNs targeted to a human gene as a control. The donor F.IX gene was designed to be used following insertion into intron 1, and thus is not expressed properly when inserted into intron 12. The human F.IX transgene is expressed at a robust level for at least 8 weeks following insertion into the mouse albumin intron 1 locus.

[0065] **Figure 8, panels A and B**, are graphs depicting the expression and functionality of the human F.IX gene in the plasma of hemophilic mice following ZFN-induced F.IX transgene insertion. The experiment described in Figure 7 was repeated in hemophilic mice using the albumin intron 1 specific ZFNs and the human F.IX donor. Two weeks following treatment, expression level in the serum (Figure 8A) and clotting time (Figure 8B) were analyzed. The expression of the human F.IX transgene in hemophilic mice was able to restore clotting time to that of normal mice.

[0066] **Figure 9** (SEQ ID NO:82) provides a segment of the human albumin gene sequence encompassing parts of exon 1 and intron 1. Horizontal bars over the sequence indicate the target sites of the zinc finger nucleases.

[0067] **Figure 10** shows an alignment of a segment of the albumin genes in intron 1 from a variety of primate species including human, *H. sapiens* (SEQ ID NO:154), cynologous monkey variants (where sequences 'C' and 'S' derive from two different genome sequence sources): *M. fascicularis*_c (SEQ ID NO:155) and *M. fascicularis*_s (SEQ ID NO:156) and rhesus, *M mulatta* (SEQ ID NO:157). The figure depicts the DNA target locations of the albumin specific TALENs (indicated by the horizontal bars above the sequence).

[0068] **Figure 11, panels A to C,** show the results of a Cel-I assay carried out on genomic DNA isolated from HepG2 cells treated with TALENs or ZFNs targeted to human albumin (Figures 11A and B) and NHEJ activity of TALENs with different gap spacings (Figure 11C). The nucleases were introduced into HepG2 cells via transient plasmid transfection and quantified 3 days later for target modification via the Cel-I assay. Two variations of the TALE DNA binding domain were used, which differed in the location of their C-terminal truncation points, the +17 version and the +63 version (see text). Pairs used are described in Table 10. In addition, three ZFN pairs were also tested and the % indels detected by the Cel 1 assay is indicated at the bottom of the lanes. Figure 11C is a graph depicting NHEJ activity in terms of the gap spacing (bp) between TALEN binding sites.

[0069] **Figure 12, panels A, B and C** depict the results of ZFN pairs directed to the rhesus macaque albumin locus. Figure 12A shows the percent of NHEJ activity for the 35396/36806 pair in comparison with the 35396/36797 pair, tested in RF/6A cells in 3 independent experiments all done using a ZFN concentration of 400 ng. Figure 12B depicts a dose titration for the two pairs, from 50 ng of each ZFN to 400

ng where the samples were analyzed at day 3 following transduction. The lower half of Figure 12B depicts another experiment comparing the two pairs at day 3 or day 10 using 400 ng of ZFN. Figure 12C depicts the results of the SELEX analysis (done at 100 mM salt concentration) of the three ZFNs that were being compared where the size of the bar above the middle line shows the results for that position that were expected (*i.e.*, a single bar with a value of 1.0 above the line would mean that every base at that position analyzed in the SELEX analysis was the expected base), while bars below the line indicate the presence of non-expected bases. Bars that are divided indicate the relative contributions of other bases.

[0070] **Figure 13, panels A and B,** demonstrate the insertion of a huGLa transgene donor (deficient in patients afflicted with Fabry's disease) into the albumin locus in mice. Figure 13A shows a Western blot against the huGLa protein encoded by the transgene, where the arrow indicates the presumed protein. Comparison of the mouse samples from those mice that received both ZFN and donor (samples 1-1, 1-2 and 1-3) with the samples that either received only ZFN (4-1, 4-2, 4-3) or those that only received the huGLa donor ("hu Fabry donor"), samples 5-1 and 5-2 leads to identification of a band that coincides with the human liver lysate control. Figure 13B depicts ELISA results using a huGLa specific ELISA kit, where samples were analyzed from mice either 14 or 30 days following virus introduction (see text for details). Error bars represent standard deviations (n=3). The results demonstrate that the mice that received both the ZFN and donor had higher amounts of huGLa signal that those that only received ZFN or only received donor.

DETAILED DESCRIPTION

[0071] Disclosed herein are compositions and methods for modifying an endogenous albumin gene, for example, for expressing a transgene in a secretory tissue. In some embodiments, the transgene is inserted into an endogenous albumin gene to allow for very high expression levels that are moreover limited to hepatic tissue. The transgene can encode any protein or peptide including those providing therapeutic benefit.

[0072] Thus, the methods and compositions of the invention can be used to express therapeutically beneficial proteins (from a transgene) from highly expressed loci in secretory tissues. For example, the transgene can encode a protein involved in

disorders of the blood, for example, clotting disorders, and a variety of other monogenic diseases. In some embodiments, the transgene can be inserted into the endogenous albumin locus such that expression of the transgene is controlled by the albumin expressional control elements, resulting in liver-specific expression of the transgene encoded protein at high concentrations. Proteins that may be expressed may include clotting factors such as Factor VII, Factor VIII, Factor IX, Factor X, Factor XI, Factor XIII, vWF and the like, antibodies, proteins relevant to lyososomal storage, insulin, alpha 1-antitrypsin, and indeed any peptide or protein that when so expressed provides benefit.

[0073] In addition, any transgene can be introduced into patient derived cells, *e.g.* patient derived induced pluripotent stem cells (iPSCs) or other types of stem cells (embryonic, hematopoietic, neural, or mesenchymal as a non-limiting set) for use in eventual implantation into secretory tissues. The transgene can be introduced into any region of interest in these cells, including, but not limited to, into an albumin gene or a safe harbor gene. These altered stem cells can be differentiated for example, into hepatocytes and implanted into the liver. Alternately, the transgene can be directed to the secretory tissue as desired through the use of viral or other delivery systems that target specific tissues. For example, use of the liver-trophic adenovirus associated virus (AAV) vector AAV8 as a delivery vehicle can result in the integration of the transgene at the desired locus when specific nucleases are co-delivered with the transgene.

General

[0074] Practice of the methods, as well as preparation and use of the compositions disclosed herein employ, unless otherwise indicated, conventional techniques in molecular biology, biochemistry, chromatin structure and analysis, computational chemistry, cell culture, recombinant DNA and related fields as are within the skill of the art. These techniques are fully explained in the literature. *See*, for example, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third

edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

Definitions

[0075] The terms "nucleic acid," "polynucleotide," and "oligonucleotide" are used interchangeably and refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogues of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (*e.g.*, phosphorothioate backbones). In general, an analogue of a particular nucleotide has the same base-pairing specificity; *i.e.*, an analogue of A will base-pair with T.

[0076] The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues or modified derivatives of corresponding naturally-occurring amino acids.

[0077] "Binding" refers to a sequence-specific, non-covalent interaction between macromolecules (*e.g.*, between a protein and a nucleic acid). Not all components of a binding interaction need be sequence-specific (*e.g.*, contacts with phosphate residues in a DNA backbone), as long as the interaction as a whole is sequence-specific. Such interactions are generally characterized by a dissociation constant (K_d) of 10^{-6} M^{-1} or lower. "Affinity" refers to the strength of binding: increased binding affinity being correlated with a lower K_d.

[0078] A "binding protein" is a protein that is able to bind non-covalently to another molecule. A binding protein can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein-binding protein, it can bind to itself (to form homodimers, homotrimers, *etc.*) and/or it can bind to one or more molecules of a different protein or proteins. A binding protein can have more than one type of binding activity. For example, zinc finger proteins have DNA-binding, RNA-binding and protein-binding activity.

[0079] A "zinc finger DNA binding protein" (or binding domain) is a protein, or a domain within a larger protein, that binds DNA in a sequence-specific manner through one or more zinc fingers, which are regions of amino acid sequence within the binding domain whose structure is stabilized through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

[0080] A "TALE DNA binding domain" or "TALE" is a polypeptide comprising one or more TALE repeat domains/units. The repeat domains are involved in binding of the TALE to its cognate target DNA sequence. A single "repeat unit" (also referred to as a "repeat") is typically 33-35 amino acids in length and exhibits at least some sequence homology with other TALE repeat sequences within a naturally occurring TALE protein. *See, e.g.,* U.S. Patent Publication No. 20110301073, incorporated by reference herein in its entirety.

[0081] Zinc finger and TALE binding domains can be "engineered" to bind to a predetermined nucleotide sequence, for example via engineering (altering one or more amino acids) of the recognition helix region of a naturally occurring zinc finger or TALE protein. Therefore, engineered DNA binding proteins (zinc fingers or TALEs) are proteins that are non-naturally occurring. Non-limiting examples of methods for engineering DNA-binding proteins are design and selection. A designed DNA binding protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP and/or TALE designs and binding data. See, for example, US Patents 6,140,081; 6,453,242; and 6,534,261; see also WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496 and U.S. Publication No. 20110301073.

[0082] A "selected" zinc finger protein or TALE is a protein not found in nature whose production results primarily from an empirical process such as phage display, interaction trap or hybrid selection. See *e.g.*, US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197; WO 02/099084 and U.S. Publication No. 20110301073.

[0083] "Recombination" refers to a process of exchange of genetic information between two polynucleotides. For the purposes of this disclosure,

"homologous recombination (HR)" refers to the specialized form of such exchange that takes place, for example, during repair of double-strand breaks in cells via homology-directed repair mechanisms. This process requires nucleotide sequence homology, uses a "donor" molecule to template repair of a "target" molecule (*i.e.*, the one that experienced the double-strand break), and is variously known as "non-crossover gene conversion" or "short tract gene conversion," because it leads to the transfer of genetic information from the donor to the target. Without wishing to be bound by any particular theory, such transfer can involve mismatch correction of heteroduplex DNA that forms between the broken target and the donor, and/or "synthesis-dependent strand annealing," in which the donor is used to re-synthesize genetic information that will become part of the target, and/or related processes. Such specialized HR often results in an alteration of the sequence of the target molecule such that part or all of the sequence of the donor polynucleotide is incorporated into the target polynucleotide.

[0084] In the methods of the disclosure, one or more targeted nucleases as described herein create a double-stranded break in the target sequence (*e.g.,* cellular chromatin) at a predetermined site, and a "donor" polynucleotide, having homology to the nucleotide sequence in the region of the break, can be introduced into the cell. The presence of the double-stranded break has been shown to facilitate integration of the donor sequence. The donor sequence may be physically integrated or, alternatively, the donor polynucleotide is used as a template for repair of the break via homologous recombination, resulting in the introduction of all or part of the nucleotide sequence as in the donor into the cellular chromatin. Thus, a first sequence in cellular chromatin can be altered and, in certain embodiments, can be converted into a sequence present in a donor polynucleotide. Thus, the use of the terms "replace" or "replacement" can be understood to represent replacement of one nucleotide sequence by another, (*i.e.*, replacement of a sequence in the informational sense), and does not necessarily require physical or chemical replacement of one polynucleotide by another.

[0085] In any of the methods described herein, additional pairs of zinc-finger or TALEN proteins can be used for additional double-stranded cleavage of additional target sites within the cell.

[0086] In certain embodiments of methods for targeted recombination and/or replacement and/or alteration of a sequence in a region of interest in cellular chromatin, a chromosomal sequence is altered by homologous recombination with an exogenous "donor" nucleotide sequence. Such homologous recombination is stimulated by the presence of a double-stranded break in cellular chromatin, if sequences homologous to the region of the break are present.

[0087] In any of the methods described herein, the first nucleotide sequence (the "donor sequence") can contain sequences that are homologous, but not identical, to genomic sequences in the region of interest, thereby stimulating homologous recombination to insert a non-identical sequence in the region of interest. Thus, in certain embodiments, portions of the donor sequence that are homologous to sequences in the region of interest exhibit between about 80 to 99% (or any value therebetween) sequence identity to the genomic sequence that is replaced. In other embodiments, the homology between the donor and genomic sequence is higher than 99%, for example if only 1 nucleotide differs as between donor and genomic sequences of over 101 contiguous base pairs. In certain cases, a non-homologous portion of the donor sequence can contain sequences not present in the region of interest, such that new sequences are introduced into the region of interest. In these instances, the non-homologous sequence is generally flanked by sequences of 50-1,000 base pairs (or any integral value therebetween) or any number of base pairs greater than 1,000, that are homologous or identical to sequences in the region of interest. In other embodiments, the donor sequence is non-homologous to the first sequence, and is inserted into the genome by non-homologous recombination mechanisms.

[0088] Any of the methods described herein can be used for partial or complete inactivation of one or more target sequences in a cell by targeted integration of donor sequence that disrupts expression of the gene(s) of interest. Cell lines with partially or completely inactivated genes are also provided.

[0089] Furthermore, the methods of targeted integration as described herein can also be used to integrate one or more exogenous sequences. The exogenous nucleic acid sequence can comprise, for example, one or more genes or cDNA molecules, or any type of coding or non-coding sequence, as well as one or more control elements (*e.g.,* promoters). In addition, the exogenous nucleic acid sequence

may produce one or more RNA molecules (*e.g.*, small hairpin RNAs (shRNAs), inhibitory RNAs (RNAis), microRNAs (miRNAs), *etc.*).

[0090] "Cleavage" refers to the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, fusion polypeptides are used for targeted double-stranded DNA cleavage.

[0091] A "cleavage half-domain" is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity). The terms "first and second cleavage half-domains;" "+ and – cleavage half-domains" and "right and left cleavage half-domains" are used interchangeably to refer to pairs of cleavage half-domains that dimerize.

[0092] An "engineered cleavage half-domain" is a cleavage half-domain that has been modified so as to form obligate heterodimers with another cleavage half-domain (*e.g.,* another engineered cleavage half-domain). *See, also,* U.S. Patent Publication Nos. 2005/0064474, 20070218528, 2008/0131962 and 2011/0201055, incorporated herein by reference in their entireties.

[0093] The term "sequence" refers to a nucleotide sequence of any length, which can be DNA or RNA; can be linear, circular or branched and can be either single-stranded or double stranded. The term "donor sequence" refers to a nucleotide sequence

that is inserted into a genome. A donor sequence can be of any length, for example between 2 and 10,000 nucleotides in length (or any integer value therebetween or thereabove), preferably between about 100 and 1,000 nucleotides in length (or any integer therebetween), more preferably between about 200 and 500 nucleotides in length.

[0094] "Chromatin" is the nucleoprotein structure comprising the cellular genome. Cellular chromatin comprises nucleic acid, primarily DNA, and protein, including histones and non-histone chromosomal proteins. The majority of eukaryotic cellular chromatin exists in the form of nucleosomes, wherein a

nucleosome core comprises approximately 150 base pairs of DNA associated with an octamer comprising two each of histones H2A, H2B, H3 and H4; and linker DNA (of variable length depending on the organism) extends between nucleosome cores. A molecule of histone H1 is generally associated with the linker DNA. For the purposes of the present disclosure, the term "chromatin" is meant to encompass all types of cellular nucleoprotein, both prokaryotic and eukaryotic. Cellular chromatin includes both chromosomal and episomal chromatin.

[0095] A "chromosome," is a chromatin complex comprising all or a portion of the genome of a cell. The genome of a cell is often characterized by its karyotype, which is the collection of all the chromosomes that comprise the genome of the cell. The genome of a cell can comprise one or more chromosomes.

[0096] An "episome" is a replicating nucleic acid, nucleoprotein complex or other structure comprising a nucleic acid that is not part of the chromosomal karyotype of a cell. Examples of episomes include plasmids and certain viral genomes.

[0097] A "target site" or "target sequence" is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule will bind, provided sufficient conditions for binding exist.

[0098] An "exogenous" molecule is a molecule that is not normally present in a cell, but can be introduced into a cell by one or more genetic, biochemical or other methods. "Normal presence in the cell" is determined with respect to the particular developmental stage and environmental conditions of the cell. Thus, for example, a molecule that is present only during embryonic development of muscle is an exogenous molecule with respect to an adult muscle cell. Similarly, a molecule induced by heat shock is an exogenous molecule with respect to a non-heat-shocked cell. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule or a malfunctioning version of a normally-functioning endogenous molecule.

[0099] An exogenous molecule can be, among other things, a small molecule, such as is generated by a combinatorial chemistry process, or a macromolecule such as a protein, nucleic acid, carbohydrate, lipid, glycoprotein, lipoprotein, polysaccharide, any modified derivative of the above molecules, or any complex comprising one or more of the above molecules. Nucleic acids include DNA and

RNA, can be single- or double-stranded; can be linear, branched or circular; and can be of any length. Nucleic acids include those capable of forming duplexes, as well as triplex-forming nucleic acids. See, for example, U.S. Patent Nos. 5,176,996 and 5,422,251. Proteins include, but are not limited to, DNA-binding proteins, transcription factors, chromatin remodeling factors, methylated DNA binding proteins, polymerases, methylases, demethylases, acetylases, deacetylases, kinases, phosphatases, integrases, recombinases, ligases, topoisomerases, gyrases and helicases.

[0100] An exogenous molecule can be the same type of molecule as an endogenous molecule, *e.g.*, an exogenous protein or nucleic acid. For example, an exogenous nucleic acid can comprise an infecting viral genome, a plasmid or episome introduced into a cell, or a chromosome that is not normally present in the cell. Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (*i.e.*, liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer. An exogenous molecule can also be the same type of molecule as an endogenous molecule but derived from a different species than the cell is derived from. For example, a human nucleic acid sequence may be introduced into a cell line originally derived from a mouse or hamster.

[0101] By contrast, an "endogenous" molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0102] A "fusion" molecule is a molecule in which two or more subunit molecules are linked, preferably covalently. The subunit molecules can be the same chemical type of molecule, or can be different chemical types of molecules. Examples of the first type of fusion molecule include, but are not limited to, fusion proteins (for example, a fusion between a ZFP or TALE DNA-binding domain and one or more activation domains) and fusion nucleic acids (for example, a nucleic acid encoding the fusion protein described *supra*). Examples of the second type of fusion

molecule include, but are not limited to, a fusion between a triplex-forming nucleic acid and a polypeptide, and a fusion between a minor groove binder and a nucleic acid.

[0103] Expression of a fusion protein in a cell can result from delivery of the fusion protein to the cell or by delivery of a polynucleotide encoding the fusion protein to a cell, wherein the polynucleotide is transcribed, and the transcript is translated, to generate the fusion protein. Trans-splicing, polypeptide cleavage and polypeptide ligation can also be involved in expression of a protein in a cell. Methods for polynucleotide and polypeptide delivery to cells are presented elsewhere in this disclosure.

[0104] A "gene," for the purposes of the present disclosure, includes a DNA region encoding a gene product (see *infra*), as well as all DNA regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions.

[0105] "Gene expression" refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (*e.g.*, mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA or any other type of RNA) or a protein produced by translation of an mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristilation, and glycosylation.

[0106] "Modulation" of gene expression refers to a change in the activity of a gene. Modulation of expression can include, but is not limited to, gene activation and gene repression. Genome editing (*e.g.,* cleavage, alteration, inactivation, random mutation) can be used to modulate expression. Gene inactivation refers to any reduction in gene expression as compared to a cell that does not include a ZFP or TALEN as described herein. Thus, gene inactivation may be partial or complete.

[0107] A "region of interest" is any region of cellular chromatin, such as, for example, a gene or a non-coding sequence within or adjacent to a gene, in which it is desirable to bind an exogenous molecule. Binding can be for the purposes of targeted DNA cleavage and/or targeted recombination. A region of interest can be present in a chromosome, an episome, an organellar genome (*e.g.*, mitochondrial, chloroplast), or an infecting viral genome, for example. A region of interest can be within the coding region of a gene, within transcribed non-coding regions such as, for example, leader sequences, trailer sequences or introns, or within non-transcribed regions, either upstream or downstream of the coding region. A region of interest can be as small as a single nucleotide pair or up to 2,000 nucleotide pairs in length, or any integral value of nucleotide pairs.

[0108] "Eukaryotic" cells include, but are not limited to, fungal cells (such as yeast), plant cells, animal cells, mammalian cells and human cells (*e.g.*, T-cells).

[0109] "Secretory tissues" are those tissues that secrete products. Examples of secretory tissues that are localized to the gastrointestinal tract include the cells that line the gut, the pancreas, and the gallbladder. Other secretory tissues include the liver, tissues associated with the eye and mucous membranes such as salivary glands, mammary glands, the prostate gland, the pituitary gland and other members of the endocrine system. Additionally, secretory tissues include individual cells of a tissue type which are capable of secretion.

[0110] The terms "operative linkage" and "operatively linked" (or "operably linked") are used interchangeably with reference to a juxtaposition of two or more components (such as sequence elements), in which the components are arranged such that both components function normally and allow the possibility that at least one of the components can mediate a function that is exerted upon at least one of the other components. By way of illustration, a transcriptional regulatory sequence, such as a promoter, is operatively linked to a coding sequence if the transcriptional regulatory sequence controls the level of transcription of the coding sequence in response to the presence or absence of one or more transcriptional regulatory factors. A transcriptional regulatory sequence is generally operatively linked in *cis* with a coding sequence, but need not be directly adjacent to it. For example, an enhancer is a transcriptional regulatory sequence that is operatively linked to a coding sequence, even though they are not contiguous.

[0111] With respect to fusion polypeptides, the term "operatively linked" can refer to the fact that each of the components performs the same function in linkage to the other component as it would if it were not so linked. For example, with respect to a fusion polypeptide in which a ZFP or TALE DNA-binding domain is fused to an activation domain, the ZFP or TALE DNA-binding domain and the activation domain are in operative linkage if, in the fusion polypeptide, the ZFP or TALE DNA-binding domain portion is able to bind its target site and/or its binding site, while the activation domain is able to up-regulate gene expression. When a fusion polypeptide in which a ZFP or TALE DNA-binding domain is fused to a cleavage domain, the ZFP or TALE DNA-binding domain and the cleavage domain are in operative linkage if, in the fusion polypeptide, the ZFP or TALE DNA-binding domain portion is able to bind its target site and/or its binding site, while the cleavage domain is able to cleave DNA in the vicinity of the target site.

[0112] A "functional fragment" of a protein, polypeptide or nucleic acid is a protein, polypeptide or nucleic acid whose sequence is not identical to the full-length protein, polypeptide or nucleic acid, yet retains the same function as the full-length protein, polypeptide or nucleic acid. A functional fragment can possess more, fewer, or the same number of residues as the corresponding native molecule, and/or can contain one or more amino acid or nucleotide substitutions. Methods for determining the function of a nucleic acid (*e.g.*, coding function, ability to hybridize to another nucleic acid) are well-known in the art. Similarly, methods for determining protein function are well-known. For example, the DNA-binding function of a polypeptide can be determined, for example, by filter-binding, electrophoretic mobility-shift, or immunoprecipitation assays. DNA cleavage can be assayed by gel electrophoresis. See Ausubel *et al., supra*. The ability of a protein to interact with another protein can be determined, for example, by co-immunoprecipitation, two-hybrid assays or complementation, both genetic and biochemical. See, for example, Fields *et al.* (1989) *Nature* **340:**245-246; U.S. Patent No. 5,585,245 and PCT WO 98/44350.

[0113] A "vector" is capable of transferring gene sequences to target cells. Typically, "vector construct," "expression vector," and "gene transfer vector," mean any nucleic acid construct capable of directing the expression of a gene of interest and which can transfer gene sequences to target cells. Thus, the term includes cloning, and expression vehicles, as well as integrating vectors.

[0114] A "reporter gene" or "reporter sequence" refers to any sequence that produces a protein product that is easily measured, preferably although not necessarily in a routine assay. Suitable reporter genes include, but are not limited to, sequences encoding proteins that mediate antibiotic resistance (*e.g.*, ampicillin resistance, neomycin resistance, G418 resistance, puromycin resistance), sequences encoding colored or fluorescent or luminescent proteins (*e.g.*, green fluorescent protein, enhanced green fluorescent protein, red fluorescent protein, luciferase), and proteins which mediate enhanced cell growth and/or gene amplification (*e.g.*, dihydrofolate reductase). Epitope tags include, for example, one or more copies of FLAG, His, myc, Tap, HA or any detectable amino acid sequence. "Expression tags" include sequences that encode reporters that may be operably linked to a desired gene sequence in order to monitor expression of the gene of interest.

Nucleases

[0115] Described herein are compositions, particularly nucleases, which are useful targeting a gene for the insertion of a transgene, for example, nucleases that are specific for albumin. In certain embodiments, the nuclease is naturally occurring. In other embodiments, the nuclease is non-naturally occurring, i.e., engineered in the DNA-binding domain and/or cleavage domain. For example, the DNA-binding domain of a naturally-occurring nuclease may be altered to bind to a selected target site (*e.g.,* a meganuclease that has been engineered to bind to site different than the cognate binding site). In other embodiments, the nuclease comprises heterologous DNA-binding and cleavage domains (*e.g.*, zinc finger nucleases; TAL-effector nucleases; meganuclease DNA-binding domains with heterologous cleavage domains).

A. DNA-binding domains

[0116] In certain embodiments, the nuclease is a meganuclease (homing endonuclease). Naturally-occurring meganucleases recognize 15-40 base-pair cleavage sites and are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cyst box family and the HNH family. Exemplary homing endonucleases include I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III. Their

recognition sequences are known. *See* also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.* (1994) Nucleic Acids Res. 22, 1125–1127; Jasin (1996) *Trends Genet.* **12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue.

[0117] In certain embodiments, the nuclease comprises an engineered (non-naturally occurring) homing endonuclease (meganuclease). The recognition sequences of homing endonucleases and meganucleases such as I-*Sce*I, I-*Ceu*I, PI-*Psp*I, PI-*Sce*, I-*Sce*IV, I-*Csm*I, I-*Pan*I, I-*Sce*II, I-*Ppo*I, I-*Sce*III, I-*Cre*I, I-*Tev*I, I-*Tev*II and I-*Tev*III are known. *See* also U.S. Patent No. 5,420,032; U.S. Patent No. 6,833,252; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379–3388; Dujon *et al.* (1989) *Gene* **82**:115–118; Perler *et al.* (1994) Nucleic Acids Res. 22, 1125–1127; Jasin (1996) *Trends Genet.* **12**:224–228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163–180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345–353 and the New England Biolabs catalogue. In addition, the DNA-binding specificity of homing endonucleases and meganucleases can be engineered to bind non-natural target sites. *See*, for example, Chevalier *et al.* (2002) *Molec. Cell* **10**:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* **31**:2952-2962; Ashworth *et al.* (2006) *Nature* **441**:656-659; Paques *et al.* (2007) *Current Gene Therapy* **7**:49-66; U.S. Patent Publication No. 20070117128. The DNA-binding domains of the homing endonucleases and meganucleases may be altered in the context of the nuclease as a whole (*i.e.*, such that the nuclease includes the cognate cleavage domain) or may be fused to a heterologous cleavage domain.

[0118] In other embodiments, the DNA-binding domain comprises a naturally occurring or engineered (non-naturally occurring) TALE DNA binding domain. *See, e.g.*, U.S. Patent Publication No. 2011/0301073, incorporated by reference in its entirety herein. The plant pathogenic bacteria of the genus *Xanthomonas* are known to cause many diseases in important crop plants. Pathogenicity of *Xanthomonas* depends on a conserved type III secretion (T3S) system which injects more than 25 different effector proteins into the plant cell. Among these injected proteins are transcription activator-like effectors (TALE) which mimic plant transcriptional activators and manipulate the plant transcriptome (see Kay *et al* (2007) *Science* 318:648-651). These proteins contain a DNA binding domain and a transcriptional

activation domain. One of the most well characterized TALEs is AvrBs3 from *Xanthomonas campestgris* pv. *Vesicatoria* (see Bonas *et al* (1989) *Mol Gen Genet* 218: 127-136 and WO2010079430). TALEs contain a centralized domain of tandem repeats, each repeat containing approximately 34 amino acids, which are key to the DNA binding specificity of these proteins. In addition, they contain a nuclear localization sequence and an acidic transcriptional activation domain (for a review see Schornack S, *et al* (2006) *J Plant Physiol* 163(3): 256-272). In addition, in the phytopathogenic bacteria *Ralstonia solanacearum* two genes, designated brg11 and hpx17 have been found that are homologous to the AvrBs3 family of *Xanthomonas* in the *R. solanacearum* biovar 1 strain GMI1000 and in the biovar 4 strain RS1000 (See Heuer *et al* (2007) *Appl and Envir Micro* 73(13): 4379-4384). These genes are 98.9% identical in nucleotide sequence to each other but differ by a deletion of 1,575 bp in the repeat domain of hpx17. However, both gene products have less than 40% sequence identity with AvrBs3 family proteins of *Xanthomonas*.

[0119] Thus, in some embodiments, the DNA binding domain that binds to a target site in a target locus (*e.g.,* albumin or safe harbor) is an engineered domain from a TALE similar to those derived from the plant pathogens Xanthomonas (see Boch *et al*, (2009) *Science* 326: 1509-1512 and Moscou and Bogdanove, (2009) *Science*326: 1501) and Ralstonia (see Heuer *et al* (2007) *Applied and Environmental Microbiology* 73(13): 4379-4384); U.S. Patent Publication No. 2011/0301073 and U.S. Patent Publication No. 20110145940.

[0120] In certain embodiments, the DNA binding domain comprises a zinc finger protein (*e.g.,* a zinc finger protein that binds to a target site in an albumin or safe-harbor gene). Preferably, the zinc finger protein is non-naturally occurring in that it is engineered to bind to a target site of choice. *See*, for example, *See*, for example, Beerli *et al.* (2002) *Nature Biotechnol.* **20:**135-141; Pabo *et al.* (2001) *Ann. Rev. Biochem.* **70:**313-340; Isalan *et al.* (2001) *Nature Biotechnol.* **19:**656-660; Segal *et al.* (2001) *Curr. Opin. Biotechnol.* **12:**632-637; Choo *et al.* (2000) *Curr. Opin. Struct. Biol.* **10:**411-416; U.S. Patent Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

[0121] An engineered zinc finger binding or TALE domain can have a novel binding specificity, compared to a naturally-occurring zinc finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual zinc finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

[0122] Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0123] In addition, as disclosed in these and other references, DNA-binding domains (*e.g.*, multi-finger zinc finger proteins or TALE domains) may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The DNA binding proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned WO 02/077227.

[0124] Selection of target sites; DNA-binding domains and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Nos. 6,140,0815; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970 WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 and WO 03/016496 and U.S. Publication No. 20110301073.

[0125] In addition, as disclosed in these and other references, DNA-binding domains (*e.g.*, multi-finger zinc finger proteins) may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein.

B. Cleavage Domains

[0126] Any suitable cleavage domain can be operatively linked to a DNA-binding domain to form a nuclease. For example, ZFP DNA-binding domains have been fused to nuclease domains to create ZFNs – a functional entity that is able to recognize its intended nucleic acid target through its engineered (ZFP) DNA binding domain and cause the DNA to be cut near the ZFP binding site via the nuclease activity. See, *e.g.*, Kim *et al.* (1996) *Proc Nat'l Acad Sci USA* 93(3):1156-1160. More recently, ZFNs have been used for genome modification in a variety of organisms. *See*, for example, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275. Likewise, TALE DNA-binding domains have been fused to nuclease domains to create TALENs. *See, e.g.,* U.S. Publication No. 20110301073.

[0127] As noted above, the cleavage domain may be heterologous to the DNA-binding domain, for example a zinc finger DNA-binding domain and a cleavage domain from a nuclease or a TALEN DNA-binding domain and a cleavage domain, or meganuclease DNA-binding domain and cleavage domain from a different nuclease. Heterologous cleavage domains can be obtained from any endonuclease or exonuclease. Exemplary endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. *See*, for example, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; and Belfort *et al.* (1997) *Nucleic Acids Res.* **25:**3379-3388. Additional enzymes which cleave DNA are known (*e.g.*, S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease; *see* also Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press,1993). One or more of

these enzymes (or functional fragments thereof) can be used as a source of cleavage domains and cleavage half-domains.

[0128] Similarly, a cleavage half-domain can be derived from any nuclease or portion thereof, as set forth above, that requires dimerization for cleavage activity. In general, two fusion proteins are required for cleavage if the fusion proteins comprise cleavage half-domains. Alternatively, a single protein comprising two cleavage half-domains can be used. The two cleavage half-domains can be derived from the same endonuclease (or functional fragments thereof), or each cleavage half-domain can be derived from a different endonuclease (or functional fragments thereof). In addition, the target sites for the two fusion proteins are preferably disposed, with respect to each other, such that binding of the two fusion proteins to their respective target sites places the cleavage half-domains in a spatial orientation to each other that allows the cleavage half-domains to form a functional cleavage domain, *e.g.*, by dimerizing. Thus, in certain embodiments, the near edges of the target sites are separated by 5-8 nucleotides or by 15-18 nucleotides. However any integral {why are we always using the qualifiers "integral" and "integer" -- are these really necessary? They just seem restrictive and their use would seem to open us up to workarounds}. number of nucleotides or nucleotide pairs can intervene between two target sites (*e.g.*, from 2 to 50 nucleotide pairs or more). In general, the site of cleavage lies between the target sites.

[0129] Restriction endonucleases (restriction enzymes) are present in many species and are capable of sequence-specific binding to DNA (at a recognition site), and cleaving DNA at or near the site of binding. Certain restriction enzymes (*e.g.*, Type IIS) cleave DNA at sites removed from the recognition site and have separable binding and cleavage domains. For example, the Type IIS enzyme *Fok* I catalyzes double-stranded cleavage of DNA, at 9 nucleotides from its recognition site on one strand and 13 nucleotides from its recognition site on the other. *See*, for example, US Patents 5,356,802; 5,436,150 and 5,487,994; as well as Li *et al.* (1992) *Proc. Natl. Acad. Sci. USA* **89:**4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. USA* **90:**2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* **91:**883-887; Kim *et al.* (1994b) *J. Biol. Chem.* **269:**31,978-31,982. Thus, in one embodiment, fusion proteins comprise the cleavage domain (or cleavage half-domain) from at least one Type IIS

restriction enzyme and one or more zinc finger binding domains, which may or may not be engineered.

[0130] An exemplary Type IIS restriction enzyme, whose cleavage domain is separable from the binding domain, is *Fok* I. This particular enzyme is active as a dimer. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95:** 10,570-10,575. Accordingly, for the purposes of the present disclosure, the portion of the *Fok* I enzyme used in the disclosed fusion proteins is considered a cleavage half-domain. Thus, for targeted double-stranded cleavage and/or targeted replacement of cellular sequences using zinc finger-*Fok* I fusions, two fusion proteins, each comprising a *Fok*I cleavage half-domain, can be used to reconstitute a catalytically active cleavage domain. Alternatively, a single polypeptide molecule containing a DNA binding domain and two *Fok* I cleavage half-domains can also be used.

[0131] A cleavage domain or cleavage half-domain can be any portion of a protein that retains cleavage activity, or that retains the ability to multimerize (*e.g.*, dimerize) to form a functional cleavage domain.

[0132] Exemplary Type IIS restriction enzymes are described in International Publication WO 07/014275, incorporated herein in its entirety. Additional restriction enzymes also contain separable binding and cleavage domains, and these are contemplated by the present disclosure. *See*, for example, Roberts *et al.* (2003) *Nucleic Acids Res.* **31:**418-420.

[0133] In certain embodiments, the cleavage domain comprises one or more engineered cleavage half-domain (also referred to as dimerization domain mutants) that minimize or prevent homodimerization, as described, for example, in U.S. Patent Publication Nos. 20050064474; 20060188987; 20080131962 and 20110201055, the disclosures of all of which are incorporated by reference in their entireties herein. Amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 of *Fok* I are all targets for influencing dimerization of the *Fok* I cleavage half-domains.

[0134] Exemplary engineered cleavage half-domains of *Fok* I that form obligate heterodimers include a pair in which a first cleavage half-domain includes mutations at amino acid residues at positions 490 and 538 of *Fok* I and a second cleavage half-domain includes mutations at amino acid residues 486 and 499.

[0135] Thus, in one embodiment, a mutation at 490 replaces Glu (E) with Lys (K); the mutation at 538 replaces Iso (I) with Lys (K); the mutation at 486 replaced Gln (Q) with Glu (E); and the mutation at position 499 replaces Iso (I) with Lys (K). Specifically, the engineered cleavage half-domains described herein were prepared by mutating positions 490 (E→K) and 538 (I→K) in one cleavage half-domain to produce an engineered cleavage half-domain designated "E490K:I538K" and by mutating positions 486 (Q→E) and 499 (I→L) in another cleavage half-domain to produce an engineered cleavage half-domain designated "Q486E:I499L". The engineered cleavage half-domains described herein are obligate heterodimer mutants in which aberrant cleavage is minimized or abolished. *See, e.g.,* U.S. Patent Publication No. 2008/0131962, the disclosure of which is incorporated by reference in its entirety for all purposes.

[0136] In certain embodiments, the engineered cleavage half-domain comprises mutations at positions 486, 499 and 496 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Gln (Q) residue at position 486 with a Glu (E) residue, the wild type Iso (I) residue at position 499 with a Leu (L) residue and the wild-type Asn (N) residue at position 496 with an Asp (D) or Glu (E) residue (also referred to as a "ELD" and "ELE" domains, respectively). In other embodiments, the engineered cleavage half-domain comprises mutations at positions 490, 538 and 537 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue, the wild type Iso (I) residue at position 538 with a Lys (K) residue, and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as "KKK" and "KKR" domains, respectively). In other embodiments, the engineered cleavage half-domain comprises mutations at positions 490 and 537 (numbered relative to wild-type FokI), for instance mutations that replace the wild type Glu (E) residue at position 490 with a Lys (K) residue and the wild-type His (H) residue at position 537 with a Lys (K) residue or a Arg (R) residue (also referred to as "KIK" and "KIR" domains, respectively). (See US Patent Publication No. 20110201055). Engineered cleavage half-domains described herein can be prepared using any suitable method, for example, by site-directed mutagenesis of wild-type cleavage half-domains (*Fok* I) as described in U.S. Patent Publication Nos. 20050064474; 20080131962 and 20110201055.

[0137] Alternatively, nucleases may be assembled *in vivo* at the nucleic acid target site using so-called "split-enzyme" technology (*see e.g.* U.S. Patent Publication No. 20090068164). Components of such split enzymes may be expressed either on separate expression constructs, or can be linked in one open reading frame where the individual components are separated, for example, by a self-cleaving 2A peptide or IRES sequence. Components may be individual zinc finger binding domains or domains of a meganuclease nucleic acid binding domain.

[0138] Nucleases can be screened for activity prior to use, for example in a yeast-based chromosomal system as described in WO 2009/042163 and 20090068164. Nuclease expression constructs can be readily designed using methods known in the art. See, *e.g.*, United States Patent Publications 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; and International Publication WO 07/014275. Expression of the nuclease may be under the control of a constitutive promoter or an inducible promoter, for example the galactokinase promoter which is activated (de-repressed) in the presence of raffinose and/or galactose and repressed in presence of glucose.

Target Sites

[0139] As described in detail above, DNA domains can be engineered to bind to any sequence of choice in a locus, for example an albumin or safe-harbor gene. An engineered DNA-binding domain can have a novel binding specificity, compared to a naturally-occurring DNA-binding domain. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual (*e.g.,* zinc finger) amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of DNA binding domain which bind the particular triplet or quadruplet sequence. *See*, for example, co-owned U.S. Patents 6,453,242 and 6,534,261, incorporated by reference herein in their entireties. Rational design of TAL-effector domains can also be performed. *See, e.g.,* U.S. Patent Publication No. 20110301073.

[0140] Exemplary selection methods applicable to DNA-binding domains, including phage display and two-hybrid systems, are disclosed in US Patents 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and

6,242,568; as well as WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 and GB 2,338,237.

[0141] Selection of target sites; nucleases and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Patent Application Publication Nos. 20050064474 and 20060188987, incorporated by reference in their entireties herein.

[0142] In addition, as disclosed in these and other references, DNA-binding domains (*e.g.*, multi-finger zinc finger proteins) may be linked together using any suitable linker sequences, including for example, linkers of 5 or more amino acids. *See, e.g.*, U.S. Patent Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual DNA-binding domains of the protein. *See, also*, U.S. Publication No. 20110301073.

Donors

[0143] As noted above, insertion of an exogenous sequence (also called a "donor sequence" or "donor"), for example for correction of a mutant gene or for increased expression of a wild-type gene. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence where it is placed. A donor sequence can contain a non-homologous sequence flanked by two regions of homology to allow for efficient HDR at the location of interest. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous to the region of interest in cellular chromatin. A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. For example, for targeted insertion of sequences not normally present in a region of interest, said sequences can be present in a donor nucleic acid molecule and flanked by regions of homology to sequence in the region of interest.

[0144] The donor polynucleotide can be DNA or RNA, single-stranded or double-stranded and can be introduced into a cell in linear or circular form. *See, e.g.*, U.S. Patent Publication Nos. 20100047805 and 20110207221. If introduced in linear form, the ends of the donor sequence can be protected (*e.g.*, from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more

dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. *See*, for example, Chang *et al.* (1987) *Proc. Natl. Acad. Sci. USA* **84:**4959-4963; Nehls *et al.* (1996) *Science* **272:**886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues.

[0145] A polynucleotide can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor polynucleotides can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (*e.g.*, adenovirus, AAV, herpesvirus, retrovirus, lentivirus and integrase defective lentivirus (IDLV)).

[0146] The donor is generally inserted so that its expression is driven by the endogenous promoter at the integration site, namely the promoter that drives expression of the albumin gene. However, it will be apparent that the donor may comprise a promoter and/or enhancer, for example a constitutive promoter or an inducible or tissue specific promoter.

[0147] The donor molecule may be inserted into an endogenous gene such that all, some or none of the endogenous gene is expressed. For example, a transgene as described herein may be inserted into an albumin locus such that some or none of the endogenous albumin sequences are expressed, for example as a fusion with the transgene. In other embodiments, the transgene (*e.g.*, with or without albumin encoding sequences) is integrated into any endogenous locus, for example a safe-harbor locus. *See, e.g.*, US patent publications 20080299580; 20080159996 and 201000218264.

[0148] When albumin sequences (endogenous or part of the transgene) are expressed with the transgene, the albumin sequences may be full-length sequences (wild-type or mutant) or partial sequences. Preferably the albumin sequences are functional. Non-limiting examples of the function of these full length or partial albumin sequences include increasing the serum half-life of the polypeptide expressed by the transgene (*e.g.*, therapeutic gene) and/or acting as a carrier.

[0149] Furthermore, although not required for expression, exogenous sequences may also include transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals.

Delivery

[0150] The nucleases, polynucleotides encoding these nucleases, donor polynucleotides and compositions comprising the proteins and/or polynucleotides described herein may be delivered *in vivo* or *ex vivo* by any suitable means.

[0151] Methods of delivering nucleases as described herein are described, for example, in U.S. Patent Nos. 6,453,242; 6,503,717; 6,534,261; 6,599,692; 6,607,882; 6,689,558; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, the disclosures of all of which are incorporated by reference herein in their entireties.

[0152] Nucleases and/or donor constructs as described herein may also be delivered using vectors containing sequences encoding one or more of the zinc finger or TALEN protein(s). Any vector systems may be used including, but not limited to, plasmid vectors, retroviral vectors, lentiviral vectors, adenovirus vectors, poxvirus vectors; herpesvirus vectors and adeno-associated virus vectors, etc. *See, also,* U.S. Patent Nos. 6,534,261; 6,607,882; 6,824,978; 6,933,113; 6,979,539; 7,013,219; and 7,163,824, incorporated by reference herein in their entireties. Furthermore, it will be apparent that any of these vectors may comprise one or more of the sequences needed for treatment. Thus, when one or more nucleases and a donor construct are introduced into the cell, the nucleases and/or donor polynucleotide may be carried on the same vector or on different vectors. When multiple vectors are used, each vector may comprise a sequence encoding one or multiple nucleases and/or donor constructs.

[0153] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding nucleases and donor constructs in cells (*e.g.*, mammalian cells) and target tissues. Non-viral vector delivery systems include DNA plasmids, naked nucleic acid, and nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-

166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Böhm (eds.) (1995); and Yu *et al., Gene Therapy* 1:13-26 (1994).

[0154] Methods of non-viral delivery of nucleic acids include electroporation, lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Sonoporation using, *e.g.*, the Sonitron 2000 system (Rich-Mar) can also be used for delivery of nucleic acids.

[0155] Additional exemplary nucleic acid delivery systems include those provided by Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) and Copernicus Therapeutics Inc, (*see* for example US6008336). Lipofection is described in *e.g.*, U.S. Patent Nos. 5,049,386; 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (*e.g.*, Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424, WO 91/16024.

[0156] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (*see, e.g.*, Crystal, *Science* 270:404-410 (1995); Blaese *et al., Cancer Gene Ther.* 2:291-297 (1995); Behr *et al., Bioconjugate Chem.* 5:382-389 (1994); Remy *et al., Bioconjugate Chem.* 5:647-654 (1994); Gao *et al., Gene Therapy* 2:710-722 (1995); Ahmad *et al., Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0157] Additional methods of delivery include the use of packaging the nucleic acids to be delivered into EnGeneIC delivery vehicles (EDVs). These EDVs are specifically delivered to target tissues using bispecific antibodies where one arm of the antibody has specificity for the target tissue and the other has specificity for the EDV. The antibody brings the EDVs to the target cell surface and then the EDV is brought into the cell by endocytosis. Once in the cell, the contents are released (*see* MacDiarmid *et al* (2009) *Nature Biotechnology* 27(7):643).

[0158] The use of RNA or DNA viral based systems for the delivery of nucleic acids encoding engineered ZFPs take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro* and the modified cells are administered to patients (*ex vivo*). Conventional viral based systems for the delivery of ZFPs include, but are not limited to, retroviral, lentivirus, adenoviral, adeno-associated, vaccinia and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0159] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system depends on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (*see, e.g.,* Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al., J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*, *Virol.* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0160] In applications in which transient expression is preferred, adenoviral based systems can be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and high levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors are also used to transduce cells with target nucleic acids, *e.g.*, in the

in vitro production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (*see, e.g.*, West *et al.*, *Virology* 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors is described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989).

[0161] At least six viral vector approaches are currently available for gene transfer in clinical trials, which utilize approaches that involve complementation of defective vectors by genes inserted into helper cell lines to generate the transducing agent.

[0162] pLASN and MFG-S are examples of retroviral vectors that have been used in clinical trials (Dunbar *et al., Blood* 85:3048-305 (1995); Kohn *et al., Nat. Med.* 1:1017-102 (1995); Malech *et al., PNAS* 94:22 12133-12138 (1997)). PA317/pLASN was the first therapeutic vector used in a gene therapy trial. (Blaese *et al., Science* 270:475-480 (1995)). Transduction efficiencies of 50% or greater have been observed for MFG-S packaged vectors. (Ellem *et al., Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al., Hum. Gene Ther.* 1:111-2 (1997).

[0163] Recombinant adeno-associated virus vectors (rAAV) are a promising alternative gene delivery systems based on the defective and nonpathogenic parvovirus adeno-associated type 2 virus. All vectors are derived from a plasmid that retains only the AAV 145 bp inverted terminal repeats flanking the transgene expression cassette. Efficient gene transfer and stable transgene delivery due to integration into the genomes of the transduced cell are key features for this vector system. (Wagner *et al., Lancet* 351:9117 1702-3 (1998), Kearns *et al., Gene Ther.* 9:748-55 (1996)). Other AAV serotypes, including AAV1, AAV3, AAV4, AAV5, AAV6, AAV8, AAV 8.2, AAV9, AAV rh10 and pseudotyped AAV such as AAV2/8, AAV2/5 and AAV2/6 can also be used in accordance with the present invention.

[0164] Replication-deficient recombinant adenoviral vectors (Ad) can be produced at high titer and readily infect a number of different cell types. Most adenovirus vectors are engineered such that a transgene replaces the Ad E1a, E1b, and/or E3 genes; subsequently the replication defective vector is propagated in human

293 cells that supply deleted gene function in *trans*. Ad vectors can transduce multiple types of tissues *in vivo*, including non-dividing, differentiated cells such as those found in liver, kidney and muscle. Conventional Ad vectors have a large carrying capacity. An example of the use of an Ad vector in a clinical trial involved polynucleotide therapy for anti-tumor immunization with intramuscular injection (Sterman *et al., Hum. Gene Ther.* 7:1083-9 (1998)). Additional examples of the use of adenovirus vectors for gene transfer in clinical trials include Rosenecker *et al., Infection* 24:1 5-10 (1996); Sterman *et al., Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al., Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al., Hum. Gene Ther.* 5:597-613 (1997); Topf *et al., Gene Ther.* 5:507-513 (1998); Sterman *et al., Hum. Gene Ther.* 7:1083-1089 (1998).

[0165] Packaging cells are used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host (if applicable), other viral sequences being replaced by an expression cassette encoding the protein to be expressed. The missing viral functions are supplied in *trans* by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. The cell line is also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

[0166] In many gene therapy applications, it is desirable that the gene therapy vector be delivered with a high degree of specificity to a particular tissue type. Accordingly, a viral vector can be modified to have specificity for a given cell type by expressing a ligand as a fusion protein with a viral coat protein on the outer surface of the virus. The ligand is chosen to have affinity for a receptor known to be present on

the cell type of interest. For example, Han *et al.*, *Proc. Natl. Acad. Sci. USA* 92:9747-9751 (1995), reported that Moloney murine leukemia virus can be modified to express human heregulin fused to gp70, and the recombinant virus infects certain human breast cancer cells expressing human epidermal growth factor receptor. This principle can be extended to other virus-target cell pairs, in which the target cell expresses a receptor and the virus expresses a fusion protein comprising a ligand for the cell-surface receptor. For example, filamentous phage can be engineered to display antibody fragments (*e.g.*, FAB or Fv) having specific binding affinity for virtually any chosen cellular receptor. Although the above description applies primarily to viral vectors, the same principles can be applied to nonviral vectors. Such vectors can be engineered to contain specific uptake sequences which favor uptake by specific target cells.

[0167] Gene therapy vectors can be delivered *in vivo* by administration to an individual patient, typically by systemic administration (*e.g.*, intravenous, intraperitoneal, intramuscular, subdermal, or intracranial infusion) or topical application, as described below. Alternatively, vectors can be delivered to cells *ex vivo*, such as cells explanted from an individual patient (*e.g.*, lymphocytes, bone marrow aspirates, tissue biopsy) or universal donor hematopoietic stem cells, followed by reimplantation of the cells into a patient, usually after selection for cells which have incorporated the vector.

[0168] Vectors (*e.g.*, retroviruses, adenoviruses, liposomes, etc.) containing nucleases and/or donor constructs can also be administered directly to an organism for transduction of cells *in vivo*. Alternatively, naked DNA can be administered. Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells including, but not limited to, injection, infusion, topical application and electroporation. Suitable methods of administering such nucleic acids are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

[0169] Vectors suitable for introduction of polynucleotides described herein include non-integrating lentivirus vectors (IDLV). *See*, for example, Ory *et al.* (1996) *Proc. Natl. Acad. Sci. USA* **93:**11382-11388; Dull *et al.* (1998) *J. Virol.* **72:**8463-

8471; Zuffery *et al.* (1998) *J. Virol.* **72**:9873-9880; Follenzi *et al.* (2000) *Nature Genetics* **25**:217-222; U.S. Patent Publication No 2009/054985.

[0170] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions available, as described below (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed., 1989).

[0171] It will be apparent that the nuclease-encoding sequences and donor constructs can be delivered using the same or different systems. For example, a donor polynucleotide can be carried by a plasmid, while the one or more nucleases can be carried by a AAV vector. Furthermore, the different vectors can be administered by the same or different routes (intramuscular injection, tail vein injection, other intravenous injection, intraperitoneal administration and/or intramuscular injection. The vectors can be delivered simultaneously or in any sequential order.

[0172] Formulations for both *ex vivo* and *in vivo* administrations include suspensions in liquid or emulsified liquids. The active ingredients often are mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients include, for example, water, saline, dextrose, glycerol, ethanol or the like, and combinations thereof. In addition, the composition may contain minor amounts of auxiliary substances, such as, wetting or emulsifying agents, pH buffering agents, stabilizing agents or other reagents that enhance the effectiveness of the pharmaceutical composition.

Applications

[0173] The methods and compositions of the invention can be used in any circumstance wherein it is desired to supply a transgene encoding one or more proteins such that the protein(s) is(are) secreted from the targeted cell. Thus, this technology is of use in a condition where a patient is deficient in some protein due to problems (*e.g.,* problems in expression level or problems with the protein expressed as sub- or non-functioning). Particularly useful with this invention is the expression of transgenes to correct or restore functionality in clotting disorders. Additionally, A1AT-deficiency disorders such as COPD or liver damage, or other disorders, conditions or diseases that can be mitigated by the supply of exogenous proteins by a

secretory organ may be successfully treated by the methods and compositions of this invention. Lysosomal storage diseases can be treated by the methods and compositions of the invention, as are metabolic diseases such as diabetes.

[0174] Proteins that are useful therapeutically and that are typically delivered by injection or infusion are also useful with the methods and compositions of the invention. By way of non-limiting examples, production of a C-peptide (e.g. Ersatta™ by Cebix) or insulin for use in diabetic therapy. A further application includes treatment of Epidermolysis Bullosa via production of collagen VII. Expression of IGF-1 in secretory tissue as described herein can be used to increase levels of this protein in patients with liver cirrhosis and lipoprotein lipase deficiency by expression of lipoprotein lipase. Antibodies may also be secreted for therapeutic benefit, for example, for the treatment of cancers, autoimmune and other diseases. Other proteins related to clotting could be produced in secretory tissue, include fibrinogen, prothrombin, tissue factor, Factor V, Factor XI, Factor XII (Hageman factor), Factor XIII (fibrin-stabilizing factor), von Willebrand factor, prekallikrein, high molecular weight kininogen (Fitzgerald factor), fibronectin, antithrombin III, heparin cofactor II, protein C, protein S, protein Z, protein Z-related protease inhibitor, plasminogen, alpha 2-antiplasmin, tissue plasminogen activator, urokinase, plasminogen activator inhibitor-1, and plasminogen activator inhibitor-2.

[0175] The following Examples relate to exemplary embodiments of the present disclosure in which the nuclease comprises a zinc finger nuclease (ZFN) or TALEN. It will be appreciated that this is for purposes of exemplification only and that other nucleases can be used, for instance homing endonucleases (meganucleases) with engineered DNA-binding domains and/or fusions of naturally occurring of engineered homing endonucleases (meganucleases) DNA-binding domains and heterologous cleavage domains.

EXAMPLES

Example 1: Design, construction and characterization of zinc finger protein nucleases (ZFN) targeted to the mouse albumin gene

[0176] Zinc finger proteins were designed to target cleavage sites within introns 1, 12 and 13 of the mouse albumin gene. Corresponding expression constructs were assembled and incorporated into plasmids, AAV or adenoviral

vectors essentially as described in Urnov *et al.* (2005) *Nature* 435(7042):646-651, Perez *et al* (2008) *Nature Biotechnology* 26(7):808-816, and as described in U.S. Patent No. 6,534,261. Table 1 shows the recognition helices within the DNA binding domain of exemplary mouse albumin specific ZFPs while Table 2 shows the target sites for these ZFPs. Nucleotides in the target site that are contacted by the ZFP recognition helices are indicated in uppercase letters; non-contacted nucleotides indicated in lowercase.

Table 1: Murine albumin-specific zinc finger nucleases helix designs

Target SBS #	Design					
	F1	F2	F3	F4	F5	F6
Intron 1 30724	TSGSLTR (SEQ ID NO:1)	RSDALST (SEQ ID NO:2)	QSATRTK (SEQ ID NO:3)	TSGHLSR (SEQ ID NO:4)	QSGNLAR (SEQ ID NO:5)	NA
Intron 1 30725	RSDHLSA (SEQ ID NO:6)	TKSNRTK (SEQ ID NO:7)	DRSNLSR (SEQ ID NO:8)	WRSSLRA (SEQ ID NO:9)	DSSDRKK (SEQ ID NO:10)	NA
Intron 1 30732	TSGNLTR (SEQ ID NO:11)	DRSTRRQ (SEQ ID NO:12)	TSGSLTR (SEQ ID NO:1)	ERGTLAR(SEQ ID NO:13)	TSANLSR (SEQ ID NO:14)	NA
Intron 1 30733	DRSALAR (SEQ ID NO:15)	RSDHLSE (SEQ ID NO:16)	HRSDRTR (SEQ ID NO:17)	QSGALAR (SEQ ID NO:18)	QSGHLSR (SEQ ID NO:19)	NS
Intron 13 30759	RSDNLST (SEQ ID NO:20)	DRSALAR (SEQ ID NO:15)	DRSNLSR (SEQ ID NO:8)	DGRNLRH (SEQ ID NO:21)	RSDNLAR (SEQ ID NO:22)	QSNALNR (SEQ ID NO:23)
Intron 13 30761	DRSNLSR (SEQ ID NO:8)	LKQVLVR (SEQ ID NO:24)	QSGNLAR (SEQ ID NO:5)	QSTPLFA (SEQ ID NO:25)	QSGALAR (SEQ ID NO:18)	NA
Intron 13 30760	DRSNLSR (SEQ ID NO:8)	DGRNLRH (SEQ ID NO:21)	RSDNLAR (SEQ ID NO:22)	QSNALNR (SEQ ID NO:23)	NA	NA
Intron 13 30767	RSDNLSV (SEQ ID NO:26)	HSNARKT (SEQ ID NO:27)	RSDSLSA (SEQ ID NO:28)	QSGNLAR (SEQ ID NO:5)	RSDSLSV (SEQ ID NO:29)	QSGHLSR (SEQ ID NO:19)
Intron 13 30768	RSDNLSE (SEQ ID NO:30)	ERANRNS (SEQ ID NO:31)	QSANRTK (SEQ ID NO:32)	ERGTLAR (SEQ ID NO:13)	RSDALTQ (SEQ ID NO:33)	NA
Intron 13 30769	TSGSLTR (SEQ ID NO:1)	DRSNLSR (SEQ ID NO:8)	DGRNLRH (SEQ ID NO:21)	ERGTLAR (SEQ ID NO:13)	RSDALTQ (SEQ ID NO:33)	NA
Intron 12 30872	QSGHLAR (SEQ ID NO:34)	RSDHLTQ (SEQ ID NO:35)	RSDHLSQ (SEQ ID NO:36)	WRSSLVA (SEQ ID NO:37)	RSDVLSE (SEQ ID NO:38)	RNQHRKT (SEQ ID NO:39)
Intron 12 30873	QSGDLTR (SEQ ID NO:40)	RSDALAR (SEQ ID NO:41)	QSGDLTR (SEQ ID NO:40)	RRDPLIN (SEQ ID NO:42)	RSDNLSV (SEQ ID NO:26)	IRSTLRD (SEQ ID NO:43)
Intron	RSDNLSV	YSSTRNS	RSDHLSA	SYWSRTV	QSSDLSR	RTDALRG

12 30876	(SEQ ID NO:26)	(SEQ ID NO:44)	(SEQ ID NO:6)	(SEQ ID NO:45)	(SEQ ID NO:46)	(SEQ ID NO:47)
Intron 12 30877	RSDNLST (SEQ ID NO:20)	QKSPLNT (SEQ ID NO:48)	TSGNLTR (SEQ ID NO:11)	QAENLKS (SEQ ID NO:49)	QSSDLSR (SEQ ID NO:46)	RTDALRG (SEQ ID NO:47)
Intron 12 30882	RSDNLSV (SEQ ID NO:26)	RRAHLNQ (SEQ ID NO:50)	TSGNLTR (SEQ ID NO:11)	SDTNRFK (SEQ ID NO:51)	RSDNLST (SEQ ID NO:20)	QSGHLSR (SEQ ID NO:19)
Intron 12 30883	DSSDRKK (SEQ ID NO:10)	DRSALSR (SEQ ID NO:52)	TSSNRKT (SEQ ID NO:53)	QSGALAR (SEQ ID NO:18)	RSDHLSR (SEQ ID NO:54)	NA

Table 2: Target sites of murine albumin-specific ZFNs

Target	SBS #	Target site
Intron 1	30724	ctGAAGGTgGCAATGGTTcctctctgct_ (SEQ ID NO:55)
Intron 1	30725	ttTCCTGTAACGATCGGgaactggcatc_ (SEQ ID NO:56)
Intron 1	30732	aaGATGCCaGTTCCCGATcgttacagga_ (SEQ ID NO:57)
Intron 1	30733	agGGAGTAGCTTAGGTCagtgaagagaa_(SEQ ID NO:58)
Intron 13	30759	acGTAGAGAACAACATCTAGattggtgg_ (SEQ ID NO:59)
Intron 13	30761	ctGTAATAGAAACTGACttacgtagctt_(SEQ ID NO:60)
Intron 13	30760	acGTAGAGAACAACatctagattggtgg_(SEQ ID NO:59)
Intron 13	30767	agGGAATGtGAAATGATTCAGatatata_(SEQ ID NO:61)
Intron 13	30768	ccATGGCCTAACAACAGtttatcttctt_(SEQ ID NO:62)
Intron 13	30769	ccATGGCCtAACAACaGTTtatcttctt_(SEQ ID NO:62)
Intron 12	30872	ctTGGCTGTGTAGGAGGGGAgtagcagt_(SEQ ID NO:63)
Intron 12	30873	ttCCTAAGTTGGCAGTGGCAtgcttaat_(SEQ ID NO:64)
Intron 12	30876	ctTTGGCTTTGAGGATTAAGcatgccac_(SEQ ID NO:65)
Intron 12	30877	acTTGGCTcCAAGATTTATAGccttaaa_(SEQ ID NO:66)
Intron 12	30882	caGGAAAGTAAGATAGGAAGgaatgtga_(SEQ ID NO:67)
Intron 12	30883	ctGGGGTAAATGTCTCCttgctcttctt_(SEQ ID NO:68)

Example 2: Activity of murine albumin-specific ZFNs

[0177] The ZFNs in Table 1 were tested for the ability to cleave their endogenous target sequences in mouse cells. To accomplish this, constructs expressing the ZFNs in Table 1 were transfected into Neuro2A cells in the pairings indicated in Figure 1. Cells were then maintained at 37°C for 3 days or subjected to a hypothermic shock (30°C, see co-owned US Patent Publication No. 20110041195). Genomic DNA was then isolated from Neuro2A cells using the DNeasy kit (Qiagen) and subjected to the Cel-I assay (Surveyor™, Transgenomics) as described in Perez *et al*, (2008) *Nat. Biotechnol.* 26: 808-816 and Guschin *et al*, (2010) *Methods Mol Biol.* 649:247-56), in order to quantify chromosomal modifications induced by ZFN-

cleavage. In this assay, PCR is used to amplify a DNA fragment bearing the ZFN target site, and then the resultant amplicon is digested with the mismatch-specific nuclease Cel-I (Yang *et al*, (2000) *Biochemistry* 39, 3533-3541), followed by resolution of intact and cleaved amplicon on an agarose gel. By quantifying the degree of amplicon cleavage, one may calculate the fraction of mutated alleles in amplicon and therefore in the original cellular pool. In these experiments, all ZFN pairs were ELD/KKR *Fok*I mutation pairs (described above).

[0178] Results from the Cel-I assay are shown in Figure 1, and demonstrate that the ZFNs are capable of inducing cleavage and consequent mutations at their respective target sites. The "percent indel" value shown beneath each lane indicates the fraction of ZFN targets that were successfully cleaved and subsequently mutated during cellular repair of the double stranded break via NHEJ. The data also demonstrate increased activity when the transduction procedure incorporates the hypothermic shock.

Example 3: Canine albumin-specific ZFNs

[0179] A pair of ZFNs targeting the canine albumin locus was constructed for use in *in vivo* models. The pair was constructed as described in Example 1, and is shown below in Table 3. The target for each ZFN is provided in Table 4.

Table 3: Canine albumin-specific zinc finger nucleases helix designs

Target	SBS #	F1	F2	F3	F4	F5
Intron 1	33115	QRSNLDS (SEQ ID NO:83)	QSSDLSR (SEQ ID NO:46)	YHWYLKK (SEQ ID NO:84)	RSDDLSV (SEQ ID NO:85)	TSSNRTK (SEQ ID NO:86)
Intron 1	34077	QSGNLAR (SEQ ID NO:5)	QYTHLVA (SEQ ID NO:87)	RSDHLST (SEQ ID NO:88)	RSDARTT (SEQ ID NO:89)	DRSALAR (SEQ ID NO:15)

Table 4: Target sites of canine albumin-specific ZFNs

Target	SBS #	Target site
Intron 1	33115	agTATTCGTTTGCTcCAAaatatttgcc (SEQ ID NO:90)
Intron 1	34077	aaGTCATGTGGAGAGAAacacaaagagt (SEQ ID NO:91)

[0180] The canine specific ZFNs were tested *in vitro* for activity essentially as described in example 2, except that the canine cell line D17 was used. As shown in Figure 2, the ZFNs were shown to generate ~30% indels in this study.

Example 4: Non-human primate albumin specific ZFNs

[0181] ZFNs targeting the albumin locus in rhesus macaque monkeys (*Macaca mulatta*) were also made. The pairs were constructed as described above and are shown below in Table 5. The targets for the ZFNs are shown in Table 6. As shown below, the human (SEQ ID NO:92) and rhesus macaque (SEQ ID NO:93) sequences for the binding site for SBS# 35396 (see below, Table 7 and 8) are perfectly conserved. The differences between the human and rhesus sequences are boxed.

```
HUMAN LEADS          35364
                                  35396
HUMAN ATTGAATTCA TAACTATCCC AAAGACCTAT CCATTGCACT ATGCTTTATT TAAAAACCAC
RHESUS ATTGAATTCA TAACTGTCCC TAAGACCTAT CCATTGCACT ATGCTTTATT TAAAAGCCAC
                           G     (NOTE: G IN SOME INSTANCES)
```

[0182] Thus, for the development of the rhesus albumin specific pair, 35396 was paired with a series of partners which were designed to replace the human 35364 partner in rhesus. These proteins are shown below (Table 5) along with their target sequences (Table 6).

Table 5: Rhesus albumin-specific zinc finger nucleases helix designs

Target	SBS #					
Rhesus		F1	F2	F3	F4	F5
Intron 1	36813	QSGNLAR (SEQ ID NO:5)	HLGNLKT (SEQ ID NO:94)	LKHHLTD (SEQ ID NO:95)	DRSNLSR (SEQ ID NO:8)	RLDNRTA (SEQ ID NO:96)
Intron 1	36808	QSGNLAR (SEQ ID NO:5)	LMQNRNQ (SEQ ID NO:97)	LKHHLTD (SEQ ID NO:95)	DRSNLSR (SEQ ID NO:8)	RSDHLTT (SEQ ID NO:98)
Intron 1	36820	QRSNLVR (SEQ ID NO:99)	LRMNLTK (SEQ ID NO:100)	LKHHLTD (SEQ ID NO:95)	DRSNLSR (SEQ ID NO:8)	RSDHLTT (SEQ ID NO:98)
Intron 1	36819	QRSNLVR (SEQ ID NO:99)	LRMNLTK (SEQ ID NO:100)	LKHHLTD (SEQ ID NO:95)	DRSNLSR (SEQ ID NO:8)	RSDHLTQ (SEQ ID NO:35)
Intron 1	36806	QSGNLAR (SEQ ID	LMQNRNQ (SEQ ID	LKHHLTD (SEQ ID	DRSNLSR (SEQ ID	RSDHLTQ (SEQ ID

		NO:5)	NO:97)	NO:95)	NO:8)	NO:35)

Table 6: Target sites of rhesus albumin-specific ZFNs

Target	SBS #	Target site
Intron 1	36813	ttAGGGACAGTTATGAAttcaatcttca_ (SEQ ID NO:101)
Intron 1	36808	ttAGGGACAGTTATGAAttcaatcttca_ (SEQ ID NO:101)
Intron 1	36820	ttAGGGACAGTTATGAAttcaatcttca_(SEQ ID NO:101)
Intron 1	36819	ttAGGGACAGTTATGAAttcaatcttca_(SEQ ID NO:101)
Intron 1	36806	ttAGGGACAGTTATGAAttcaatcttca_(SEQ ID NO:101)

[0183] The rhesus albumin specific ZFNs were tested in pairs to determine the pair with the greatest activity. In each pair, SBS#35396 was tested with the potential partners shown in Tables 5 and 6 in the rhesus cell line RF/6A using the methods described above.

[0184] The resultant activity, as determined by percent of mismatch detected using the Cel-I assay is shown in the body of the matrix (Table 7), and demonstrate that the ZFNs pairs have activity against the rhesus albumin locus.

Table 7: Activity at the rhesus macaque albumin locus

	36813	36808	36820	36819	36806
35396	21%	26%	23%	30%	20.5%

[0185] Two pairs were examined more extensively, comparing sequence specificity by SELEX analysis and by a titration of each pair for activity *in vitro*. The results demonstrate that the 35396/36806 pair was the most desirable lead pair (see Figure 12).

[0186] Comparison of the sequence of the human albumin locus with the sequences of other non-human primates demonstrates that similar pairs may be developed for work in other primates such as cynologous monkeys (see, Figure 3A and 3B).

Example 5: *In vivo* cleavage by ZFNs in mice

[0187] To deliver the albumin-specific ZFNs to the liver *in vivo*, the normal site of albumin production, we generated a hepatotropic adeno-associated virus

vector, serotype 8 expressing the albumin-specific ZFNs from a liver-specific enhancer and promoter (Shen *et al*, *ibid* and Miao *et al, ibid*). Adult C57BL/6 mice were subjected to genome editing at the albumin gene as follows: adult mice were treated by i.v. (intravenous) injection with 1 x 10^{11} v.g. (viral genomes)/mouse of either ZFN pair 1 (SBS 30724 and SBS 30725), or ZFN pair 2 (SBS 30872 and SBS 30873) and sacrificed seven days later. The region of the albumin gene encompassing the target site for pair 1 was amplified by PCR for the Cel-I mismatch assay using the following 2 PCR primers:

Cel1 F1: 5' CCTGCTCGACCATGCTATACT 3' (SEQ ID NO:69)
Cel1R1: 5' CAGGCCTTTGAAATGTTGTTC 3' (SEQ ID NO:70)

[0188] The region of the albumin gene encompassing the target site for pair 2 was amplified by PCR for the Cel-I assay using these PCR primers:

mAlb set4F4: 5' AAGTGCAAAGCCTTTCAGGA 3' (SEQ ID NO:71)
mAlb set4R4: 5' GTGTCCTTGTCAGCAGCCTT 3' (SEQ ID NO:72)

[0189] As shown in Figure 4, the ZFNs induce indels in up to 17% of their target sites *in vivo* in this study.

[0190] The mouse albumin specific ZFNs SBS30724 and SBS30725 which target a sequence in intron 1 were also tested in a second study. Genes for expressing the ZFNs were introduced into an AAV2/8 vector as described previously (Li *et al* (2011) *Nature* 475 (7355): 217). To facilitate AAV production in the baculovirus system, a baculovirus containing a chimeric serotype 8.2 capsid gene was used. Serotype 8.2 capsid differs from serotype 8 capsid in that the phopholipase A2 domain in capsid protein VP1 of AAV8 has been replaced by the comparable domain from the AAV2 capsid creating a chimeric capsid. Production of the ZFN containing virus particles was done either by preparation using a HEK293 system or a baculovirus system using standard methods in the art (See Li *et al*, *ibid*, see *e.g.* US6723551). The virus particles were then administered to normal male mice (n=6) using a single dose of 200 microliter of 1.0e11 total vector genomes of either AAV2/8 or AAV2/8.2 encoding the mouse albumin-specific ZFN. 14 days post administration

of rAAV vectors, mice were sacrificed, livers harvested and processed for DNA or total proteins using standard methods known in the art. Detection of AAV vector genome copies was performed by quantitative PCR. Briefly, qPCR primers were made specific to the bGHpA sequences within the AAV as follows:

Oligo200 (Forward) 5'-GTTGCCAGCCATCTGTTGTTT-3' (SEQ ID NO:102)

Oligo201 (Reverse) 5'-GACAGTGGGAGTGGCACCTT-3' (SEQ ID NO:103)

Oligo202 (Probe) 5'-CTCCCCCGTGCCTTCCTTGACC-3'(SEQ ID NO:104)

[0191] Cleavage activity of the ZFN was measured using a Cel-I assay performed using a LC-GX apparatus (Perkin Elmer), according to manufacturer's protocol. Expression of the ZFNs *in vivo* was measured using a FLAG-Tag system according to standard methods.

[0192] As shown in Figure 5 (for each mouse in the study) the ZFNs were expressed, and cleave the target in the mouse liver gene. The % indels generated in each mouse sample is provided at the bottom of each lane. The type of vector and their contents are shown above the lanes. Mismatch repair following ZFN cleavage (indicated % indels) was detected at nearly 16% in some of the mice.

[0193] The mouse specific albumin ZFNs were also tested for *in vivo* activity when delivered via use of a variety of AAV serotypes including AAV2/5, AAV2/6, AAV2/8 and AAV2/8.2. In these AAV vectors, all the ZFN encoding sequence is flanked by the AAV2 ITRs, contain, and then encapsulated using capsid proteins from AAV5, 6, or 8, respectively. The 8.2 designation is the same as described above. The SBS30724 and SBS30725 ZFNs were cloned into the AAV as described previously (Li *et al, ibid*), and the viral particles were produced either using baculovirus or a HEK293 transient transfection purification as described above. Dosing was done in normal mice in a volume of 200 μL per mouse via tail injection, at doses from 5e10 to 1e12 vg per dose. Viral genomes per diploid mouse genome were analyzed at days 14, and are analyzed at days 30 and 60. In addition, ZFN directed cleavage of the albumin locus was analyzed by Cel-I assay as described previously at day 14 and is analyzed at days 30 and 60.

[0194] As shown in Figure 6, cleavage was observed at a level of up to 21% indels. Also included in Figure are the samples from the previous study as a comparison (far right, "mini-mouse" study-D14and a background band ("unspecific band").

Example 6: *In vivo* co-delivery of a donor nucleic acid and albumin ZFNs.

[0195] Insertion of human Factor IX: ZFNs were used to target integration of the gene for the clotting protein Factor IX (F.IX) into the albumin locus in adult wild-type mice. In these experiments, the mice were treated by I.V. injection with either 1 x 10^{11} v.g./mouse albumin-specific ZFN pair 1 targeting intron 1 + donor ("mAlb (intron1)"), 1 x 10^{11} v.g./mouse albumin-specific ZFN pair 2 targeting intron 12 + donor ("mAlb(intron12)") or a ZFN set that targets a human gene plus donor as a control ("Control"). The ZFN pair #1 was 30724/30725, targeting intron 1, and ZFN pair 2 was 30872/30873, targeting exon 12. In these experiments, the F.IX donor transgene was integrated via end capture following ZFN-induced cleavage. Alternatively, the F.IX transgene was inserted into a donor vector such that the transgene was flanked by arms with homology to the site of cleavage. In either case, the F.IX transgene was the "SA – wild-type hF9 exons 2-8" cassette (see co-owned US patent application 61/392,333).

[0196] Transduced mice were then sampled for serum human F.IX levels, which were elevated (see Figure 7, showing stabilized expression of human F.IX for at least eight weeks following insertion into intron 1). The expressed human F.IX is also functional, as evidenced by the reduction in clotting time in hemophilic mice with a human F.IX transgene targeted into the albumin locus (see Figure 8). Notably, within two weeks following transgene insertion, the clotting time is not significantly different than clotting time in a wild type mouse. When the intron 1 specific donor was inserted into the intron 12 locus, correct splicing to result in expression of the huF.IX cannot occur. The lack of signal in this sample verifies that the signal from the intron 1 donor being integrated into the intron 1 site is truly from correct transgene integration, and not from random integration and expression at another non-specific site.

[0197] Insertion of human alpha galactosidase (huGLa): Similar to the insertion of the human F.IX gene, the gene encoding human alpha galatosidase (deficient in patients with Fabry's disease) was inserted into the mouse albumin locus. The ZFN pair 30724/30725 was used as described above using an alpha galactosidase transgene in place of the F.IX transgene. In this experiment, 3 mice were treated with an AAV2/8 virus containing the ZFN pair at a dose of 3.0e11 viral genomes per mouse and an AAV2/8 virus containing the huGLa donor at 1.5e12 viral genomes per

mouse. Control animals were given either the ZFN containing virus alone or the huGLa donor virus alone. Western blots done on liver homogenates showed an increase in alpha galactosidase-specific signal, indicating that the alpha galactosidase gene had been integrated and was being expressed (Figure 13A). In addition, an ELISA was performed on the liver lysate using a human alpha galactosidase assay kit (Sino) according to manufacturer's protocol. The results, shown in Figure 13B, demonstrated an increase in signal in the mice that had been treated with both the ZFNs and the huGLa donor.

Example 7: Design of human albumin specific ZFNs.

[0198] To design ZFNs with specificity for the human albumin gene, the DNA sequence of human albumin intron 1 was analyzed using previously described methods to identify target sequences with the best potential for ZFN binding. Regions throughout the intron (loci 1-5) were chosen and several ZFNs were designed to target these regions region (for example, see Figure 9 which shows the binding sites of ZFNs from loci 1-3). In this analysis, five loci were identified to target in the albumin intron1 (see Figure 3B). The target and helices are shown in Tables 8 and 9.

Table 8: Human albumin-specific zinc finger nucleases helix designs

Target SBS #	Design					
	F1	F2	F3	F4	F5	F6
Intron 1 35393	QSSDLSR (SEQ ID NO:46)	LRHNLRA (SEQ ID NO:105)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)
Intron 1 35394	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)
Intron 1 35396	QSSDLSR (SEQ ID NO:46)	LKWNLRT (SEQ ID NO:109)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)
Intron 1 35398	QSSDLSR (SEQ ID NO:46)	LRHNLRA (SEQ ID NO:105)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)
Intron 1 35399	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)
Intron 1 35405	QSSDLSR (SEQ ID NO:46)	WKWNLRA (SEQ ID NO:110)	DQSNLRA (SEQ ID NO:106)	RPYTLRL (SEQ ID NO:107)	QSSDLSR (SEQ ID NO:46)	HRSNLNK (SEQ ID NO:108)

Intron 1 35361	QSGNLAR (SEQ ID NO:5)	LMQNRNQ (SEQ ID NO:9	LKQHLNE (SEQ ID NO:111)	TSGNLTR (SEQ ID NO:11)	RRYYLRL (SEQ ID NO:112)	N/A
Intron 1 35364	QSGNLAR (SEQ ID NO:5)	HLGNLKT (SEQ ID NO:94)	LKQHLNE (SEQ ID NO:111)	TSGNLTR (SEQ ID NO:11)	RRDWRRD (SEQ ID NO:113)	N/A
Intron 1 35370	QSGNLAR (SEQ ID NO:5)	LMQNRNQ (SEQ ID NO:9	LKQHLNE (SEQ ID NO:111)	TSGNLTR (SEQ ID NO:11)	RRDWRRD (SEQ ID NO:113)	N/A
Intron 1 35379	QRSNLVR (SEQ ID NO:99)	TSSNRKT (SEQ ID NO:53)	LKHHLTD (SEQ ID NO:95)	TSGNLTR (SEQ ID NO:11)	RRDWRRD (SEQ ID NO:113)	N/A
Intron 1 35458	DKSYLRP (SEQ ID NO:114)	TSGNLTR (SEQ ID NO:11)	HRSARKR (SEQ ID NO:115)	QSSDLSR (SEQ ID NO:46)	WRSSLKT (SEQ ID NO:116)	N/A
Intron 1 35480	TSGNLTR (SEQ ID NO:11)	HRSARKR (SEQ ID NO:115)	QSGDLTR (SEQ ID NO:40)	NRHHLKS (SEQ ID NO:116)	N/A	N/A
Intron 1 35426	QSGDLTR (SEQ ID NO:40)	QSGNLHV (SEQ ID NO:117)	QSAHRKN (SEQ ID NO:118)	STAALSY (SEQ ID NO:119)	TSGSLSR (SEQ ID NO:120)	RSDALAR (SEQ ID NO:41)
Intron 1 35428	QSGDLTR (SEQ ID NO:40)	QRSNLNI (SEQ ID NO:121)	QSAHRKN (SEQ ID NO:118)	STAALSY (SEQ ID NO:119)	DRSALSR (SEQ ID NO:52)	RSDALAR (SEQ ID NO:41)
Intron 1 34931	QRTHLTQ (SEQ ID NO:122)	DRSNLTR (SEQ ID NO:123)	QSGNLAR (SEQ ID NO:5)	QKVNRAG (SEQ ID NO:124)	N/A	N/A
Intron 1 33940	RSDNLSV (SEQ ID NO:26)	QNANRIT (SEQ ID NO:125)	DQSNLRA (SEQ ID NO:106)	QSAHRIT (SEQ ID NO:126)	TSGNLTR (SEQ ID NO:11)	HRSARKR (SEQ ID NO:115)

Table 9: Target sites of Human albumin-specific ZFNs

Target	SBS #	Target site
Intron 1 (locus 2)	35393	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35394	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35396	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35398	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35399	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35405	ccTATCCATTGCACTATGCTttatttaa (SEQ ID NO:127)
Intron 1 (locus 2)	35361	ttTGGGATAGTTATGAAttcaatcttca (SEQ ID NO:128)
Intron 1	35364	ttTGGGATAGTTATGAAttcaatcttca (SEQ ID NO:128)

(locus 2)		
Intron 1 (locus 2)	35370	ttTGGGATAGTTATGAAttcaatcttca (SEQ ID NO:128)
Intron 1 (locus 2)	35379	ttTGGGATAGTTATGAAttcaatcttca (SEQ ID NO:128)
Intron 1 (locus 3)	35458	ccTGTGCTGTTGATCTCataaatagaac (SEQ ID NO:129)
Intron 1 (locus 3)	35480	ccTGTGCTGTTGATctcataaatagaac (SEQ ID NO:129)
Intron 1 (locus 3)	35426	ttGTGGTTTTTAAAtAAAGCAtagtgca(SEQ ID NO:130)
Intron 1 (locus 3)	35428	ttGTGGTTTTTAAAtAAAGCAtagtgca(SEQ ID NO:130)
Intron 1 (locus 4)	34931	acCAAGAAGACAGActaaaatgaaaata (SEQ ID NO:131)
Intron 1 (locus 4)	33940	ctGTTGATAGACACTAAAAGagtattag (SEQ ID NO:132)

[0199] These nucleases were tested in pairs to determine the pair with the highest activity. The resultant matrices of tested pairs are shown in Tables 10 and 11, below where the ZFN used for the right side of the dimer is shown across the top of each matrix, and the ZFN used for the left side of the dimer is listed on the left side of each matrix. The resultant activity, as determined by percent of mismatch detected using the Cel-I assay is shown in the body of both matrices:

Table 10: Activity of Human albumin-specific ZFNs (%mutated targets)

	35393	35394	35396	35398	35399	35405	ave.
35361	18	19	25	22	23	21	21
35364	n.d.	24	23	19	21	21	22
35370	21	19	22	n.d.	22	23	21
35379	21	21	n.d.	19	19	21	20

Table 11: Activity of Human albumin-specific ZFNs (% mutated targets))

	35458	35480	ave.
35426	4.5	7	3
35428	4.9	6	3.6

(note: 'n.d.' means the assay on this pair was not done)

[0200] Thus, highly active nucleases have been developed that recognize target sequences in intron 1 of human albumin.

Example 8: Design of albumin specific TALENs

[0201] TALENs were designed to target sequences within human albumin intron 1. Base recognition was achieved using the canonical RVD-base correspondences (the "TALE code": NI for A, HD for C, NN for G (NK in half repeat), NG for T). TALENs were constructed as previously described (see co-owned U.S. Patent Publicaiton No. 20110301073). Targets for a subset of TALENs were conserved in cynomolgus monkey and rhesus macaque albumin genes (see figure 10). The TALENs were constructed in the "+17" and "+63" TALEN backbones as described in US20110301073. The targets and numeric identifiers for the TALENs tested are shown below in Table 12.

Table 12: Albumin specific TALENs

SBS #	site	# of RVDs	SEQ ID NO:
102249	gtTGAAGATTGAATTCAta	15	133
102250	gtTGAAGATTGAATTCATAac	17	133
102251	gtGCAATGGATAGGTCTtt	15	134
102252	atAGTGCAATGGATAGGtc	15	135
102253	atTGAATTCATAACTATcc	15	136
102254	atTGAATTCATAACTATCCca	17	137
102255	atAAAGCATAGTGCAATGGat	17	138
102256	atAAAGCATAGTGCAATgg	15	139
102257	ctATGCTTTATTTAAAAac	15	140
102258	ctATGCTTTATTTAAAAACca	17	141
102259	atTTATGAGATCAACAGCAca	17	142
102260	ctATTTATGAGATCAACAGca	17	158
102261	ttCATTTTAGTCTGTCTTCtt	17	143
102262	atTTTAGTCTGTCTTCTtg	15	144
102263	ctAATACTCTTTTAGTGTct	16	145
102264	atCTAATACTCTTTTAGTGtc	17	146
102265	atAATTGAACATCATCCtg	15	147
102266	atAATTGAACATCATCCTGag	17	148
102267	atATTGGGCTCTGATTCCTac	17	149
102268	atATTGGGCTCTGATTCct	15	150
102269	ttTTTCTGTAGGAATCAga	15	159
102270	ttTTTCTGTAGGAATCAGag	16	151
102271	ttATGCATTTGTTTCAAaa	15	152
102272	atTATGCATTTGTTTCAaa	15	153

[0202] The TALENs were then tested in pairs in HepG2 cells for the ability to induce modifications at their endogenous chromosomal targets, and the results

showed that many proteins bearing the +17 truncation point were active. Similarly, many TALENs bearing the +63 truncation point were also active (see Table 13 and Figure 11). Note that the pair numbers shown in Table 13 correspond with the pair numbers shown above the lanes in Figure 11. Side by side comparisons with three sets of non-optimized albumin ZFNs showed that the TALENs and ZFNs have activities that are in the same approximate range.

Table 13: TALEN-induced target modification in HepG2-C3a cells

Sample pair	TALEN C17	%modification, C17	TALEN C63	% modification, C63	Gap
1	102251:102249	15	102251:102249	0	12
2	102251:102250	0	102251:102250	0	10
3	102252:102249	0	102252:102249	8.3	15
4	102252:102250	32	102252:102250	8.0	13
5	102255:102253	38	102255:102253	21	13
6	102255:102254	43	102255:102254	0	11
7	102256:102253	0	102256:102253	23	15
8	102256:102254	28	102256:102254	16	13
9	102259:102257	18	102259:102257	15	13
10	102259:102258	15	102259:102258	0	11
11	102260:102257	15	102260:102257	13	15
12	102260:102258	24	102260:102258	11	13
13	102263:102261	0	102263:102261	16	17
14	102263:102262	0	102263:102262	15	16
15	102264:102261	0	102264:102261	22	18
16	102264:102262	0	102264:102262	17	17
20	102267:102265	47	102267:102265	9.8	13
21	102267:102266	4.7	102267:102266	0	11
22	102268:102265	4.2	102268:102265	7.9	15
23	102268:102266	10	102268:102266	0	13
24	102271:102269	14	102271:102269	0	12
25	102271:102270	0	102271:102270	0	11
26	102272:102269	0	102272:102269	0	13
27	102272:102270	0	102272:102270	0	12
ZFNs					
17	35361:35396	31	35361:35396	29	6
18	35426:35458	10	35426:35458	7	6
19	34931:33940	7.3	34931:33940	7	6

[0203] As noted previously (see co-owned U.S. Patent Publication No. 20110301073), the C17 TALENs have greater activity when the gap size between the

two TALEN target sites is approximately 11- 15 bp, while the C63 TALENs sustain activity at gap sizes up to 18 bp (see Figure 10, 11C and Table 13).

[0204] All patents, patent applications and publications mentioned herein are hereby incorporated by reference in their entirety.

[0205] Although disclosure has been provided in some detail by way of illustration and example for the purposes of clarity of understanding, it will be apparent to those skilled in the art that various changes and modifications can be practiced without departing from the spirit or scope of the disclosure. Accordingly, the foregoing descriptions and examples should not be construed as limiting.

CLAIMS

What is claimed is:

1. A non-naturally occurring fusion protein comprising a zinc finger protein or a TAL-effector domain protein that binds to an endogenous albumin gene and a cleavage domain, wherein the fusion protein modifies the endogenous albumin gene, and wherein the zinc finger protein comprises 4, 5 or 6 zinc finger domains comprising a recognition helix region, wherein the zinc finger proteins comprise the recognition helix regions shown in a single row of Table 1, Table 3, Table 5 or Table 8 or the TAL-effector domain protein that binds to a target site as shown in Table 12.

2. A polynucleotide encoding one or more fusion proteins of claim 1.

3. An isolated cell comprising one or more fusion proteins according to claim 1 or one or more polynucleotides according to claim 2.

4. The cell of claim 3, wherein the cell is a stem cell.

5. The cell of claim 4, wherein the stem cell is selected from the group consisting of an embryonic stem cell (ESC), an induced pluripotent stem cell (iPSC), a hepatic stem cell and a liver stem cell.

6. A kit comprising a fusion protein according to claim 1 or a polynucleotide according to claim 2.

7. An *in vitro* method of cleaving an endogenous albumin gene in a cell, the method comprising:

introducing, into the cell, one or more expression vectors comprising at least one fusion protein according to claim 1 or at least one polynucleotide according to claim 2, under conditions such that the one or more fusion proteins are expressed and the albumin gene is cleaved.

8. The method of claim 7, wherein the polynucleotide comprises an AAV vector.

9. The method of claim 7 or claim 8, wherein the cell is a liver cell.

10. A non-naturally occurring fusion protein of claim 1 or a polynucleotide of claim 2 for use in therapy.

11. A non-naturally occurring fusion protein of claim 1 or a polynucleotide of claim 2, and a donor polynucleotide comprising a transgene encoding a therapeutic protein for use in therapy.

12. Use of one or more expression vectors comprising at least one fusion protein according to claim 1 or at least one polynucelotide according to claim 2 in the manufacture of a medicament for treating a lysosomal storage disease, epidermolysis bullosa, AAT deficient emphysema or a blood disorder, wherein the vector or the polynucleotide is introduced into the cell under conditions such that the one or more fusion proteins are expressed and an endogenous albumin gene is cleaved.

13. The use of claim 12, wherein the blood disorder is a clotting disorder.

14. The use of claim 12 or 13, wherein the medicament further comprises a transgene encoding a therapeutic protein, wherein the transgene is integrated into the albumin gene following cleavage of the albumin gene by a nuclease.

15. The use of claim 14, wherein expression of the transgene is driven by control sequences of the endogenous albumin.

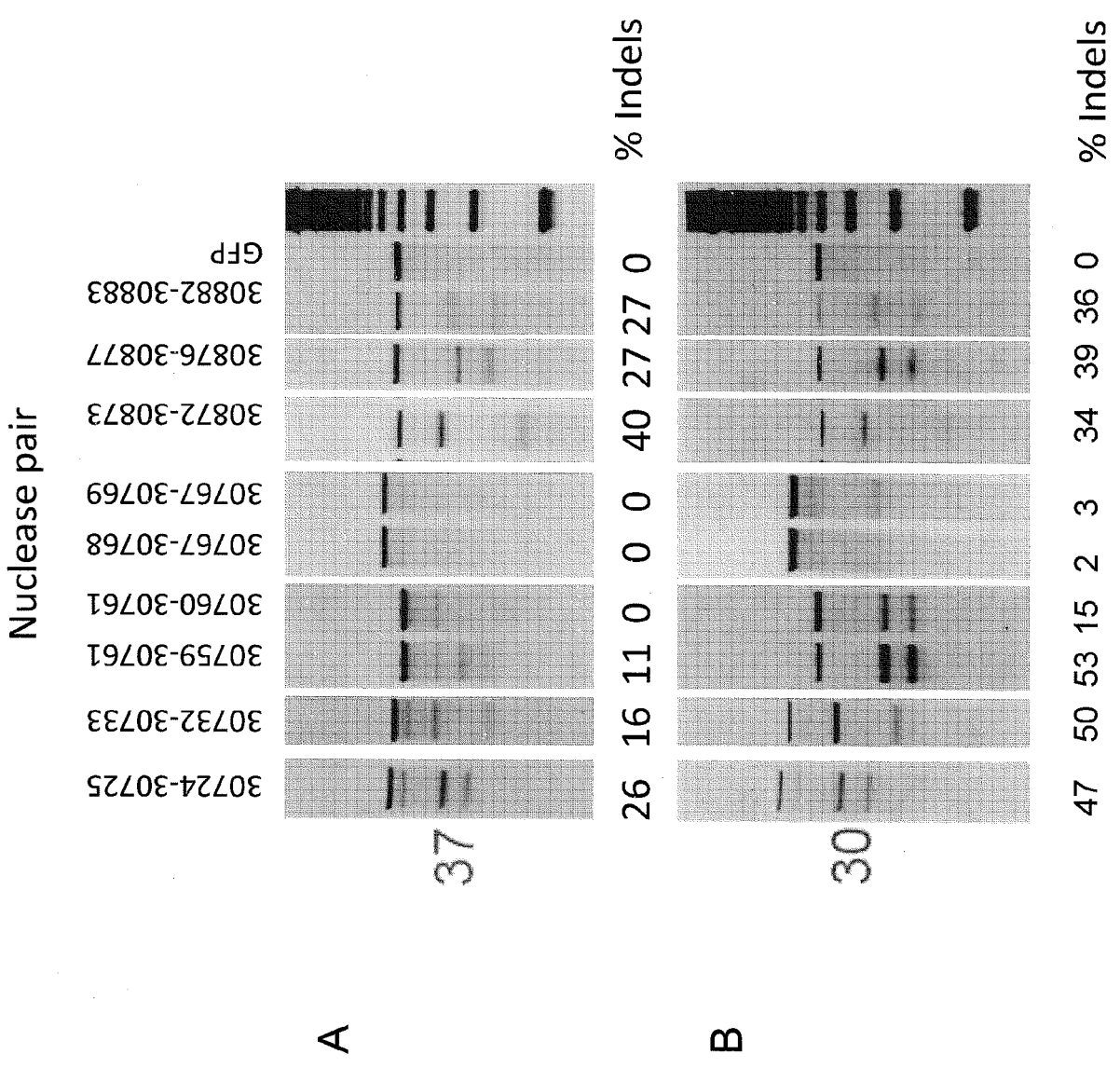

A
B
Nuclease pair
30724-30725
30732-30733
30759-30761
30760-30761
30767-30768
30767-30769
30872-30873
30876-30877
30882-30883
GFP
% Indels
26 16 11 0 0 0 40 27 27 0
% Indels
47 50 53 15 2 3 34 39 36 0
37
30

Figure 1

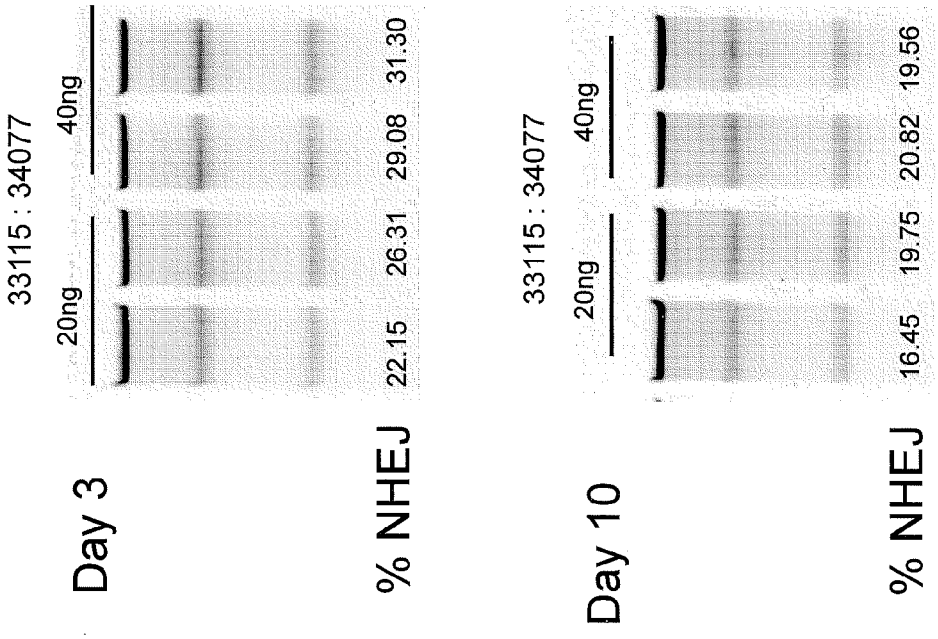

A
Day 3
33115 : 34077
20ng
40ng
% NHEJ
22.15
26.31
29.08
31.30
B
Day 10
33115 : 34077
20ng
40ng
% NHEJ
16.45
19.75
20.82
19.56

Figure 2

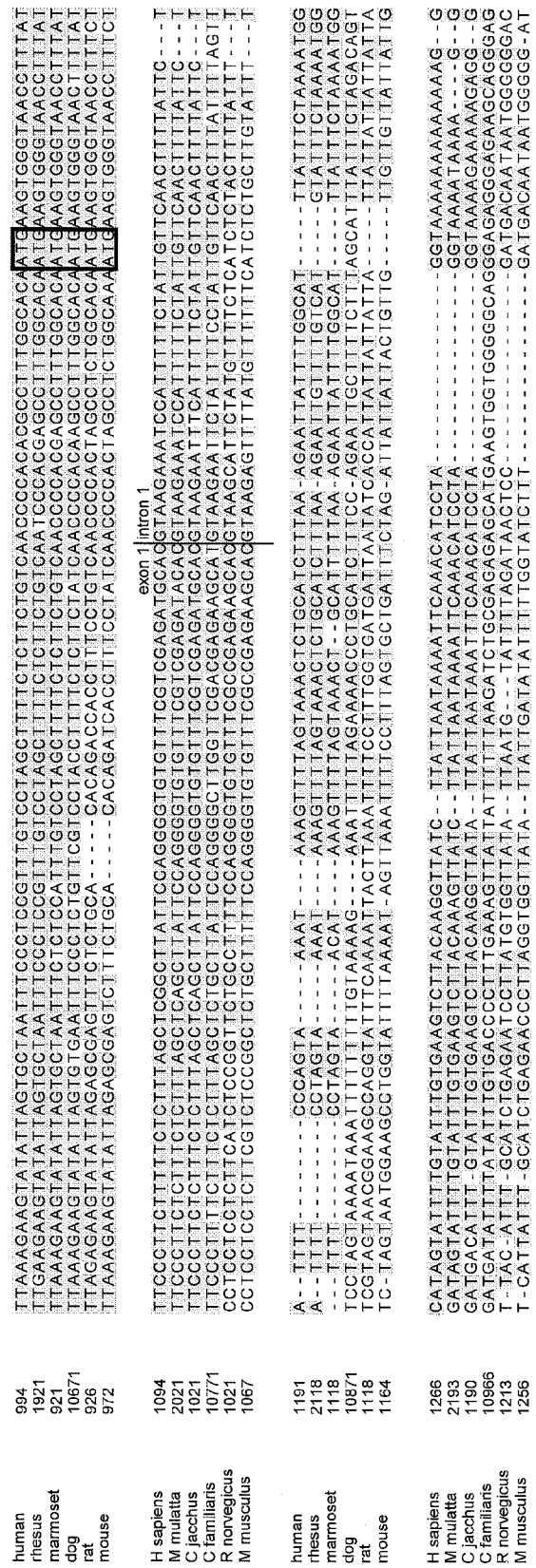


Figure 3A

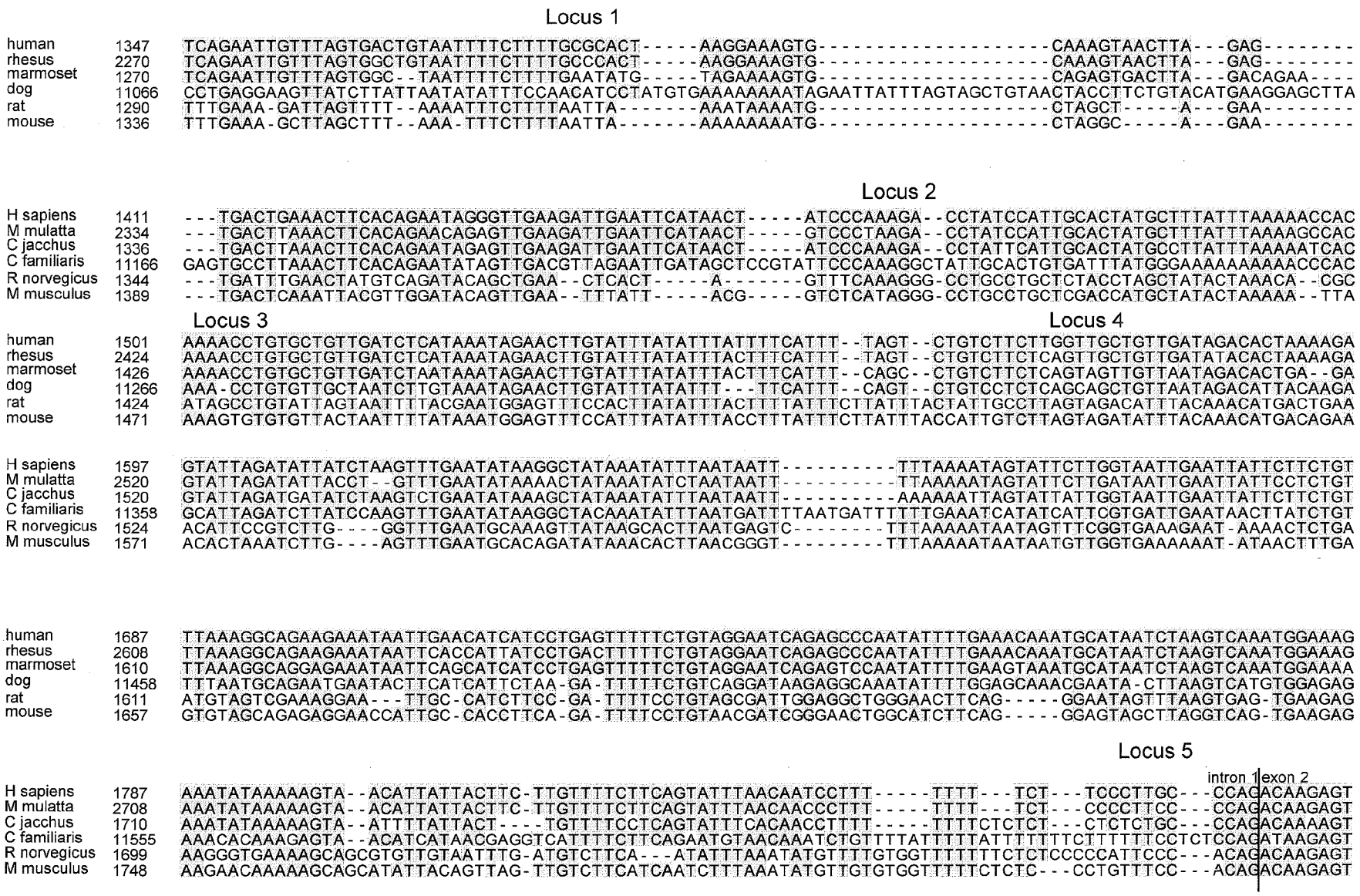

Locus 1
human 1347 TCAGAATTGTTTAGTGACTGTAATTTTCTTTTGCGCACT-----AAGGAAAGTG----------CAAAGTAACTTA---GAG-------
rhesus 2270 TCAGAATTGTTTAGTGGCTGTAATTTTCTTTTGCCCACT-----AAGGAAAGTG----------CAAAGTAACTTA---GAG-------
marmoset 1270 TCAGAATTGTTTAGTGGC--TAATTTTCTTTTGAATATG-----TAGAAAAGTG----------CAGAGTGACTTA---GACAGAA----
dog 11066 CCTGAGGAAGTTATCTTATTAATATATTTCCAACATCCTATGTGAAAAAAAATAGAATTATTTAGTAGCTGTAACTACCTTCTGTACATGAAGGAGCTTA
rat 1290 TTTGAAA-GATTAGTTTT--AAAATTTCTTTTAATTA-------AAATAAAATG----------CTAGCT-----A---GAA-------
mouse 1336 TTTGAAA-GCTTAGCTTT--AAA-TTTCTTTTAATTA-------AAAAAAAATG----------CTAGGC-----A---GAA-------
Locus 2
H sapiens 1411 ---TGACTGAAACTTCACAGAATAGGGTTGAAGATTGAATTCATAACT-----ATCCCAAAGA--CCTATCCATTGCACTATGCTTTATTTAAAAACCAC
M mulatta 2334 ---TGACTTAAACTTCACAGAACAGAGTTGAAGATTGAATTCATAACT-----GTCCCTAAGA--CCTATCCATTGCACTATGCTTTATTTAAAAGCCAC
C jacchus 1336 ---TGACTTAAACTTCACAGAATAGAGTTGAAGATTGAATTCATAACT-----ATCCCAAAGA--CCTATTCATTGCACTATGCCTTATTTAAAAATCAC
C familiaris 11166 GAGTGCCTTAAACTTCACAGAATATAGTTGACGTTAGAATTGATAGCTCCGTATTCCCAAAGGCTATTGCACTGTGATTTATGGGAAAAAAAAAACCCAC
R norvegicus 1344 ---TGATTTGAACTATGTCAGATACAGCTGAA--CTCACT-----A-------GTTTCAAAGGG-CCTGCCTGCTCTACCTAGCTATACTAAACA--CGC
M musculus 1389 ---TGACTCAAATTACGTTGGATACAGTTGAA--TTTATT-----ACG-----GTCTCATAGGG-CCTGCCTGCTCGACCATGCTATACTAAAAA--TTA
Locus 3
Locus 4
human 1501 AAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTATTTTCATTT--TAGT--CTGTCTTCTTGGTTGCTGTTGATAGACACTAAAAGA
rhesus 2424 AAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTACTTTCATTT--TAGT--CTGTCTTCTCAGTTGCTGTTGATATACACTAAAAGA
marmoset 1426 AAAACCTGTGCTGTTGATCTAATAAATAGAACTTGTATTTATATTTACTTTCATTT--CAGC--CTGTCTTCTCAGTAGTTGTTAATAGACACTGA--GA
dog 11266 AAA-CCTGTGTTGCTAATCTTGTAAATAGAACTTGTATTTATATTT---TTCATTT--CAGT--CTGTCCTCTCAGCAGCTGTTAATAGACATTACAAGA
rat 1424 ATAGCCTGTATTAGTAATTTTACGAATGGAGTTTCCACTTATATTTACTTTTATTTCTTATTTACTATTGCCTTAGTAGACATTTACAAACATGACTGAA
mouse 1471 AAAGTGTGTGTTACTAATTTTATAAATGGAGTTTCCATTTATATTTACCTTTATTTCTTATTTACCATTGTCTTAGTAGATATTTACAAACATGACAGAA
H sapiens 1597 GTATTAGATATTATCTAAGTTTGAATATAAGGCTATAAATATTTAATAATT----------TTTAAAATAGTATTCTTGGTAATTGAATTATTCTTCTGT
M mulatta 2520 GTATTAGATATTACCT--GTTTGAATATAAAACTATAAATATCTAATAATT----------TTAAAAATAGTATTCTTGATAATTGAATTATTCCTCTGT
C jacchus 1520 GTATTAGATGATATCTAAGTCTGAATATAAAGCTATAAATATTTAATAATT----------AAAAAATTAGTATTATTGGTAATTGAATTATTCTTCTGT
C familiaris 11358 GCATTAGATCTTATCCAAGTTTGAATATAAGGCTACAAATATTTAATGATTTTAATGATTTTTTGAAATCATATCATTCGTGATTGAATAACTTATCTGT
R norvegicus 1524 ACATTCCGTCTTG----GGTTTGAATGCAAAGTTATAAGCACTTAATGAGTC--------TTTAAAAATAATAGTTTCGGTGAAAGAAT-AAAACTCTGA
M musculus 1571 ACACTAAATCTTG----AGTTTGAATGCACAGATATAAACACTTAACGGGT---------TTTAAAAATAATAATGTTGGTGAAAAAAT-ATAACTTTGA
human 1687 TTAAAGGCAGAAGAAATAATTGAACATCATCCTGAGTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAATGGAAAG
rhesus 2608 TTAAAGGCAGAAGAAATAATTCACCATTATCCTGACTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAATGGAAAG
marmoset 1610 TTAAAGGCAGGAGAAATAATTCAGCATCATCCTGAGTTTTTCTGTAGGAATCAGAGTCCAATATTTTGAAGTAAATGCATAATCTAAGTCAAATGGAAAA
dog 11458 TTTAATGCAGAATGAATACTTCATCATTCTAA-GA-TTTTTCTGTCAGGATAAGAGGCAAATATTTTGGAGCAAACGAATA-CTTAAGTCATGTGGAGAG
rat 1611 ATGTAGTCGAAAGGAA---TTGC-CATCTTCC-GA-TTTTCCTGTAGCGATTGGAGGCTGGGAACTTCAG-----GGAATAGTTTAAGTGAG-TGAAGAG
mouse 1657 GTGTAGCAGAGAGGAACCATTGC-CACCTTCA-GA-TTTTCCTGTAACGATCGGGAACTGGCATCTTCAG-----GGAGTAGCTTAGGTCAG-TGAAGAG
Locus 5
intron 1
exon 2
H sapiens 1787 AAATATAAAAAGTA--ACATTATTACTTC-TTGTTTTCTTCAGTATTTAACAATCCTTT-----TTTT---TCT---TCCCTTGC---CCAGACAAGAGT
M mulatta 2708 AAATATAAAAAGTA--ACATTATTACTTC-TTGTTTTCTTCAGTATTTAACAACCCTTT-----TTTT---TCT---CCCCTTCC---CCAGACAAGAGT
C jacchus 1710 AAATATAAAAAGTA--ATTTTATTACT----TGTTTTCCTCAGTATTTCACAACCTTTT-----TTTTCTCTCT---CTCTCTGC---CCAGACAAAAGT
C familiaris 11555 AAACACAAAGAGTA--ACATCATAACGAGGTCATTTTCTTCAGAATGTAACAAATCTGTTTTATTTTTATTTTTTTCTTTTTTCCTCTCCAGATAAGAGT
R norvegicus 1699 AAGGGTGAAAAGCAGCGTGTTGTAATTTG-ATGTCTTCA---ATATTTAAATATGTTTTGTGGTTTTTTTCTCTCCCCCATTCCC---ACAGACAAGAGT
M musculus 1748 AAGAACAAAAAGCAGCATATTACAGTTAG-TTGTCTTCATCAATCTTTAAATATGTTGTGTGGTTTTTCTCTC---CCTGTTTCC---ACAGACAAGAGT

Figure 3B

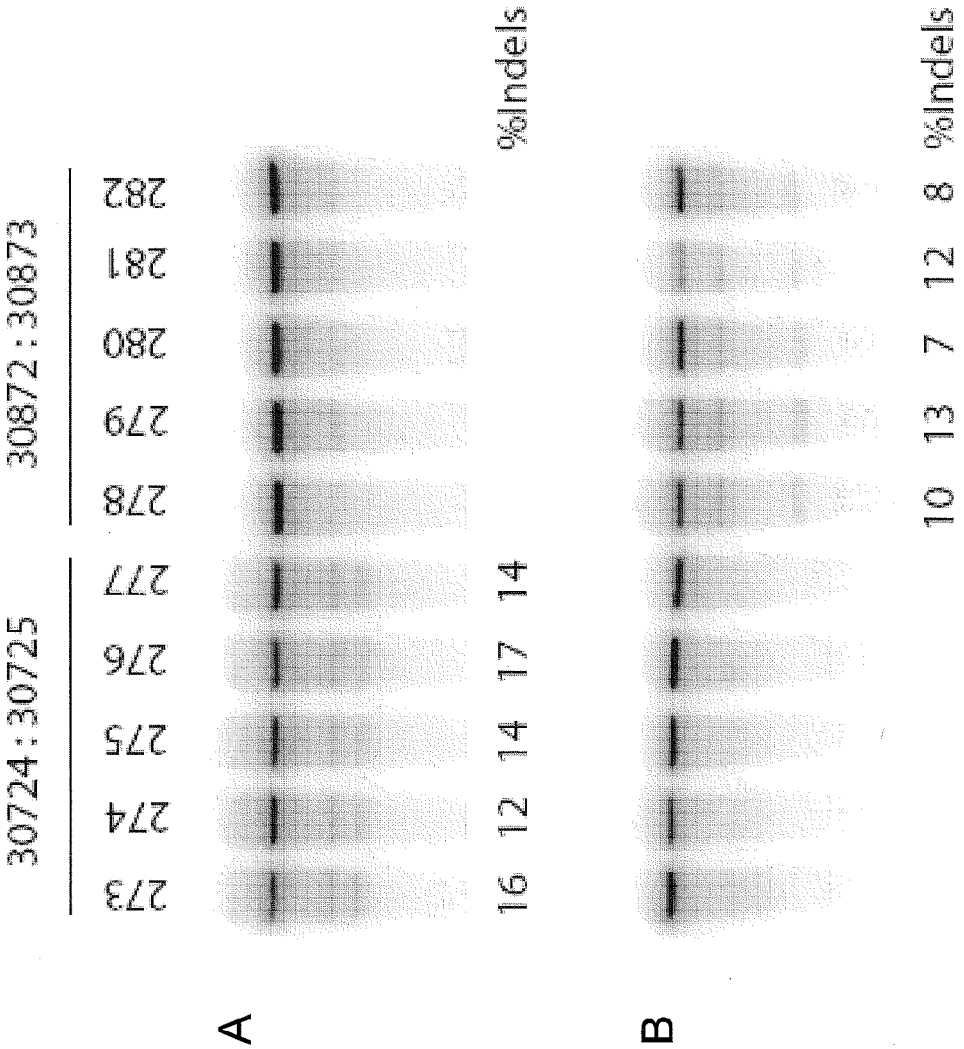

A
B
30724:30725
30872:30873
273
274
275
276
277
278
279
280
281
282
%Indels
16 12 14 17 14
10 13 7 12 8

Figure 4

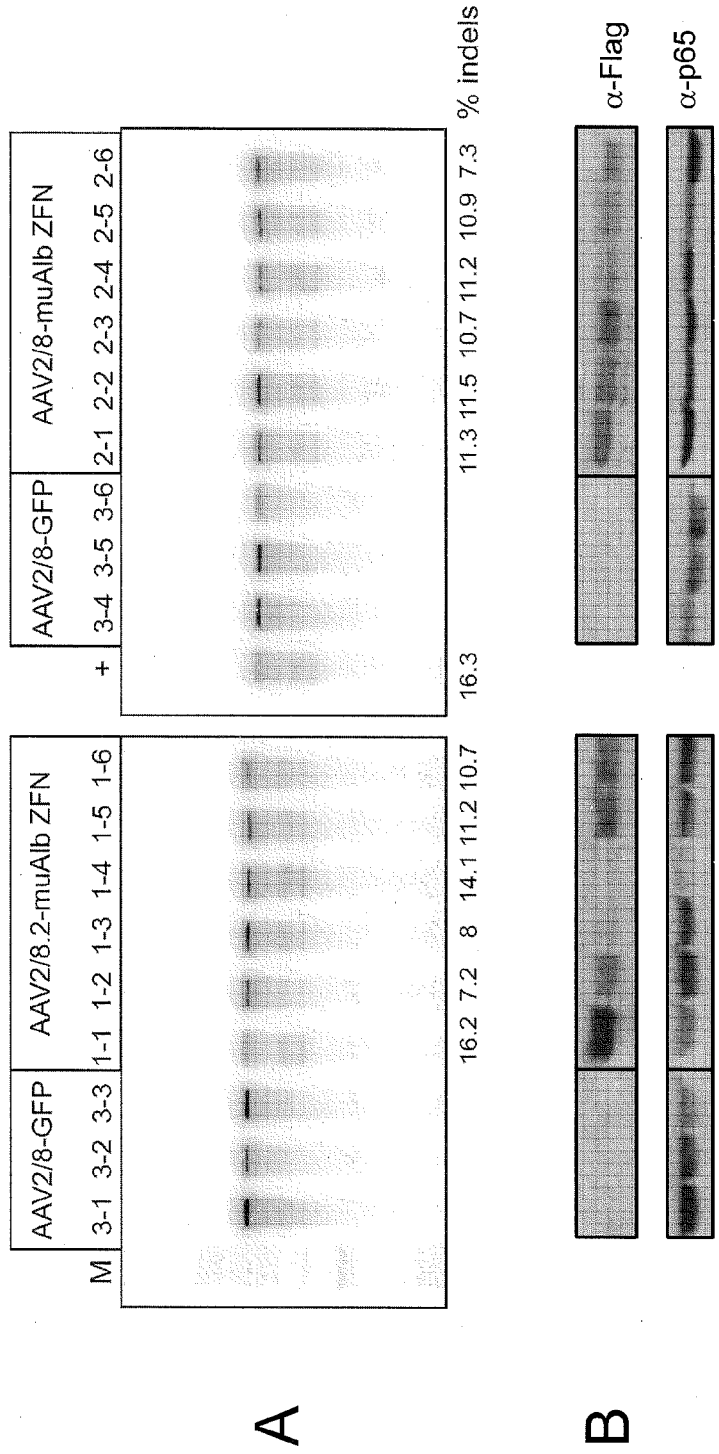

A
M
AAV2/8-GFP
3-1 3-2 3-3
AAV2/8.2-muAlb ZFN
1-1 1-2 1-3 1-4 1-5 1-6
16.2 7.2 8 14.1 11.2 10.7
+
AAV2/8-GFP
3-4 3-5 3-6
AAV2/8-muAlb ZFN
2-1 2-2 2-3 2-4 2-5 2-6
16.3 11.3 11.5 10.7 11.2 10.9 7.3
% indels
B
α-Flag
α-p65

Figure 5

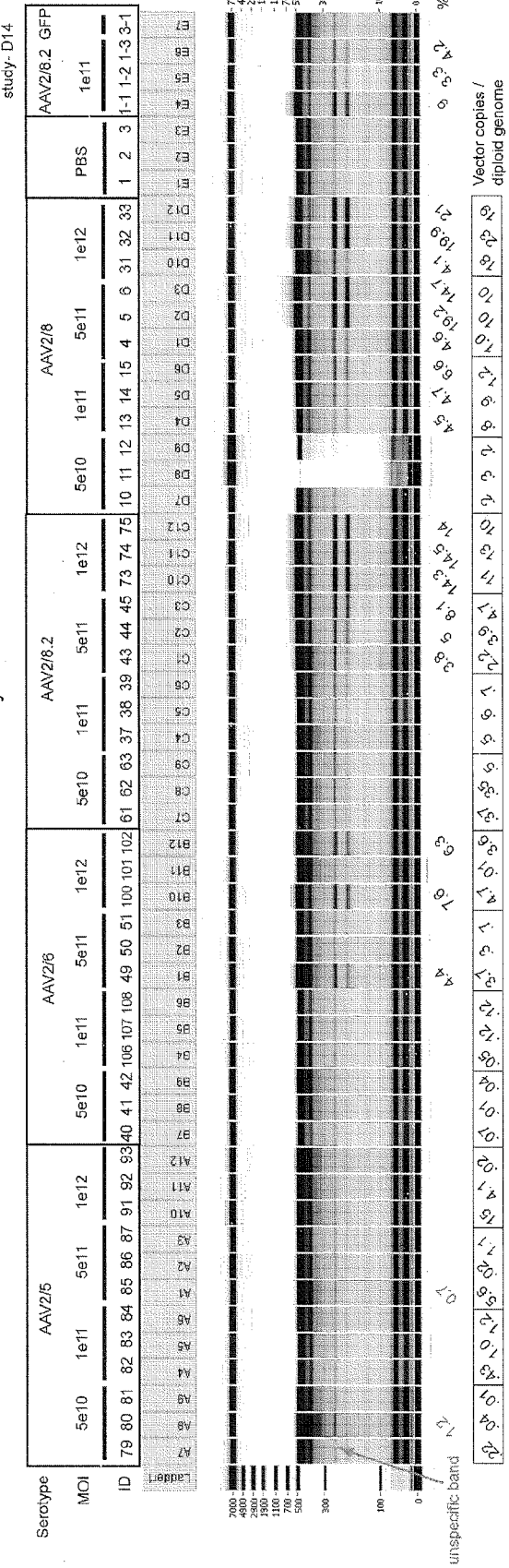

IVS19321 - mouse study - D14
Mini-mouse study- D14
Serotype
MOI
ID
AAV2/5
AAV2/6
AAV2/8.2
AAV2/8
PBS
AAV2/8.2 GFP
5e10
1e11
5e11
1e12
Ladder
unspecific band
% indels
Vector copies / diploid genome

Figure 6

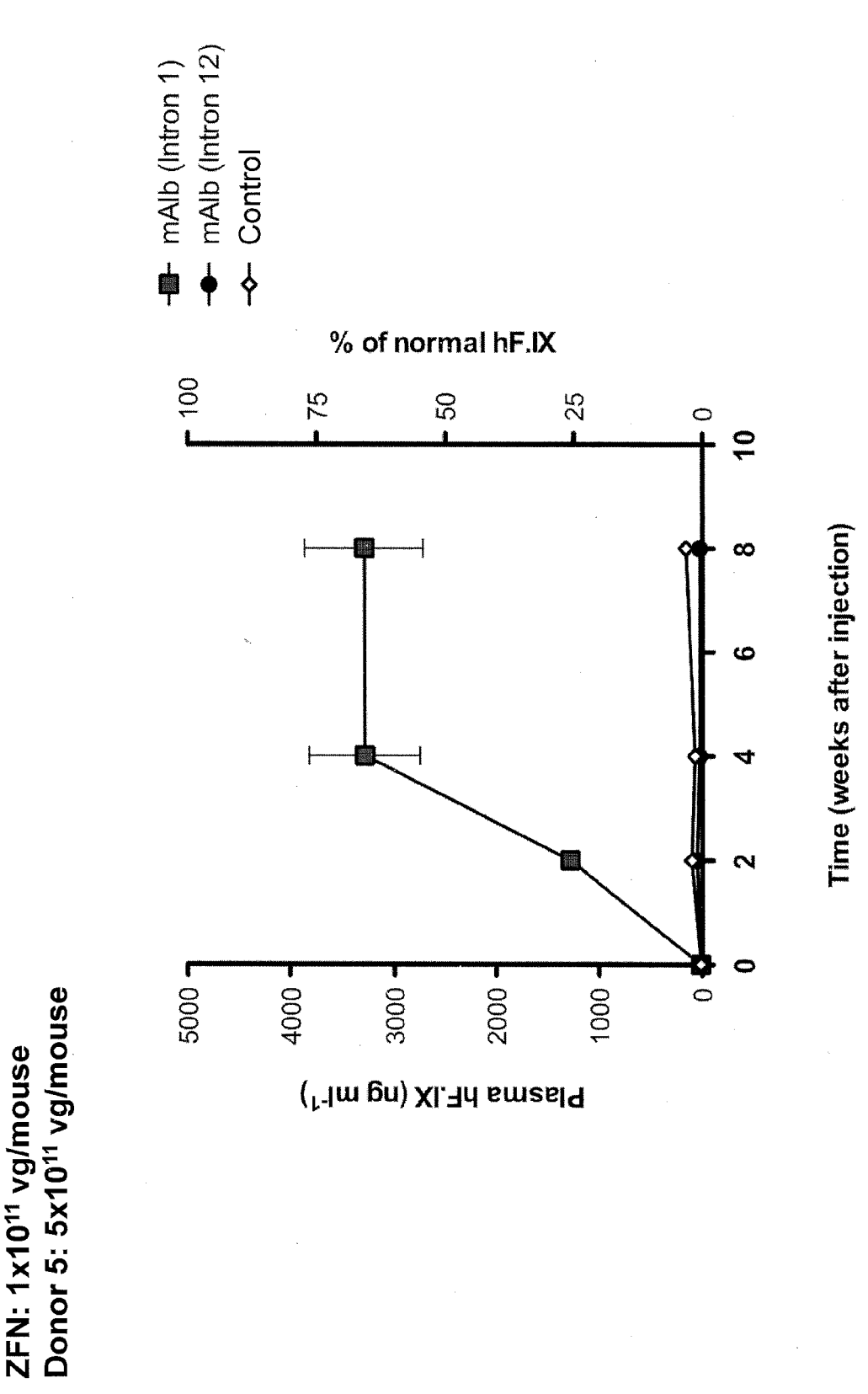
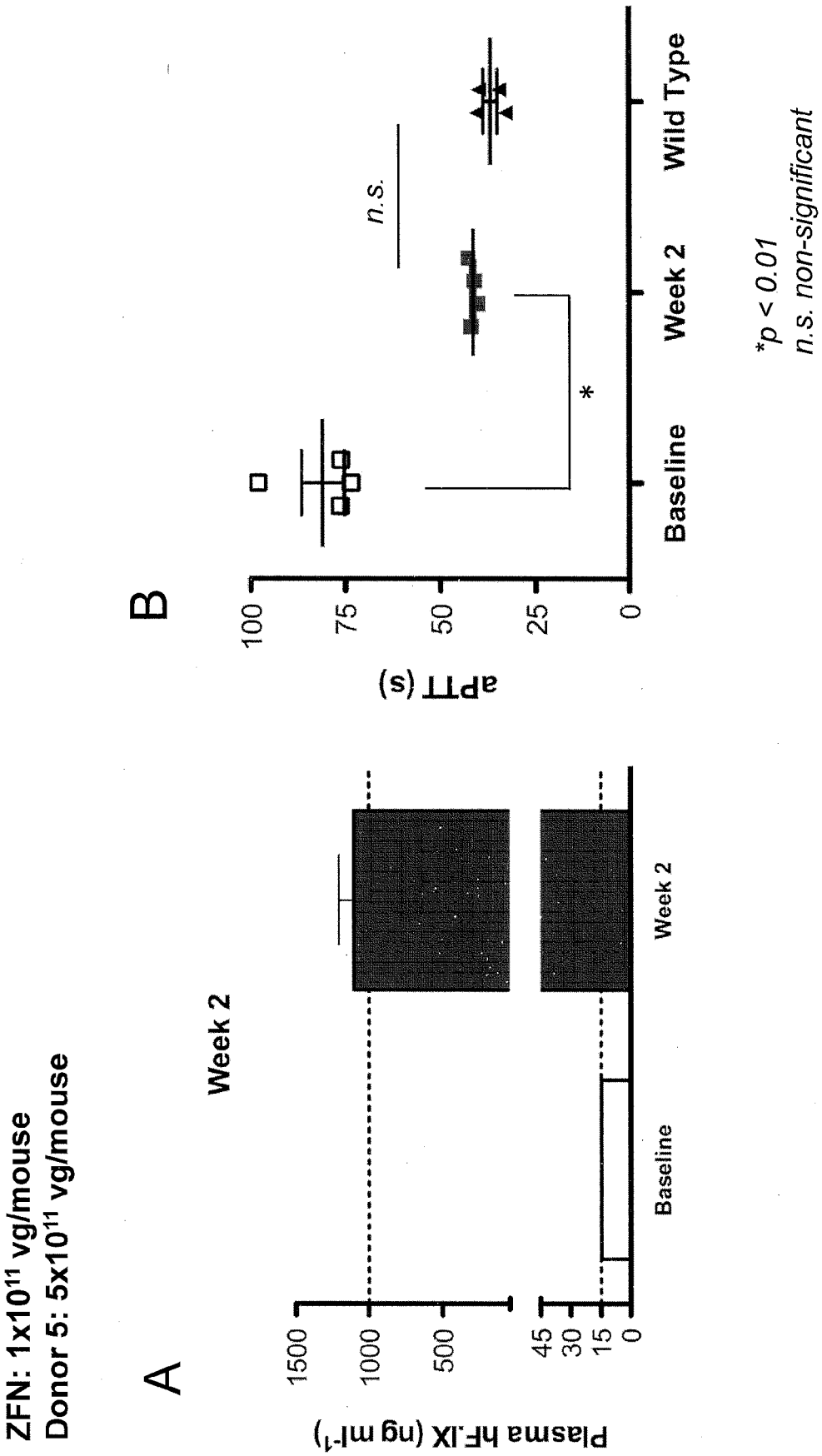

ZFN: 1x10^11 vg/mouse
Donor 5: 5x10^11 vg/mouse
mAlb (Intron 1)
mAlb (Intron 12)
Control
% of normal hF.IX
100
75
50
25
0
Plasma hF.IX (ng ml-1)
5000
4000
3000
2000
1000
0
0
2
4
6
8
10
Time (weeks after injection)

Figure 7



*p < 0.01
n.s. non-significant

Figure 8

HUMAN ALBUMIN, EXON 1/INTRON 1 BORDER REGION

EXON 1 | INTRON 1

1141 GTCGAGATGCAC|GTAAGAAA TCCATTTTTC TATTGTTCAA CTTTTATTCT ATTTTCCCAG

1201 TAAAATAAAG TTTTAGTAAA CTCTGCATCT TTAAAGAATT ATTTTGGCAT TTATTTCTAA

1261 AATGGCATAG TATTTTGTAT TTGTGAAGTC TTACAAGGTT ATCTTATTAA TAAAATTCAA

1321 ACATCCTAGG TAAAAAAAAA AAAAGGTCAG AATTGTTTAG TGACTGTAAT TTTCTTTTGC

1381 GCACTAAGGA AAGTGCAAAG TAACTTAGAG TGACTGAAAC TTCACAGAAT AGGGTTGAAG

35361, 35364, 35370, 35379 35426, 35428

35393, 35394, 35396, 35398, 35399, 35405

1441 ATTGAATTCA TAACTATCCC AAAGACCTAT CCATTGCACT ATGCTTTATT TAAAAACCAC

35458, 35480

1501 AAAACCTGTG CTGTTGATCT CATAAATAGA ACTTGTATTT ATATTTATTT TCATTTTAGT

2101 ATGGCTATTG AGTACTTCAA ATATGACAAG TGCAACTGAG AAACAAAAAC TTAAATTGTA

Figure 9

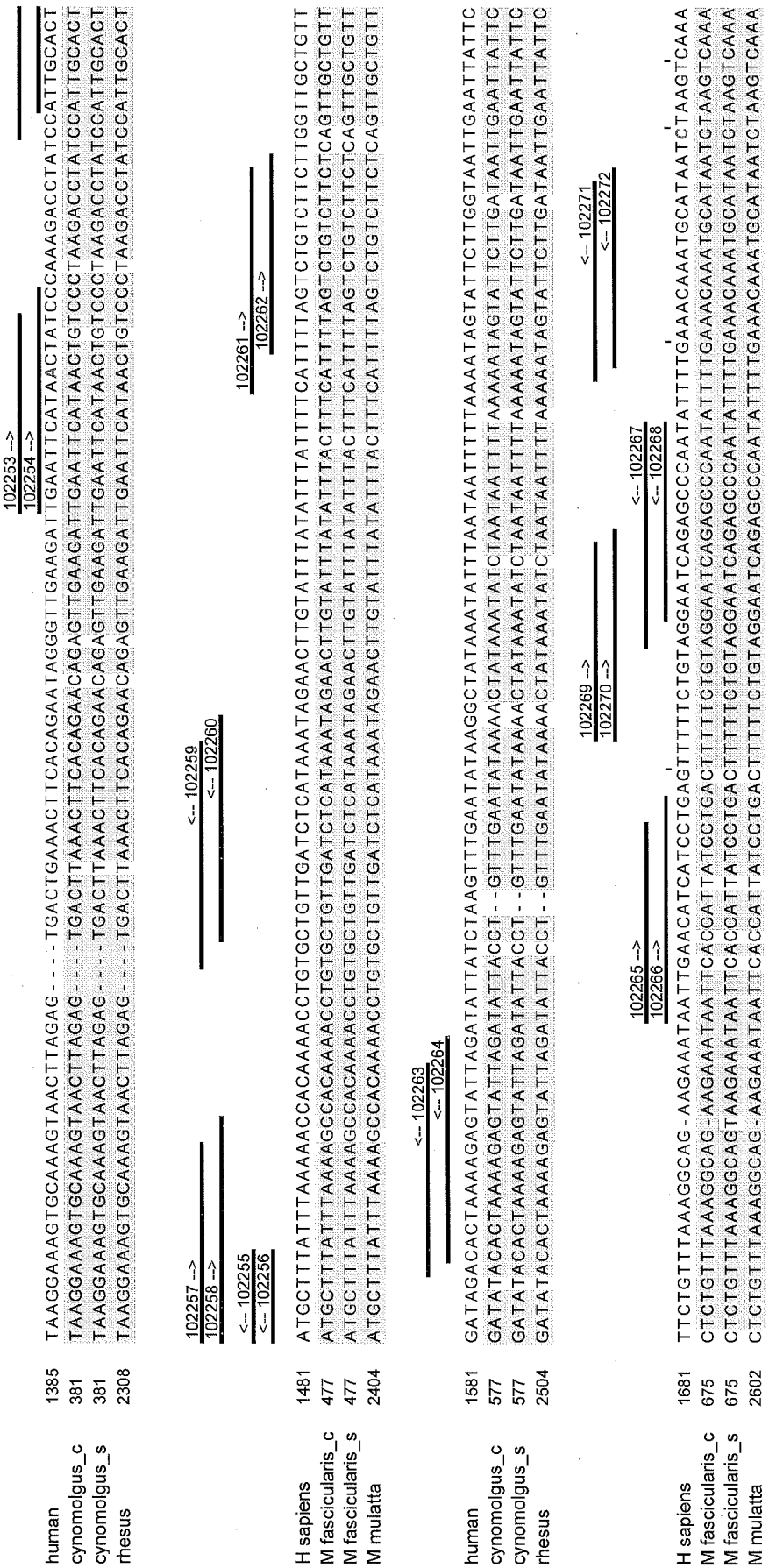

102249 -->
102250 -->
<-- 102251
<-- 102252
102253 -->
102254 -->
human 1385 TAAGGAAAGTGCAAAGTAACTTAGAG----TGACTGAAACTTCACAGAATAGGGTTGAAGATTGAATTCATAACTATCCCAAAGACCTATCCATTGCACT
cynomolgus_c 381 TAAGGAAAGTGCAAAGTAACTTAGAG----TGACTTAAACTTCACAGAACAGAGTTGAAGATTGAATTCATAACTGTCCCTAAGACCTATCCATTGCACT
cynomolgus_s 381 TAAGGAAAGTGCAAAGTAACTTAGAG----TGACTTAAACTTCACAGAACAGAGTTGAAGATTGAATTCATAACTGTCCCTAAGACCTATCCATTGCACT
rhesus 2308 TAAGGAAAGTGCAAAGTAACTTAGAG----TGACTTAAACTTCACAGAACAGAGTTGAAGATTGAATTCATAACTGTCCCTAAGACCTATCCATTGCACT
102257 -->
102258 -->
<-- 102259
<-- 102260
<-- 102255
<-- 102256
102261 -->
102262 -->
H sapiens 1481 ATGCTTTATTTAAAAACCACAAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTTCATTTTAGTCTGTCTTCTTGGTTGCTGTT
M fascicularis_c 477 ATGCTTTATTTAAAAGGCACAAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTTACTTTCATTTTAGTCTGTCTTCTCAGTTGCTGTT
M fascicularis_s 477 ATGCTTTATTTAAAAGCCACAAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTTACTTTCATTTTAGTCTGTCTTCTCAGTTGCTGTT
M mulatta 2404 ATGCTTTATTTAAAAGCCACAAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTATTTATATTTTACTTTCATTTTAGTCTGTCTTCTCAGTTGCTGTT
<-- 102263
<-- 102264
human 1581 GATAGACACTAAAAGAGTATTATCTAAGTTTGAATATAAGGCTATAAATATTTAATAATTTTTAAAATAGTATTCTTGGTAATTGAATTATTC
cynomolgus_c 577 GATATACACTAAAAGAGTATTAGATATTACCT--GTTTGAATATAAAACTATAAATATCTAATAATTTTTAAAAATAGTATTCTTGATAATTGAATTATTC
cynomolgus_s 577 GATATACACTAAAAGAGTATTAGATATTACCT--GTTTGAATATAAAACTATAAATATCTAATAATTTTTAAAAATAGTATTCTTGATAATTGAATTATTC
rhesus 2504 GATATACACTAAAAGAGTATTAGATATTACCT--GTTTGAATATAAAACTATAAATATCTAATAATTTTTAAAAATAGTATTCTTGATAATTGAATTATTC
102269 -->
102270 -->
<-- 102271
<-- 102272
102265 -->
102266 -->
<-- 102267
<-- 102268
H sapiens 1681 TTCTGTTTAAAGGCAG-AAGAAATAATTGAACATCATCCTGAGTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAA
M fascicularis_c 675 CTCTGTTTAAAGGCAG-AAGAAATAATTCACCATTATCCTGACTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAA
M fascicularis_s 675 CTCTGTTTAAAGGCAGTAAGAAATAATTCACCATTATCCTGACTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAA
M mulatta 2602 CTCTGTTTAAAGGCAG-AAGAAATAATTCACCATTATCCTGACTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTCAAA

Figure 10

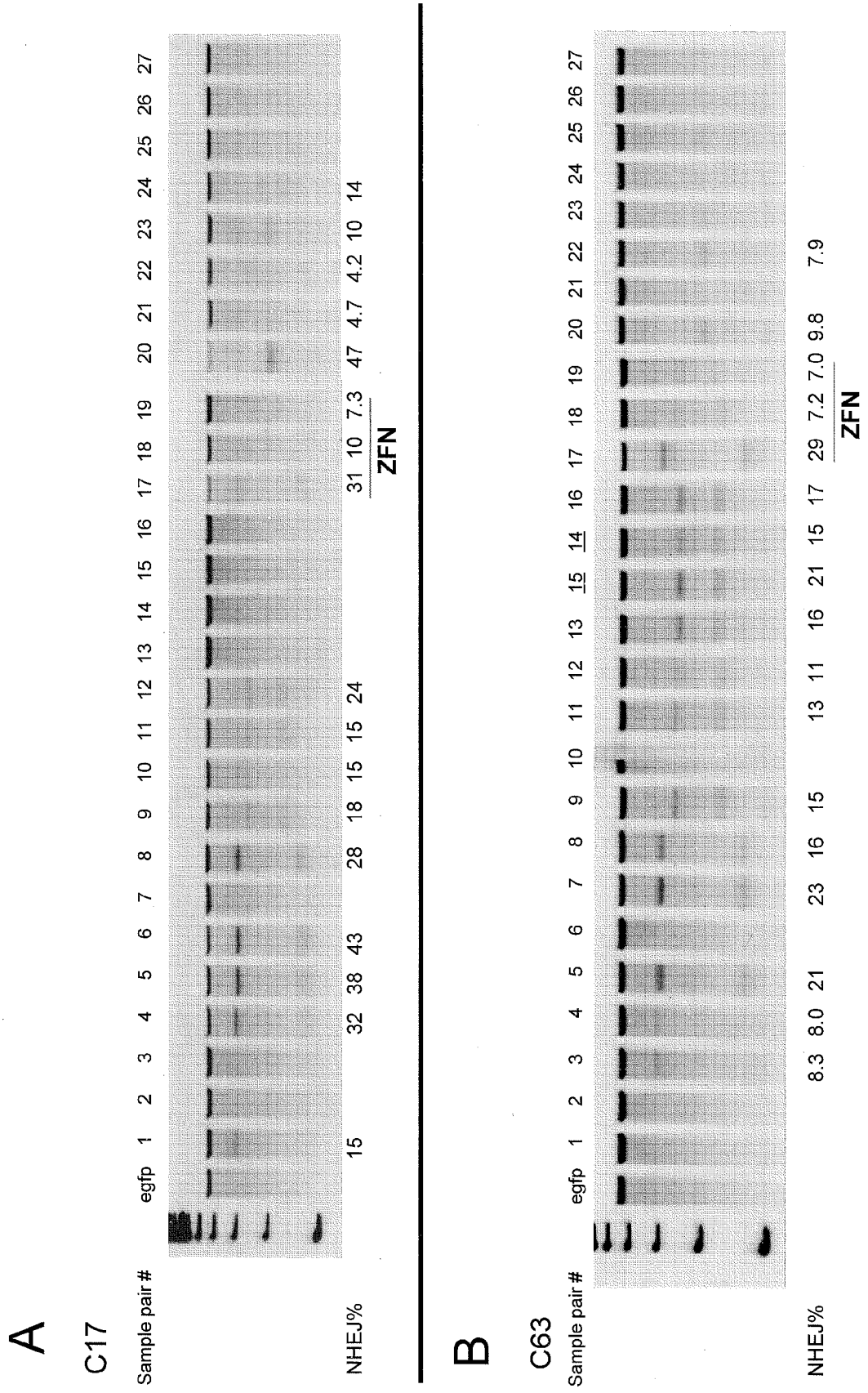

A
C17
Sample pair #
egfp 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
NHEJ%
15 32 38 43 28 18 15 15 24 31 10 7.3 47 4.7 4.2 10 14
ZFN
B
C63
Sample pair #
egfp 1 2 3 4 5 6 7 8 9 10 11 12 13 15 14 16 17 18 19 20 21 22 23 24 25 26 27
NHEJ%
8.3 8.0 21 23 16 15 13 11 16 21 15 17 29 7.2 7.0 9.8 7.9
ZFN

Figure 11

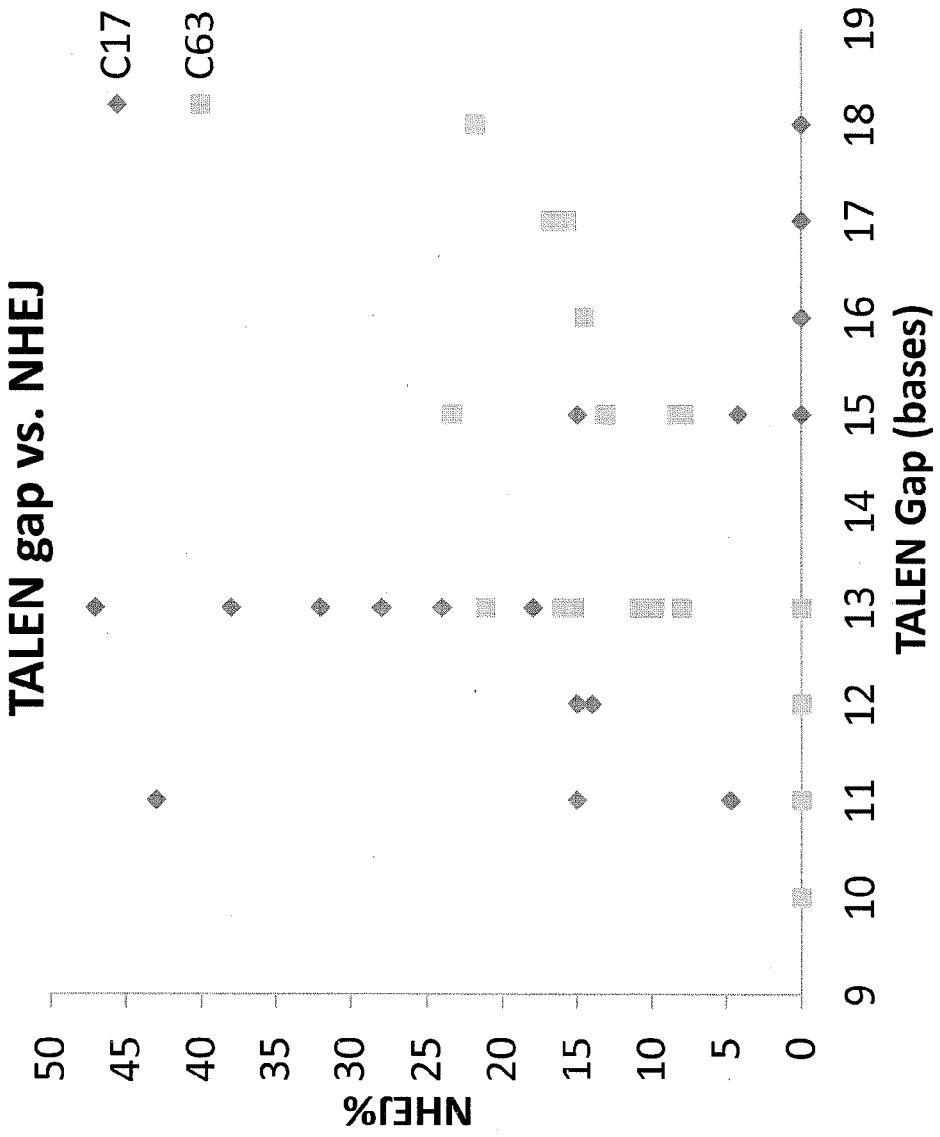

TALEN gap vs. NHEJ
C17
C63
NHEJ%
0
5
10
15
20
25
30
35
40
45
50
9
10
11
12
13
14
15
16
17
18
19
TALEN Gap (bases)

Figure 11C

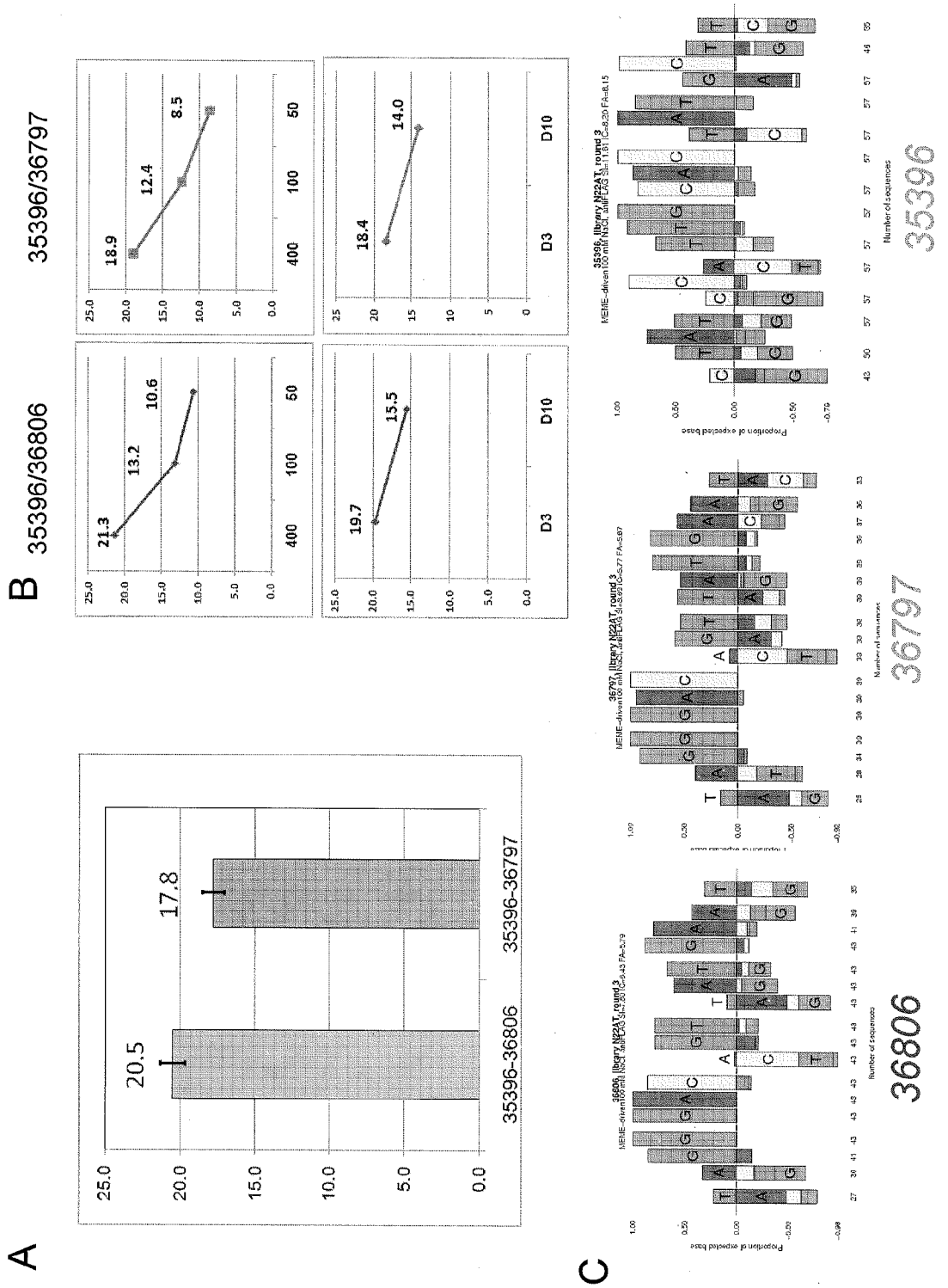
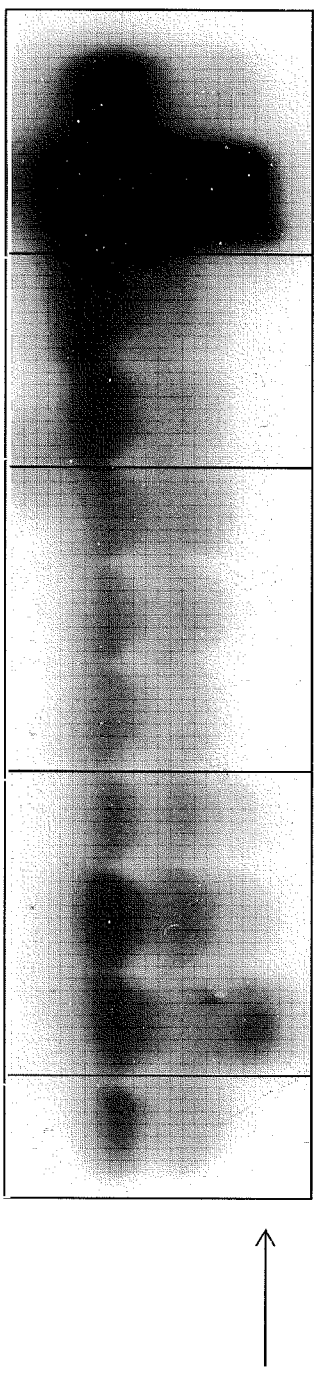
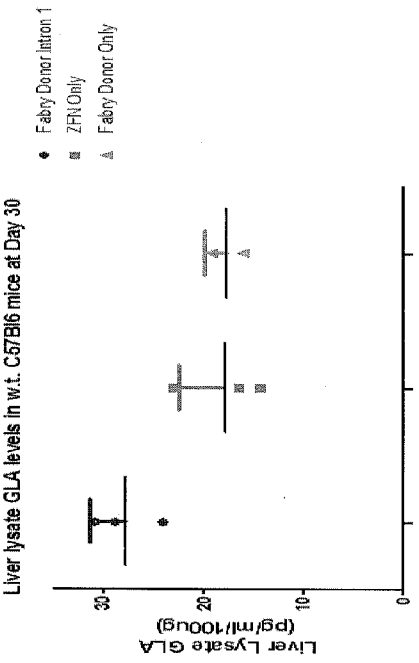

A
20.5
17.8
35396-36806
35396-36797
B
35396/36806
35396/36797
21.3
13.2
10.6
18.9
12.4
8.5
19.7
15.5
18.4
14.0
D3
D10
C
36806
36797
35396

Figure 12


A
Naive
mu alb ZFN + hu Fabry donor
1-1
1-2
1-3
4-1
mu alb ZFN
4-2
4-3
hu Fabry donor
5-1
5-2
human liver lysate


Liver lysate GLA levels in w.t. C57Bl6 mice at Day 30
Fabry Donor Intron 1
ZFN Only
Fabry Donor Only
Liver Lysate GLA
(pg/ml/100ug)
30
20
10
0

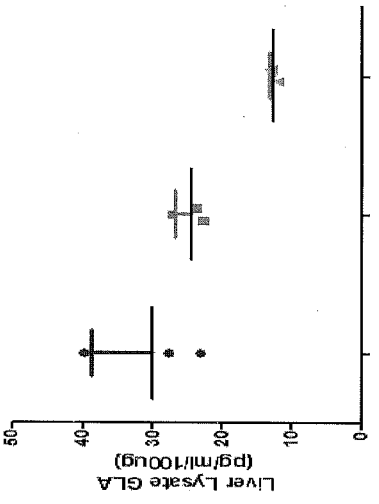

B
Liver lysate GLA levels in w.t. C57Bl6 mice at Day 14
Fabry Donor Intron 1
ZFN Only
Fabry Donor Only
Liver Lysate GLA
(pg/ml/100ug)
50
40
30
20
10
0

Figure 13