GENE THERAPY VECTORS AND CYTOSINE DEAMINASES

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Application Serial No. 61/100,666, filed September 26, 2008, U.S. Provisional Application Serial No. 61/120,618, filed December 8, 2008 and U.S. Provisional Application Serial No. 61/186,823, filed June 13, 2009 the disclosures of which are incorporated herein by reference.

TECHNICAL FIELD

[0002] This disclosure relates to modified cytosine deaminases (CDs). The disclosure further relates to cells expressing such modified CDs and methods of using such modified CDs in the treatment of disease and disorders.

BACKGROUND

[0003] The yeast, or bacterial, cytosine deaminase converts the innocuous antibiotic pro-drug 5-FC into the cytotoxic chemotherapeutic agent 5-fluorouracil (5-FU). Humans (and mammals in general) are not known to posess a naturally occurring gene encoding an enzyme with significant cytosine deaminase activity. Yeast and bacterial cytosine deaminase have gained recognition in the treatment of cancers using gene delivery and viral vectors for the delivery of the enzyme followed by treatment with 5-FC, which is then converted by the enzyme to a cytotoxic drug (Miller *et al.*, Can Res 62:773-780 2002; Kievit *et al.*, Can Res 59:1417-1421 1999).

SUMMARY

[0004] Provided herein are polypeptides that convert 5-FC to 5-FU. Also provided are nucleic acid molecules that encode such polypeptides, cells expressing such polypeptides, vectors containing such polynucleotide and polypeptides and methods of synthesizing 5-FU or derivatives thereof from a suitable using such polypeptides. Accordingly, in various embodiments, isolated or recombinant polypeptides comprising a cytosine deaminase having a sequence as set forth in SEQ ID NO:2 is provided. In other embodiments, the polypeptide comprise SEQ ID NO:2 with a mutation selected from the

group consisting of A23L, V108I, I140L and any combination thereof. In yet another embodiment, the the polypeptide comprises the sequence as set forth in SEQ ID NO:4 and includes up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: (a) 23, 108 and 140. In general, polypeptides provided herein display cytosine deaminase activity useful for converting 5-FC to 5-FU.

[0005] The disclosure also provides polypeptide that comprise a sequence that is at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO:4, wherein the polypeptide has a leucine at position 23, an isoleucine at position 108 and a leucine at position 140, and wherein the polypeptide has cytosine deaminase activity. In yet another embodiment, the polypeptide comprises the sequence as set forth in SEQ ID NO:4.

[0006] The disclosure further provides fusion constructs comprising any one of the foregoing polypeptides operably linked to a uracil phosphoribosyltransferase (UPRT) or an orotate phosphoribosyltransferase (OPRT). In one embodiment, the fusion construct comprises a first polypeptide comprising SEQ ID NO:2 with a mutation selected from the group consisting of A23L, V108I, I140L and any combination thereof linked to a second polypeptide having UPRT or OPRT activity. In yet another embodiment, the fusion construct comprises a first polypeptide having the sequence as set forth in SEQ ID NO:4 and includes up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: 23, 108 and 140 wherein the first polypeptide is linked to a second polypeptide having UPRT or OPRT activity. The disclosure also provides a fusion construct comprising a first polypeptide having cytosine deaminase activity and that comprise a sequence that is at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO:4, wherein the polypeptide has a leucine at position 23, an isoleucine at position 108 and a leucine at position 140, and wherein the first polypeptide is linked to a second polypeptide comprising UPRT or OPRT activity. In one embodiment, the polypeptides having UPRT or OPRT activity comprises a sequence as set forth in SEQ ID NO:8 and 10, respectively, or variants thereof. In yet another embodiment, a first polypeptide comprising cytosine deaminase activity is linked to a polypeptide having UPRT or OPRT activity, wherein the

polypeptides are separated by a peptide linker. In another embodiment, the fusion construct comprises a sequence selected from the group consisting of SEQ ID NO:12, 14, 16 or 18.

[0007] The disclosure further provides polynucleotides encoding any of the foregoing polypeptides. For example, the disclosure provides a polynucleotide encoding a polypeptide comprising SEQ ID NO:2 with a mutation selected from the group consisting of A23L, V108I, I140L and any combination thereof. In yet another embodiment, the polynucleotide encodes a polypeptide comprising a sequence as set forth in SEQ ID NO:4 and includes up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: (a) 23, 108 and 140. In a further embodiment, the disclosure provides a polynucleotide encoding a polypeptide comprising a sequence that is at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO:4, wherein the polypeptide has a leucine at position 23, an isoleucine at position 108 and a leucine at position 140, and wherein the polypeptide has cytosine deaminase activity. In yet another embodiment, the polynucleotide encodes a polypeptide comprising SEQ ID NO: 4. In a further embodiment, the polynucleotide encodes a polypeptide comprising a sequence as set forth in SEQ ID NO:4, 6, 8, 10, 11, 12, or 13.

[0008] The disclosure provides a human codon optimized polynucleotide encoding a cytosine deaminase. In one embodiment, the human codon optimized polynucleotide comprises a sequence as set forth in SEQ ID NO:3. In yet another embodiment, the polynucleotide comprises a sequence as set forth in SEQ ID NO:5, 7, or 9.

[0009] In other embodiments, a cytosine deaminase or fusion construct of the disclosure is delivered using a gene delivery system (GDS). In another aspect, the polynucleotide encoding a cytosine deaminase is delivered with a GDS that is a viral or viral derived vector. The viral vector can be replicating or non-replicating, and can be an adenoviral vector, a measles vector, a herpes vector, a retroviral vector (including a lentiviral vector), a rhabdoviral vector such as a Vesicular Stomatitis viral vector, a reovirus vector, a Seneca Valley Virus vector, a poxvirus vector (including animal pox or vaccinia derived vectors), a parvovirus vector (including an AAV vector), an alphavirus vector or other

viral vector known to one skilled in the art. In one embodiment, the viral vector can be a replication competent retroviral vector capable of infecting replicating mammalian cells. The replication competent retroviral vector can comprise an Orthoretrovirus or more typically a gamma retrovirus vector. In one aspect, a replication competent retroviral vector comprises an internal ribosomal entry site (IRES) 5' to the polynucleotide encoding a cytosine deaminase of the disclosure. In one embodiment, the polynucleotide encoding a cytosine deaminase is 3' to a ENV polynucleotide of a retroviral vector.

[0010] In other embodiments, host cells transfected with a cytosine deaminase or fusion construct of the disclosure, or a vector that includes a polynucleotide or fusion construct of the disclosure, are provided. Host cells include eukaryotic cells such as yeast cells, insect cells, or animal cells. Host cells also include procaryotic cells such as bacterial cells.

[0011] The disclosure also provides methods of treating a cell proliferative disorder including a cancer comprising administering a polynucleotide or polypeptide of the disclosure to a subject and contacting the subject with a cytotoxic drug comprising 5-fluorocytosine (5-FC).

[0012] The disclosure provides a recombinant replication competent retrovirus (RCR) comprising: a retroviral GAG protein; a retroviral POL protein; a retroviral envelope; a retroviral polynucleotide comprising Long-Terminal Repeat (LTR) sequences at the 3' end of the retroviral polynucleotide sequence, a promoter sequence at the 5' end of the retroviral polynucleotide, said promoter being suitable for expression in a mammalian cell, a gag nucleic acid domain, a pol nucleic acid domain and an env nucleic acid domain; a cassette comprising an internal ribosome entry site (IRES) operably linked to a heterologous polynucleotide, wherein the cassette is positioned 5' to the 3' LTR and 3' to the env nucleic acid domain encoding the retroviral envelope; and cis-acting sequences necessary for reverse transcription, packaging and integration in a target cell,wherein the RCR maintains higher replication competency after 6 passages compared to a pACE vector (SEQ ID NO:21). In one embodiment, the the retroviral

polynucleotide sequence is derived from murine leukemia virus (MLV), Moloney murine leukemia virus (MoMLV),Feline leukemia virus or Gibbon ape leukemia virus (GALV). In another embodiment, the MLV is an amphotropic MLV. In yet another embodiment, the retrovirus is an oncoretrovirus. In yet another embodiment, the target cell is a cell having a cell proliferative disorder. The cell proliferative disorder can be selectedfrom the group consisting of, but is not limited to, lung cancer, colon-rectum cancer, breast cancer, prostate cancer, urinary tract cancer, uterine cancer, brain cancer, head and neck cancer, pancreatic cancer, melanoma, stomach cancer and ovarian cancer, rheumatoid arthritis and other auto-immune diseases. In one embodiment, the promoter comprises a CMV promoter having a sequence as set forth in SEQ ID NO:19, 20, or 22 from nucleotide 1 to about nucleotide 582 and may include modification to one or more nucleic acid bases and which is capable of directing and initiating transcription. In yet a further embodiment, the promoter comprises a sequence as set forth in SEQ ID NO:19, 20, or 22 from nucleotide 1 to about nucleotide 582. In a further embodiment, the promoter comprises a CMV-R-U5 domain polynucleotide. In one embodiment, the CMV-R-U5 domain comprise the immediately early promoter from human cytomegalovirus linked to an MLV R-U5 region. In yet another embodiment, the CMV-R-U5 domain polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22 from about nucleotide 1 to about nucleotide 1202 or sequences that are at least 95% identical to a sequence as set forth in SEQ ID NO:19, 20, or 22, wherein the polynucleotide promotes transcription of a nucleic acid molecule operably linked thereto. In another embodiment, the gag and pol of the polynucleotide are derived from an oncoretrovirus. The gag nucleic acid domain can comprise a sequence from about nucleotide number 1203 to about nucleotide 2819 of SEQ ID NO: 19 or a sequence having at least 95%, 98%, 99% or 99.8% identity thereto. The pol domain can comprise a sequence from about nucleotide number 2820 to about nucleotide 6358 of SEQ ID NO:19 or a sequence having at least 95%, 98%, 99% or 99.9% identity thereto. In one embodiment, the env domain encodes an amphoteric env protein. The env domain can comprise a sequence from about nucleotide number 6359 to about nucleotide 8323 of SEQ ID NO:19 or a

sequence having at least 95%, 98%, 99% or 99.8% identity thereto. The IRES domain of the vector can be any IRES, however, in one embodiment the IRES is derived from an encephalomyocarditis virus. In a further embodiment, the IRES comprises a sequence from about nucleotide number 8327 to about nucleotide 8876 of SEQ ID NO:19 or a sequence having at least 95%, 98%, or 99% identity thereto. In yet a further embodiment, the heterologous polynucleotide comprises a polynucleotide having a sequence as set forth in SEQ ID NO:3, 5, 11, 13, 15 or 17. In a further embodiment, the heterologous sequence encodes a polypeptide comprising a sequence as set forth in SEQ ID NO:4. The heterologous nucleic acid is human codon optimized and encodes a polypeptide as set forth in SEQ ID NO:4. In a further embodiment, the heterologous nucleic acid comprises a sequence as set forth in SEQ ID NO: 19 from about nucleotide number 8877 to about 9353. In one embodiment, the 3′ LTR is derived from an oncoretrovirus. In a further embodiment, the 3′ LTR comprises a U3-R-U5 domain. In yet a further embodiment, the 3′ LTR comprises a sequence as set forth in SEQ ID NO:19 from about nucleotide 9405 to about 9998 or a sequence that is at least 95%, 98% or 99.5% identical thereto.

[0013] The disclosure provides a polynucleotide comprising a sequence as set forth in SEQ ID NO:19, 20, or 22.

[0014] The disclosure provides an isolated polynucleotide comprising from 5′ to 3′: a CMV-R-U5 fusion of the immediate early promoter from human cytomegalovirus to an MLV R-U5 region; a PBS, primer binding site for reverse transcriptase; a 5′ splice site; ψ packaging signal; a gag coding sequence for MLV group specific antigen; a pol coding sequence for MLV polymerase polyprotein; a 3′ splice site; a 4070A env coding sequence for envelope protein of MLV strain 4070A; an internal ribosome entry site (IRES) from encephalomyocarditis virus; a modified cytosine deaminase coding sequence; a polypurine tract; and a U3-R-U5 MLV long terminal repeat.

[0015] The disclosure provides a method of treating a subject with a cell proliferative disorder comprising contacting the subject with a polynucleotide a polypeptide of the disclosure having cytosine deaminase activity under conditions such that the

polynucleotide is expressed and contacting the subject with 5-fluorocytosine.

[0016] The disclosure also provides a method of treating a cell proliferative disorder in a subject comprising contacting the subject with a retrovirus of the disclosure, wherein the heterologous nucleic acid sequence encodes a therapeutic protein that inhibits proliferation of a neoplastic cell. In one embodiment, the retrovirus comprises a polynucleotide encoding a polypeptide having a sequence as set forth in SEQ ID NO:4, 12, 14, 16, or 18.

[0017] The details of one or more embodiments of the disclosure are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0018] **Figure 1A-D** shows (a) a schematic of a recombinant retroviral vector of the disclosure; (b) a plasmid map of a polynucleotide of the disclosure; (c and d) a sequence of a polynucleotide of the disclosure (SEQ ID NO:19).

[0019] **Figure 2A-D** shows schemes for the generation of various embodiments of the disclosure comprising polypeptides with CD, OPRT and UPRT activity.

[0020] **Figure 3** shows that higher levels of yCD2 protein are observed compared to wild type yCD protein in infected U-87 cells.

[0021] **Figure 4** shows the stability of a vector comprising a CD polypeptide of the disclosure and the comparison to other vectors of the disclosure.

[0022] **Figure 5A-B** show that a cytosine deaminase activity and vector of the disclosure provide comparable or better expression, and hence killing, of infected rat RG2(5A) or U87 cells (5B) compared to wild type yCD activity (T5.0007),when infected cells are exposed to increasing levels of 5-FC.

[0023] **Figure 6** shows that the specific activity in infected U87 cells of T5.0002 (yCD2) is greater than T5.0001 (partially optimized yCD) which is greater than T5.0007 (wt yCD).

[0024] Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

[0025] As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the agent" includes reference to one or more agents known to those skilled in the art, and so forth.

[0026] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0027] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0028] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice of the disclosed methods and compositions, the exemplary methods, devices and materials are described herein.

[0029] The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior disclosure.

[0030] A cytosine deaminase (EC 3.5.4.1) is an enzyme that catalyzes the chemical reaction

$$\text{cytosine} + H_2O \rightarrow \text{uracil} + NH_3$$

Thus, the two substrates of this enzyme are cytosine and H_2O, whereas its two products are uracil and NH_3. This enzyme belongs to the family of hydrolases, those acting on carbon-nitrogen bonds

other than peptide bonds, specifically in cyclic amidines. The systematic name of this enzyme class is cytosine aminohydrolase. This enzyme is also called isocytosine deaminase. This enzyme participates in pyrimidine metabolism.

[0031] More particularly, cytosine deaminase is an enzyme involved in the metabolic pathway for pyrimidines, through which exogenous cytosine is transformed, via hydrolytic deamination, into uracil. Cytosine deaminase (CDase or CD) activities have been demonstrated in prokaryotes and lower eukaryotes, but they are absent in mammals (Koechlin *et al.*, 1966, Biochem. Pharmacol. 15, 435-446; Polak *et al.*, 1976, Chemotherapy 22, 137-153). The *FCY1* gene of *Saccharomyces cerevisiae* (*S. cerevisiae*) and the *coda* gene of *E. coli*, which encode, respectively, the CDase of these two organisms, are known and their sequences are published (EP 402 108; Erbs *et al.*, 1997, Curr. Genet. 31, 1-6; WO 93/01281). CDase also deaminates a cytosine analogue, 5-fluorocytosine (5-FC) to 5-fluorouracil (5-FU), which is a highly cytotoxic compound, in particular when it is converted to 5-fluoro-UMP (5-FUMP). Cells which lack CDase activity, due either to an inactivating mutation of the gene encoding the enzyme or to their natural deficiency for this enzyme (for example mammalian cells) are resistant to 5-FC (Jund and Lacroute, 1970, J. Bacteriol. 102, 607-615; Kilstrup *et al.*, 1989, J. Bacteriol., 171, 2124-2127). On the other hand, it has been demonstrated that it is possible to transmit 5-FC sensitivity to mammalian cells into which the sequence encoding a CDase activity has been transferred (Huber *et al.*, 1993, Cancer Res. 53, 4619-4626; Mullen *et al.*, 1992, Proc. Natl. Acad. Sci. USA 89, 33-37; WO 93/01281). Accordingly, the use of CD is advantageous in the context of gene therapy, in particular anticancer gene therapy.

[0032] However, 5-FC sensitivity varies a great deal depending on the cell lines. Low sensitivity is observed, for example, in PANC-1 (carcinoma of the pancreas) and SK-BR-3 (breast adenocarcinoma) human tumour lines transduced with a retrovirus expressing the *coda* gene of E. coli (Harris *et al.*, 1994, Gene Therapy 1, 170-175). This phenomenon is explained by the absence or poor endogenous conversion of the 5-FU formed by the enzymatic action of the CDase, to cytotoxic 5-FUMP. This step, which is

normally carried out in mammalian cells by orotate phosphorybosyltransferase (OPRTase), may be absent in certain tumours and thus make gene therapy based on CDase ineffective. In prokaryotes and lower eukaryotes, uracil is transformed into UMP through the action of uracil phosphoribosyltransferase (UPRTase activity). This enzyme also converts 5-FU to 5-FUMP. Importantly, bacterial uracil phosphoribosyltransferase (UPRT) is functionally equivalent to orotate phosphoribosyltransferase (OPRT) or uridine-5'-monophosphate synthase of mammalian cells. These enzymes mediate the conversion of 5-fluorouracil (5-FU) (a fluorinated analog of uracil) to 5-fluorouridine 5' monophosphate (5-FUMP). 5-fluorouridine 5' monophosphate is subsequently converted to 5-FdUDP and FdUTP via the mammalian de novo pyrimidine pathway. Each 5-FdUTP is an irreversible inhibitor of thymidylate synthase (Thy-A) and results in dTTP starvation. It is widely accepted that this conversion is one of the requisite pathways to achieve cytotoxic effects of 5-fluorouracil and that bacterially uracil phosphoribosyltransferase of bacterial origin is able to convert 5-fluorouracil to the same active metabolite as does mammalian orotate phosphoribosyltransferase. In the absence of UPRTase activity or OPRTase activity, the 5-FU, originating from the deamination of the 5-FC by the CDase, is not transformed into cytotoxic 5-FUMP (Jund and Lacroute, 1970, J. Bacteriol. 102, 607-615).

[0033] As described herein, the disclosure provides polynucleotides encoding polypeptides and polypeptides comprising a fusion of a first polypeptide having cytosine deaminase activity and a second polypeptide comprising UPRT or OPRT activity. Such fusion constructs are useful for converting uracil, or a derivative thereof, into a monophosphate analogue, and in particular 5-FU into 5-FUMP.

[0034] As will be described in more detail below, the disclosure is based, at least in part, on the generation and expression of novel polypeptides that catalyze the conversion 5-FC to 5-FU. In one embodiment, novel polypeptides that have been engineered to have improved catalytic properties for the conversion of 5-FC to 5-FU are provided. Such polypeptides include variants that have been altered to include amino acid substitutions at specified residues. While

these variants will be described in more detail below, it is understood that polypeptides of the disclosure may contain one or more modified amino acids. The presence of modified amino acids may be advantageous in, for example, (a) increasing polypeptide *in vivo* half-life, (b) reducing or increasing polypeptide antigenicity, and (c) increasing polypeptide storage stability. Amino acid(s) are modified, for example, co-translationally or post-translationally during recombinant production (*e.g.*, N-linked glycosylation at N--X--S/T motifs during expression in mammalian cells) or modified by synthetic means. Accordingly, A "mutant", "variant" or "modified" protein, enzyme, polynucleotide, gene, or cell, means a protein, enzyme, polynucleotide, gene, or cell, that has been altered or derived, or is in some way different or changed, from a parent protein, enzyme, polynucleotide, gene, or cell. A mutant or modified protein or enzyme is usually, although not necessarily, expressed from a mutant polynucleotide or gene.

[0035] A "mutation" means any process or mechanism resulting in a mutant protein, enzyme, polynucleotide, gene, or cell. This includes any mutation in which a protein, enzyme, polynucleotide, or gene sequence is altered, and any detectable change in a cell arising from such a mutation. Typically, a mutation occurs in a polynucleotide or gene sequence, by point mutations, deletions, or insertions of single or multiple nucleotide residues. A mutation includes polynucleotide alterations arising within a protein-encoding region of a gene as well as alterations in regions outside of a protein-encoding sequence, such as, but not limited to, regulatory or promoter sequences. A mutation in a gene can be "silent", *i.e.*, not reflected in an amino acid alteration upon expression, leading to a "sequence-conservative" variant of the gene. This generally arises when one amino acid corresponds to more than one codon.

[0036] Non-limiting examples of a modified amino acid include a glycosylated amino acid, a sulfated amino acid, a prenlyated (*e.g.*, farnesylated, geranylgeranylated) amino acid, an acetylated amino acid, an acylated amino acid, a pegylated amino acid, a biotinylated amino acid, a carboxylated amino acid, a phosphorylated amino acid, and the like. References adequate to guide one of skill in the

modification of amino acids are replete throughout the literature. Example protocols are found in Walker (1998) Protein Protocols on CD-ROM (Humana Press, Towata, N.J.).

[0037] Recombinant methods for producing, isolating, and using the modified CD polypeptides and polynucleotides of the disclosure are described herein. In addition to recombinant production, the polypeptides may be produced by direct peptide synthesis using solid-phase techniques (*e.g.*, Stewart *et al.* (1969) Solid-Phase Peptide Synthesis (WH Freeman Co, San Francisco); and Merrifield (1963) J. Am. Chem. Soc. 85: 2149-2154). Peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer.

[0038] By way of illustration, the nucleic acid sequences encoding the UPRTases of E. coli (Anderson *et al.*, 1992, Eur. J. Biochem 204, 51-56), of Lactococcus lactis (Martinussen and Hammer, 1994, J. Bacteriol. 176, 6457-6463), of Mycobacterium bovis (Kim *et al.*, 1997, Biochem Mol. Biol. Int 41, 1117-1124) and of Bacillus subtilis (Martinussen *et al.*, 1995, J. Bacteriol. 177, 271-274) can be used in the context of the disclsoure. However, use of a yeast UPRTase, and in particular that encoded by the FUR1 gene of *S. cerevisiae*, the sequence of which disclosed in Kern *et al.* (1990, Gene 88, 149-157) is incorporated herein of reference. By way of indication, the sequences of the genes and those of the corresponding UPRTases can be found in the literature and the specialized data banks (SWISSPROT, EMBL, Genbank Medline, etc.).

[0039] A "protein" or "polypeptide", which terms are used interchangeably herein, comprises one or more chains of chemical building blocks called amino acids that are linked together by chemical bonds called peptide bonds. An "enzyme" means any substance, preferably composed wholly or largely of protein, that catalyzes or promotes, more or less specifically, one or more chemical or biochemical reactions. The term "enzyme" can also refer to a catalytic polynucleotide (*e.g.*, RNA or DNA). A "native" or "wild-type" protein, enzyme, polynucleotide, gene, or cell, means a

protein, enzyme, polynucleotide, gene, or cell that occurs in nature.

[0040] Accordingly, in various embodiments, isolated or recombinant polypeptides comprising SEQ ID NO:2 with a mutation selected from the group consisting of A23L, V108I, I140L and any combination thereof are provided. In yet another embodiment, the the polypeptide comprises the sequence as set forth in SEQ ID NO:4 and includes up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: (a) 23, 108 and 140. In general, polypeptides provided herein display cytosine deaminase activity useful for converting 5-FC to 5-FU.

[0041] The disclosure also provides polypeptide that comprise a sequence that is at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO:4, wherein the polypeptide has a leucine at position 23, an isoleucine at position 108 and a leucine at position 140, and wherein the polypeptide has cytosine deaminase activity. In yet another embodiment, the polypeptide comprises the sequence as set forth in SEQ ID NO:4.

[0042] The disclosure further provides fusion constructs comprising any one of the foregoing polypeptides operably linked to a uracil phosphoribosyltransferase (UPRT) or an oratate phosphoribosyltransferase (OPRT). In one embodiment, the fusion construct comprises a first polypeptide comprising SEQ ID NO:2 with a mutation selected from the group consisting of A23L, V108I, I140L and any combination thereof linked to a second polypeptide having UPRT or OPRT activity. In yet another embodiment, the fusion construct comprises a first polypeptide having the sequence as set forth in SEQ ID NO:4 and includes up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: 23, 108 and 140 wherein the first polypeptide is linked to a second polypeptide having UPRT or OPRT activity. The disclosure also provides a fusion construct comprising a first polypeptide having cytosine deaminase activity and that comprise a sequence that is at least 80%, 90%, 95%, 98% or 99% identical to SEQ ID NO:4, wherein the polypeptide has a leucine at position 23, an isoleucine at position 108 and a leucine at position 140, and wherein the first polypeptide is linked to a second polypeptide comprising UPRT or

OPRT activity. In one embodiment, the polypeptides having UPRT or OPRT activity comprises a sequence as set forth in SEQ ID NO:8 and 10, respectively, or variants thereof. In yet another embodiment, a first polypeptide comprising cytosine deaminase activity is linked to a polypeptide having UPRT or OPRT activity, wherein the polypeptides are separated by a peptide linker. In another embodiment, the fusion construct comprises a sequence selected from the group consisting of SEQ ID NO:11, 12 or 13.

[0043] The term "operably linked" or "operably associated" refers to a functional linkage or association between a regulatory sequence and the polynucleotide regulated by the regulatory sequence or between two distinct polypeptides or polynucleotides encoding such polypeptides.

[0044] A fusion construct comprising a polypeptide having CD activity and a polypeptide comprising UPRT or OPRT activity can be engineered to contain a cleavage site to aid in protein recovery or other linker moiety. Typically a linker will be a peptide linker moiety. The length of the linker moiety is chosen to optimize the biological activity of the CD polypeptide and the UPRT or OPRT polypeptide and can be determined empirically without undue experimentation. The linker moiety should be long enough and flexible enough to allow a substrate to interact with the CD polypeptide and a substrated with the UPRT or OPRT polypeptide. A linker moiety is a peptide between about one and 30 amino acid residues in length, typically between about two and 15 amino acid residues. Examples of linker moieties are --Gly--Gly--, GGGGS (SEQ ID NO:22), $(GGGGS)_N$ (SEQ ID NO:23), $(SGGGG)_N$ (SEQ ID NO: 24), GKSSGSGSESKS (SEQ ID NO:25), GSTSGSGKSSEGKG (SEQ ID NO:26), GSTSGSGKSSEGSGSTKG (SEQ ID NO:27), GSTSGSGKPGSGEGSTKG (SEQ ID NO:28), or EGKSSGSGSESKEF (SEQ ID NO:29). Linking moieties are described, for example, in Huston *et al.*, Proc. Nat'l Acad. Sci 85:5879, 1988; Whitlow *et al.*, Protein Engineering 6:989, 1993; and Newton *et al.*, Biochemistry 35:545, 1996. Other suitable peptide linkers are those described in U.S. Pat. Nos. 4,751,180 and 4,935,233, which are hereby incorporated by reference. A DNA sequence encoding a desired peptide linker can be inserted between, and in the same reading frame as, a polynucleotide encoding a CD

polypeptide followed by a UPRT or OPRT polypeptide, using any suitable conventional technique. For example, a chemically synthesized oligonucleotide encoding the linker can be ligated between two coding polynucleotides. In particular embodiments, a fusion polypeptide comprises from two to four separate domains (*e.g.*, a CD domain and a UPRT or OPRT) are separated by peptide linkers.

[0045] "Conservative amino acid substitution" or, simply, "conservative variations" of a particular sequence refers to the replacement of one amino acid, or series of amino acids, with essentially identical amino acid sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a percentage of amino acids in an encoded sequence result in "conservative variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid.

[0046] Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one conservative substitution group includes Alanine (A), Serine (S), and Threonine (T). Another conservative substitution group includes Aspartic acid (D) and Glutamic acid (E). Another conservative substitution group includes Asparagine (N) and Glutamine (Q). Yet another conservative substitution group includes Arginine (R) and Lysine (K). Another conservative substitution group includes Isoleucine, (I) Leucine (L), Methionine (M), and Valine (V). Another conservative substitution group includes Phenylalanine (F), Tyrosine (Y), and Tryptophan (W).

[0047] Thus, "conservative amino acid substitutions" of a listed polypeptide sequence (*e.g.*, SEQ ID NOs: 2, 4, 6, 8, 10, 11, 12, 13 etc.) include substitutions of a percentage, typically less than 10%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Accordingly, a conservatively substituted variation of a polypeptide of the disclosure can contain 100, 75, 50, 25, or 10 substitutions with a conservatively substituted variation of the same conservative substitution group.

[0048] The "activity" of an enzyme is a measure of its ability to catalyze a reaction, *i.e.*, to "function", and may be expressed as the rate at which the product of the reaction is produced. For example, enzyme activity can be represented as the amount of product produced per unit of time or per unit of enzyme (*e.g.*, concentration or weight), or in terms of affinity or dissociation constants. As used interchangeably herein a "cytosine deaminase activity", "biological activity of cytosine deaminse" or "functional activity of a cytosine deaminse", refers to an activity exerted by a cytosine deaminase protein, or polypeptide of the disclosure, on a cytosine deaminase substrate, as determined *in vivo*, or *in vitro*, according to standard techniques. Assays for measuring cytosine deaminase activity are known in the art. For example, an cytosine deaminase activity can measure by determining the rate of conversion of 5-FC to 5-FU or cytosine to uracil. The detection of 5-FC, 5-FU, cytosine and uracil can be performed by chromatography and other methods known in the art.

[0049] One of skill in the art will appreciate that many conservative variations of the nucleic acid constructs which are disclosed yield a functionally identical construct. For example, as discussed above, owing to the degeneracy of the genetic code, "silent substitutions" (*i.e.*, substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. It will be appreciated by those skilled in the art that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding modified cytosine deaminase polypeptides of the disclosure may be produced, some of which bear substantial identity to the nucleic acid sequences explicitly disclosed herein. For instance, codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the disclosure where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that U in an RNA sequence corresponds to T in a DNA sequence.

[0050] Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the polyeptides provided herein.

[0051] "Conservative variants" are proteins or enzymes in which a given amino acid residue has been changed without altering overall conformation and function of the protein or enzyme including, but not limited to, replacement of an amino acid with one having similar properties, including polar or non-polar character, size, shape and charge. Amino acids other than those indicated as conserved may differ in a protein or enzyme so that the percent protein or amino acid sequence similarity between any two proteins of similar function may vary and can be, for example, at least 30%, at least 50%, at least 70%, at least 80%, or at least 90%, as determined according to an alignment scheme. As referred to herein, "sequence similarity" means the extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. "Sequence identity" herein means the extent to which two nucleotide or amino acid sequences are invariant. "Sequence alignment" means the process of lining up two or more sequences to achieve maximal levels of identity (and, in the case of amino acid sequences, conservation) for the purpose of assessing the degree of similarity. Numerous methods for aligning sequences and assessing similarity/identity are known in the art such as, for example, the Cluster Method, wherein similarity is based on the MEGALIGN algorithm, as well as BLASTN, BLASTP, and FASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). When using all of these programs, the preferred settings are those that results in the highest sequence similarity. For example, the "identity" or "percent identity" with respect to a particular pair of aligned amino acid sequences can refer to the percent amino acid sequence identity that is obtained by ClustalW analysis (version W 1.8 available from European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such

number of identical matches by the greater of (i) the length of the aligned sequences, and (ii) 96, and using the following default ClustalW parameters to achieve slow/accurate pairwise alignments--Gap Open Penalty: 10; Gap Extension Penalty: 0.10; Protein weight matrix: Gonnet series; DNA weight matrix: IUB; Toggle Slow/Fast pairwise alignments=SLOW or FULL Alignment.

[0052] Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (*e.g.*, BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, *e.g.*, in Dayhoff *et al.* (1978) "A model of evolutionary change in proteins" in "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff*et al.* (1992) Proc. Nat'l. Acad. Sci. USA 89: 10915-10919. The BLOSUM62 matrix (FIG. 10) is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, *e.g.*, gapped BLAST 2.0, described in Altschul *et al.* (1997) Nucl. Acids Res. 25: 3389-3402, and made available to the public at the National Center for Biotechnology Information (NCBI) Website (www.ncbi.nlm.nih.gov). Optimal alignments, including multiple alignments, can be prepared using, *e.g.*, PSI-BLAST, available through the NCB1 website and described by Altschul *et al.* (1997) Nucl. Acids Res. 25:3389-3402.

[0053] With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue

"corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence as determined by simply counting from the N-terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[0054] Non-conservative modifications of a particular polypeptide are those which substitute any amino acid not characterized as a conservative substitution. For example, any substitution which crosses the bounds of the six groups set forth above. These include substitutions of basic or acidic amino acids for neutral amino acids (*e.g.*, Asp, Glu, Asn, or Gln for Val, Ile, Leu or Met), aromatic amino acid for basic or acidic amino acids (*e.g.*, Phe, Tyr or Trp for Asp, Asn, Glu or Gln) or any other substitution not replacing an amino acid with a like amino acid. Basic side chains include lysine (K), arginine (R), histidine (H); acidic side chains include aspartic acid (D), glutamic acid (E); uncharged polar side chains include glycine (G), asparagine(N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), cysteine (C); nonpolar side chains include alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), phenylalanine (F), methionine (M), tryptophan (W); beta-branched side chains include threonine (T), valine (V), isoleucine (I); aromatic side chains include tyrosine (Y), phenylalanine (F), tryptophan (W), histidine (H).

[0055] Accordingly, some amino acid residues at specific positions in a polypeptide are "excluded" from conservative amino acid substitutions. For example, the disclosure provides a polypeptide comprising a sequence as set forth in SEQ ID NO:4. In some embodiments, the polypeptide may be altered with conservative amino acid substitutions as described above, however, certain residues are desirably left unsubstituted such as residues at positions 23, 108, and 140.

[0056] A "parent" protein, enzyme, polynucleotide, gene, or cell, is any protein, enzyme, polynucleotide, gene, or cell, from which any other protein, enzyme, polynucleotide, gene, or cell, is derived or made, using any methods, tools or techniques, and whether or not the parent is itself native or mutant. A parent polynucleotide or gene encodes for a parent protein or enzyme.

[0057] A polynucleotide, polypeptide, or other component is "isolated" when it is partially or completely separated from components with which it is normally associated (other proteins, nucleic acids, cells, synthetic reagents, etc.). A nucleic acid or polypeptide is "recombinant" when it is artificial or engineered, or derived from an artificial or engineered protein or nucleic acid. For example, a polynucleotide that is inserted into a vector or any other heterologous location, *e.g.*, in a genome of a recombinant organism, such that it is not associated with nucleotide sequences that normally flank the polynucleotide as it is found in nature is a recombinant polynucleotide. A protein expressed *in vitro* or *in vivo* from a recombinant polynucleotide is an example of a recombinant polypeptide. Likewise, a polynucleotide sequence that does not appear in nature, for example a variant of a naturally occurring gene, is recombinant.

[0058] In other embodiments, isolated polynucleotides are provided that encode a cytosine deaminase polypeptide or fusion construct of the disclosure. In one aspect, the disclosure provides an isolated or recombinant polynucleotides referred to herein as "CD optimized polynucleotides", "CD polynucleotides" or "CD-fusion polynucleotides." The terms "polynucleotide," "nucleotide sequence," and "nucleic acid molecule" are used to refer to a polymer of nucleotides (A, C, T, U, G, etc. or naturally occurring or

artificial nucleotide analogues), *e.g.*, DNA or RNA, or a representation thereof, *e.g.*, a character string, etc., depending on the relevant context. A given polynucleotide or complementary polynucleotide can be determined from any specified nucleotide sequence.

[0059] One embodiment of the disclosure pertains to isolated polynucleotides that encode a cytosine deamine or a mutant cytosine deaminase polypeptide or biologically active portions thereof. As used herein, the term "nucleic acid molecule" or "polynucleotide" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded.

[0060] In one embodiment, a CD optimized polynucleotide or CD polynucleotide comprises recombinant, engineered or isolated forms of naturally occurring nucleic acids isolated from an organism, *e.g.*, a bacterial or yeast strain. Exemplary CD polynucleotides include those that encode the wild-type polypeptides set forth in SEQ ID NO: 2. In another embodiment of the disclosure, CD polynucleotides are produced by diversifying, *e.g.*, recombining and/or mutating one or more naturally occurring, isolated, or recombinant CD polynucleotides. As described in more detail elsewhere herein, it is possible to generate diversified CD polynucleotides encoding CD polypeptides with superior functional attributes, *e.g.*, increased catalytic function, increased stability, or higher expression level, than a CD polynucleotide used as a substrate or parent in the diversification process. Exemplary polynucleotides include SEQ ID NO:3 and 5 and those that encode the CD variant polypeptides of the disclosure.

[0061] The polynucleotides of the disclosure have a variety of uses in, for example, recombinant production (*i.e.*, expression) of the CD polypeptides of the disclosure and as substrates for further diversity generation, *e.g.*, recombination reactions or mutation reactions to produce new and/or improved CD homologues, and the like.

[0062] It is important to note that certain specific, substantial and credible utilities of CD polynucleotides do not

require that the polynucleotide encode a polypeptide with substantial CD activity or even variant CD activity. For example, CD polynucleotides that do not encode active enzymes can be valuable sources of parental polynucleotides for use in diversification procedures to arrive at CD polynucleotide variants, or non-CD polynucleotides, with desirable functional properties (*e.g.*, high k_{cat} or k_{cat}/K_m, low K_m, high stability towards heat or other environmental factors, high transcription or translation rates, resistance to proteolytic cleavage, etc.).

[0063] CD polynucleotides, including nucleotide sequences that encode CD polypeptides and variants thereof, fragments of CD polypeptides, related fusion proteins, or functional equivalents thereof, are used in recombinant DNA molecules that direct the expression of the CD polypeptides in appropriate host cells, such as bacterial cells. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can also be used to clone and express the CD polynucleotides. The term "host cell", as used herein, includes any cell type which is susceptible to transformation with a nucleic acid construct. The term "transformation" means the introduction of a foreign (*i.e.*, extrinsic or extracellular) gene, DNA or RNA sequence to a host cell, so that the host cell will express the introduced gene or sequence to produce a desired substance, typically a protein or enzyme coded by the introduced gene or sequence. The introduced gene or sequence may include regulatory or control sequences, such as start, stop, promoter, signal, secretion, or other sequences used by the genetic machinery of the cell. A host cell that receives and expresses introduced DNA or RNA has been "transformed" and is a "transformant" or a "clone." The DNA or RNA introduced to a host cell can come from any source, including cells of the same genus or species as the host cell, or cells of a different genus or species.

[0064] As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms preferentially use a subset of these codons. The codons that are utilized most often in a

species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons (see, *e.g.*, Zhang *et al.* (1991) Gene 105:61-72). Codons can be substituted to reflect the preferred codon usage of the host, a process sometimes called "codon optimization" or "controlling for species codon bias."

[0065] Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (see also, Murray *et al.* (1989) Nucl. Acids Res. 17:477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *S. cerevisiae* and mammals are UAA and UGA, respectively. The preferred stop codon for monocotyledonous plants is UGA, whereas insects and *E. coli* prefer to use UAA as the stop codon (Dalphin *et al.* (1996) Nucl. Acids Res. 24: 216-218).

[0066] Codon usage bias refers to differences among organisms in the frequency of occurrence of codons in protein-coding DNA sequences (genes). A codon is a series of three nucleotides (triplets) that encodes a specific amino acid residue in a polypeptide chain. Because there are four nucleotides in DNA, adenine (A), guanine (G), cytosine (C) and thymine (T), there are 64 possible triplets encoding 20 amino acids, and three translation termination (nonsense) codons. Because of this redundancy, all but two amino acids are encoded by more than one triplet. Different organisms often show particular preferences for one of the several codons that encode the same amino acid. How these preferences arise is a much debated area of molecular evolution.

[0067] It is generally acknowledged that codon preferences reflect a balance between mutational biases and natural selection for translational optimization. Optimal codons in fast-growing microorganisms, like *Escherichia coli* or *Saccharomyces cerevisiae* (baker's yeast), reflect the composition of their respective genomic tRNA pool. It is thought that optimal codons help to achieve faster translation rates and high accuracy. As a result of these factors, translational selection is expected to be stronger in highly

expressed genes, as is indeed the case for the above-mentioned organisms. In other organisms that do not show high growing rates or that present small genomes, codon usage optimization is normally absent, and codon preferences are determined by the characteristic mutational biases seen in that particular genome. Examples of this are *Homo sapiens* (human) and *Helicobacter pylori*. Organisms that show an intermediate level of codon usage optimization include *Drosophila melanogaster* (fruit fly), *Caenorhabditis elegans* (nematode worm) or *Arabidopsis thaliana* (wall cress).

[0068] The term "codon optimized sequences" generally refers to nucleotide sequences that have been optimized for a particular host species by replacing any codons having a usage frequency of less than about 20%. Nucleotide sequences that have been optimized for expression in a given host species by elimination of spurious polyadenylation sequences, elimination of exon/intron splicing signals, elimination of transposon-like repeats and/or optimization of GC content in addition to codon optimization are referred to herein as an "expression enhanced sequences."

[0069] **Table 1:** The human codon usage and codon preference. For each codon, the table displays the frequency of usage of each codon (per thousand) in human coding regions (first column) and the relative frequency of each codon among synonymous codons (second column).

The Human Codon Usage Table

Gly	GGG	17.08	0.23	Arg	AGG	12.09	0.22	Trp	TGG	14.74	1.00	Arg	CGG	10.40	0.19
Gly	GGA	19.31	0.26	Arg	AGA	11.73	0.21	End	TGA	2.64	0.61	Arg	CGA	5.63	0.10
Gly	GGT	13.66	0.18	Ser	AGT	10.18	0.14	Cys	TGT	9.99	0.42	Arg	CGT	5.16	0.09
Gly	GGC	24.94	0.33	Ser	AGC	18.54	0.25	Cys	TGC	13.86	0.58	Arg	CGC	10.82	0.19
Glu	GAG	38.82	0.59	Lys	AAG	33.79	0.60	End	TAG	0.73	0.17	Gln	CAG	32.95	0.73
Glu	GAA	27.51	0.41	Lys	AAA	22.32	0.40	End	TAA	0.95	0.22	Gln	CAA	11.94	0.27
Asp	GAT	21.45	0.44	Asn	AAT	16.43	0.44	Tyr	TAT	11.80	0.42	His	CAT	9.56	0.41
Asp	GAC	27.06	0.56	Asn	AAC	21.30	0.56	Tyr	TAC	16.48	0.58	His	CAC	14.00	0.59
Val	GTG	28.60	0.48	Met	ATG	21.86	1.00	Leu	TTG	11.43	0.12	Leu	CTG	39.93	0.43
Val	GTA	6.09	0.10	Ile	ATA	6.05	0.14	Leu	TTA	5.55	0.06	Leu	CTA	6.42	0.07

Val	GTT	10.30	0.17	Ile	ATT	15.03	0.35	Phe	TTT	15.36	0.43	Leu	CTT	11.24	0.12
Val	GTC	15.01	0.25	Ile	ATC	22.47	0.52	Phe	TTC	20.72	0.57	Leu	CTC	19.14	0.20
Ala	GCG	7.27	0.10	Thr	ACG	6.80	0.12	Ser	TCG	4.38	0.06	Pro	CCG	7.02	0.11
Ala	GCA	15.50	0.22	Thr	ACA	15.04	0.27	Ser	TCA	10.96	0.15	Pro	CCA	17.11	0.27
Ala	GCT	20.23	0.28	Thr	ACT	13.24	0.23	Ser	TCT	13.51	0.18	Pro	CCT	18.03	0.29
Ala	GCC	28.43	0.40	Thr	ACC	21.52	0.38	Ser	TCC	17.37	0.23	Pro	CCC	20.51	0.33

[0070] Accordingly, in some embodiments of the disclosure a polynucleotide comprises a molecule codon optimized for translation in a human cell. For example, SEQ ID NO:3 and 5 compise sequences that have been optimized for producing cytosine deaminase in a human host cell.

[0071] The disclosure thus provides a human codon optimized polynucleotide encoding a polypeptide having cytosine deaminase activity. A human codon optimized polynucleotide of the disclosure (*e.g.*, SEQ ID NO:5) may further include additional mutations resulting in conservative amino acid substitution and or improved activity or stability of the encoded polypeptide. For example, mutations at positions 23, 108 and 140 of the polypeptide comprising SEQ ID NO:6 provide increased thermostability. Accordingly, the disclosure provides a human codon optimized polynucleotide encoding a polypeptide having cytosine deaminase activity and increased thermostabiliyt. In one embodiment, the human codon optimized polynucleotide comprises a sequence encoding a polypeptide of SEQ ID NO:4, wherein the codons encoding the amino acids of the polypeptide have been optimized for expression in a human cell. In another embodiment, a polynucleotide of the disclosure comprises SEQ ID NO:3 or 5.

[0072] In one embodiment, the disclosure provides a polynucleotide comprising a cytosine deaminase polynucleotide or a codon optimized polynucleotide or mutant linked to a hterologous polynucleotide encoding a polypeptide having UPRT or OPRT activity.

[0073] A polynucleotide of the disclosure comprising, *e.g.*, a sequence encoding the polypeptide of SEQ ID NO:4, 8, 10, 12, 14, 16, or 18; or having the nucleotide sequence of set forth in any of SEQ ID NOs:3, 5, 7, 9, 11, 13, 15, or 17, or a portion thereof, can be

isolated and generated using standard molecular biology techniques and the sequence information provided herein.

[0074] A polynucleotide of the disclosure can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer. In some embodiments, an isolated nucleic acid molecule of the disclosure comprises a nucleic acid molecule which is a complement of a nucleotide sequence encoding a polypeptide set forth in any of SEQ NOs:2, 4, 6, 8, 10, 12, 14, 16, or 18, or having the nucleotide sequence of set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, or 17.

[0075] In another embodiment, an isolated polynucleotide of the disclosure comprises a nucleic acid molecule that hybridizes under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence encoding a polypeptide set forth in any of SEQ NOs:4, 12, 14, 16 or 18, or consisting of the nucleotide sequence set forth in any of SEQ ID NOs:3, 5, 11, 13, 15, or 17. In other embodiments, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 nucleotides in length. Nucleic acid molecules are "hybridizable" to each other when at least one strand of one polynucleotide can anneal to another polynucleotide under defined stringency conditions. Stringency of hybridization is determined, *e.g.*, by (a) the temperature at which hybridization and/or washing is performed, and (b) the ionic strength and polarity (*e.g.*, formamide) of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two polynucleotides contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stringency (such as, for example, in an aqueous solution of 0.5 X SSC at 65°C) requires that the sequences exhibit some high degree of complementarity over their entire sequence. Conditions of intermediate stringency (such as, for

example, an aqueous solution of 2 X SSC at 65°C) and low stringency (such as, for example, an aqueous solution of 2 X SSC at 55°C), require correspondingly less overall complementarity between the hybridizing sequences (1 X SSC is 0.15 M NaCl, 0.015 M Na citrate). Nucleic acid molecules that hybridize include those which anneal under suitable stringency conditions and which encode polypeptides or enzymes having the same function, such as the ability to catalyze the conversion of 5-Fluorocytosine (5-FC) to 5-flurouracil (5-FU), of the disclosure. Further, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. In some cases, an isolated nucleic acid molecule of the disclosure hybridizes under stringent conditions to a nucleic acid sequence encoding a polypeptide set forth in any of SEQ ID NOs: 4, 12, 14, 16, or 18, or having the nucleotide sequence set forth in any of SEQ ID NOs: 3, 5, 11, 13, 15, or 17. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

[0076] The skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences of any polynucleotides encoding a polypeptide set forth in any of SEQ ID NOs: 2, 4, 6, 8, 10, 12, 14, 16, or 18, or having the nucleotide sequence set forth in any of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, or 17, thereby leading to changes in the amino acid sequence of the encoded proteins. In some cases the alteration will lead to altered function of the polyepeptide. In other cases the change will not alter the functional ability of the encoded polypeptide. In general, substitutions that do not alter the function of a polyeptide include nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the parent sequence

without altering the biological activity of the resulting polypeptide, *e.g.*, catalyzing the conversion of 5-FC to 5-FU.

[0077] Also contemplated are those situations where it is desirable to alter the activity of a parent polypeptide such that the polypeptide has new or increased activity on a particular substrate or increased stability or reduced degradation. It is understood that these amino acid substitutions will generally not constitute "conservative" substitutions. Instead, these substitutions constitute non-conservative substitutions introduced in to a sequence in order to obtain a new or improved activity. For example, SEQ ID NO:1 provides the parent nucleic acid sequence for *S. cervisae* cytosine deaminase (SEQ ID NO:2 provides the corresponding polypeptide). SEQ ID NO:3 provides the nucleic acid sequence of a mutant sequence that includes amino acid substitutions that impart increased stability to the polypeptide. Accordingly, the nucleic acid molecule encoding the amino acid sequence of SEQ ID NO:2 provides a "parent" nucleic acid molecule from which mutations can be made to obtain a nucleic acid molecule that encodes a modified polypeptide that includes amino acid substitutions.

[0078] It is also understood that a modified polypeptide can constitute a "parent" polypeptide from which additional substitutions can be made. Accordingly, a parent polypeptide, and a nucleic acid molecule that encodes a parent polypeptide, includes modified polypeptides and not just "wild-type" sequences. For example, the polynucleotide of SEQ ID NO:5 is a modified polynucleotide with respect to SEQ ID NO:1 (*i.e.*, the "parent" polynucleotide). Similarly, the polynucleotide of SEQ ID NO:3 is a modified polynucleotide with respect to SEQ ID NO:5. Accordingly, SEQ ID NO:5 is the parent sequence of SEQ ID NO:3.

[0079] Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling *et al.* (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale *et al.* (1996) "Oligonucleotide-directed random mutagenesis using the phosphorothioate method" Methods Mol. Biol. 57:369-374; Smith (1985) "*In vitro* mutagenesis" Ann. Rev. Genet. 19:423-462; Botstein & Shortle (1985) "Strategies and applications of *in vitro* mutagenesis" Science 229:1193-1201; Carter (1986) "Site-directed mutagenesis"

Biochem. J. 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Proc. Natl. Acad. Sci. USA 82:488-492; Kunkel *et al.* (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" Methods in Enzymol. 154, 367-382; and Bass *et al.* (1988) "Mutant Trp repressors with new DNA-binding specificities" Science 242:240-245); oligonucleotide-directed mutagenesis (Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" Nucleic Acids Res. 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" Methods in Enzymol. 100:468-500; and Zoller & Smith (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" Methods in Enzymol. 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor *et al.* (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" Nucl. Acids Res. 13: 8749-8764; Taylor *et al.* (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" Nucl. Acids Res. 13: 8765-8787; Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers *et al.* (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" Nucl. Acids Res. 16:791-802; and Sayers *et al.* (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucl. Acids Res. 16: 803-814); mutagenesis using gapped duplex DNA (Kramer *et al.* (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed

construction of mutations via gapped duplex DNA" 154:350-367; Kramer *et al.* (1988) "Improved enzymatic *in vitro* reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; and Fritz *et al.* (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*" Nucl. Acids Res. 16: 6987-6999).

[0080] Additional suitable methods include point mismatch repair (Kramer *et al.* (1984) "Point Mismatch Repair" Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter *et al.* (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh & Henikoff (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), restriction-selection and restriction-purification (Wells *et al.* (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar *et al.* (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14: 6361-6372; Wells *et al.* (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; and Grundstrom *et al.* (1985) "Oligonucleotide-directed mutagenesis by microscale `shot-gun` gene synthesis" Nucl. Acids Res. 13: 3305-3316); double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455; and "Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[0081] Additional details regarding various diversity generating methods can be found in the following U.S. patents, PCT publications, and EPO publications: U.S. Pat. No. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for *In vitro* Recombination;" U.S. Pat. No. 5,811,238 to Stemmer *et al.* (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer *et al.* (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, *et al.* (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 to Minshull, *et al.* (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Crameri, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO 97/20078 by Stemmer and Crameri "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen *et al.* "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen *et al.* "Antigen Library Immunization;" WO 99/41369 by Punnonen *et al.* "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen *et al.* "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Crameri, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer *et al.*, "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt *et al.*, "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre *et al.* "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/13487 by Stemmer *et al.*, "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection;" WO 00/00632, "Methods for Generating Highly Diverse Libraries;" WO 00/09679, "Methods for Obtaining *in vitro* Recombined Polynucleotide Sequence Banks and Resulting Sequences;" WO 98/42832 by Arnold *et al.*, "Recombination of

Polynucleotide Sequences Using Random or Defined Primers;" WO 99/29902 by Arnold *et al.*, "Method for Creating Polynucleotide and Polypeptide Sequences;" WO 98/41653 by Vind, "An *in vitro* Method for Construction of a DNA Library;" WO 98/41622 by Borchert *et al.*, "Method for Constructing a Library Using DNA Shuffling;" WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination;" WO 00/18906 by Patten *et al.*, "Shuffling of Codon-Altered Genes;" WO 00/04190 by del Cardayre *et al.* "Evolution of Whole Cells and Organisms by Recursive Recombination;" WO 00/42561 by Crameri *et al.*, "Oligonucleotide Mediated Nucleic Acid Recombination;" WO 00/42559 by Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations;" WO 00/42560 by Selifonov *et al.*, "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics;" WO 01/23401 by Welch *et al.*, "Use of Codon-Varied Oligonucleotide Synthesis for Synthetic Shuffling;" and WO 01/64864 "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation" by Affholter.

[0082] Also provided are recombinant constructs comprising one or more of the nucleic acid sequences as broadly described above. The constructs comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), a viral vector or the like, into which a polynucleotide of the disclosure has been inserted, in a forward or reverse orientation. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. In one embodiment, the viral vector is a retroviral vector.

[0083] The terms "vector", "vector construct" and "expression vector" mean the vehicle by which a DNA or RNA sequence (*e.g.* a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (*e.g.* transcription and translation) of the introduced sequence. Vectors typically comprise the DNA of a transmissible agent, into which foreign DNA encoding a protein is inserted by restriction enzyme technology. A common type of vector is a "plasmid", which generally is a self-contained molecule of double-stranded DNA that can readily accept additional (foreign) DNA

and which can readily introduced into a suitable host cell. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. Non-limiting examples include pKK plasmids (Clonetech), pUC plasmids, pET plasmids (Novagen, Inc., Madison, Wis.), pRSET or pREP plasmids (Invitrogen, San Diego, Calif.), or pMAL plasmids (New England Biolabs, Beverly, Mass.), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, *e.g.*, antibiotic resistance, and one or more expression cassettes.

[0084] The terms "express" and "expression" mean allowing or causing the information in a gene or DNA sequence to become manifest, for example producing a protein by activating the cellular functions involved in transcription and translation of a corresponding gene or DNA sequence. A DNA sequence is expressed in or by a cell to form an "expression product" such as a protein. The expression product itself, *e.g.* the resulting protein, may also be said to be "expressed" by the cell. A polynucleotide or polypeptide is expressed recombinantly, for example, when it is expressed or produced in a foreign host cell under the control of a foreign or native promoter, or in a native host cell under the control of a foreign promoter.

[0085] Polynucleotides provided herein can be incorporated into any one of a variety of expression vectors suitable for expressing a polypeptide. Suitable vectors include chromosomal, nonchromosomal and synthetic DNA sequences, *e.g.*, derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated viruses, retroviruses and many others. Any vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used. These nucleic acid vectors can be delivered using a non-viral delivery system such as described in P.Midoux et al. Brit J Pharm

157: 166-178 2009 or Kamimura & Liu J AAPS 10:589-595 (2008) or a viral system as described below.

[0086] In another aspect, the polynucleotide encoding a cytosine deaminase or mutant thereof is delivered with a gene delivery system that is a viral or viral derived vector. The viral vector can be replicating or non-replicating, and can be an adenoviral vector, a measles vector, a herpes vector, a retroviral vector (including a lentiviral vector), a rhabdoviral vector such as a Vesicular Stomatitis viral vector, a reovirus vector, a Seneca Valley Virus vector, a poxvirus vector (including animal pox or vaccinia derived vectors), a parvovirus vector (including an AAV vector), an alphavirus vector or other viral vector known to one skilled in the art (see also, *e.g.*, *Concepts in Genetic Medicine*, ed. Boro Dropulic and Barrie Carter, Wiley, 2008, Hoboken, NJ.; *The Development of Human Gene Therapy*, ed. Theodore Friedmann, Cold Springs Harbor Laboratory Press, Cold springs Harbor, New York, 1999; *Gene and Cell Therapy*, ed. Nancy Smyth Templeton, Marcel Dekker Inc., New York, New York, 2000 and *Gene Therapy: Therapeutic Mechanism and Strategies*, ed. Nancy Smyth Templetone and Danilo D Lasic, Marcel Dekker, Inc., New York, New York, 2000; the disclosures of which are incorporated herein by reference).

[0087] In one embodiment, the viral vector can be a replication competent retroviral vector capable of infecting only replicating mammalian cells. Retroviruses have been classified in various ways but the nomenclature has been standardized in the last decade (see ICTVdB - The Universal Virus Database, v 4 on the World Wide Web (www) at ncbi.nlm.nih.gov/ICTVdb/ICTVdB/ and the text book "Retroviruses" Eds Coffin, Hughs and Varmus, Cold Spring Harbor Press 1997; the disclosure of which are incorporated herein by reference). The replication competent retroviral vector can comprise an Orthoretrovirus or more typically a gamma retrovirus vector. In one aspect, a replication competent retroviral vector comprises an internal ribosomal entry site (IRES) 5' to the polynucleotide encoding a cytosine deaminase. In one embodiment, the polynucleotide encoding a cytosine deaminase is 3' to an ENV polynucleotide of a retroviral vector.

[0088] Accordingly, in other embodiments, vectors that include a polynucleotide of the disclosure are provided. In other embodiments, host cells transfected with a nucleic acid molecule of the disclosure, or a vector that includes a polynucleotide of the disclosure, are provided. Host cells include eucaryotic cells such as yeast cells, insect cells, or animal cells. Host cells also include prokaryotic cells such as bacterial cells.

[0089] As previously discussed, general texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other relevant topics, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology Volume 152, (Academic Press, Inc., San Diego, Calif.) ("Berger"); Sambrook *et al.*, Molecular Cloning--A Laboratory Manual, 2d ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel"). Examples of protocols sufficient to direct persons of skill through *in vitro* amplification methods, including the polymerase chain reaction (PCR), the ligase chain reaction (LCR), Qβ-replicase amplification and other RNA polymerase mediated techniques (*e.g.*, NASBA), *e.g.*, for the production of the homologous nucleic acids of the disclosure are found in Berger, Sambrook, and Ausubel, as well as in Mullis *et al.* (1987) U.S. Pat. No. 4,683,202; Innis *et al.*, eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press Inc. San Diego, Calif.) ("Innis"); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3: 81-94; Kwoh *et al.* (1989) Proc. Natl. Acad. Sci. USA 86: 1173; Guatelli *et al.* (1990) Proc. Nat'l. Acad. Sci. USA 87: 1874; Lomell *et al.* (1989) J. Clin. Chem 35: 1826; Landegren *et al.* (1988) Science 241: 1077-1080; Van Brunt (1990) Biotechnology 8: 291-294; Wu and Wallace (1989) Gene 4:560; Barringer *et al.* (1990) Gene 89:117; and Sooknanan and Malek (1995) Biotechnology 13: 563-564. Improved methods for cloning *in vitro* amplified nucleic acids are described in Wallace *et al.*, U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng *et al.* (1994) Nature 369: 684-

685 and the references cited therein, in which PCR amplicons of up to 40 kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, *e.g.*, Ausubel, Sambrook and Berger, all supra.

[0090] Also provided are engineered host cells that are transduced (transformed or transfected) with a vector provided herein (*e.g.*, a cloning vector or an expression vector), as well as the production of polypeptides of the disclosure by recombinant techniques. The vector may be, for example, a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying a coding polynucleotide. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, *e.g.*, Sambrook, Ausubel and Berger, as well as *e.g.*, Freshney (1994) Culture of Animal Cells: A Manual of Basic Technique, 3rd ed. (Wiley-Liss, New York) and the references cited therein.

[0091] Vectors can be employed to transform an appropriate host to permit the host to express a protein or polypeptide. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *B. subtilis*, Streptomyces, and *Salmonella typhimurium*; fungal cells, such as *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Neurospora crassa*; insect cells such as Drosophila and *Spodoptera frugiperda*; mammalian cells such as CHO, COS, BHK, HEK 293 or Bowes melanoma; or plant cells or explants, etc.

[0092] In bacterial systems, a number of expression vectors may be selected depending upon the use intended cytosine deaminase polypeptide. For example, when large quantities of CD polypeptide or fragments thereof are needed for commercial production or for induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified can be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in

which the CD polypeptide coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) J. Biol. Chem. 264: 5503-5509); pET vectors (Novagen, Madison Wis.); and the like.

[0093] Similarly, in the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used for production of the CD polypeptides of the disclosure. For reviews, see Ausubel (*supra*) and Grant *et al.* (1987) Methods in Enzymology 153:516-544.

[0094] The disclosure also provides replication competent retroviral vectors having increased stability relative to prior retroviral vectors. Such increased stability during infection and replication is important for the treatment of cell proliferative disorders. The combination of transduction efficiency, transgene stability and target selectivity is provided by the replication competent retrovirus. The compositions and methods provide insert stability and maintains transcription activity of the transgene and the translational viability of the encoded polypeptide.

[0095] Retroviruses can be transmitted horizontally and vertically. Efficient infectious transmission of retroviruses requires the expression on the target cell of receptors which specifically recognize the viral envelope proteins, although viruses may use receptor-independent, nonspecific routes of entry at low efficiency. In addition, the target cell type must be able to support all stages of the replication cycle after virus has bound and penetrated. Vertical transmission occurs when the viral genome becomes integrated in the germ line of the host. The provirus will then be passed from generation to generation as though it were a cellular gene. Hence endogenous proviruses become established which frequently lie latent, but which can become activated when the host is exposed to appropriate agents.

[0096] As mentioned above, the integrated DNA intermediate is referred to as a provirus. Prior gene therapy or gene delivery systems use methods and retroviruses that require transcription of the provirus and assembly into infectious virus while in the presence of an appropriate helper virus or in a cell line containing

appropriate sequences enabling encapsidation without coincident production of a contaminating helper virus. As described below, a helper virus is not required for the production of the recombinant retrovirus of the disclosure, since the sequences for encapsidation are provided in the genome thus providing a replication competent retroviral vector for gene delivery or therapy.

[0097] A retroviral genome useful in the methods and compositions of the disclosure comprises a proviral DNA having at least three genes: the *gag*, the *pol*, and the *env*, these genes may be flanked by one or two long terminal (LTR) repeat, or in the provirus are flanked by two long terminal repeat (LTR) and sequences containing cis-acting sequences such as *psi*. The *gag* gene encodes the internal structural (matrix, capsid, and nucleocapsid) proteins; the *pol* gene encodes the RNA-directed DNA polymerase (reverse transcriptase), protease and integrase; and the *env* gene encodes viral envelope glycoproteins. The 5' and/or 3' LTRs serve to promote transcription and polyadenylation of the virion RNAs. The LTR contains all other cis-acting sequences necessary for viral replication. Lentiviruses have additional genes including *vif*, *vpr*, *tat*, *rev*, *vpu*, *nef*, and *vpx* (in HIV-1, HIV-2 and/or SIV).

[0098] Adjacent to the 5' LTR are sequences necessary for reverse transcription of the genome (the tRNA primer binding site) and for efficient encapsidation of viral RNA into particles (the *Psi* site). If the sequences necessary for encapsidation (or packaging of retroviral RNA into infectious virion) are missing from the viral genome, the result is a cis defect which prevents encapsidation of genomic viral RNA. This type of modified vector is what has typically been used in prior gene delivery systems (*i.e.*, systems lacking elements which are required for encapsidation of the virion).

[0099] In one embodiment, the disclosure provides a recombinant retrovirus capable of infecting a non-dividing cell, a dividing cell, or a cell having a cell proliferative disorder. The recombinant replication competent retrovirus of the disclosure comprises a polynucleotide sequence encoding a viral GAG, a viral POL, a viral ENV, a heterologous polynucleotide preceded by an internal ribosome entry site (IRES) enpasulated within a virion.

[00100] The heterologous nucleic acid sequence is operably linked to an IRES. As used herein, the term "heterologous" nucleic acid sequence or transgene refers to (i) a sequence that does not normally exist in a wild-type retrovirus, (ii) a sequence that originates from a foreign species, or (iii) if from the same species, it may be substantially modified from its original form. Alternatively, an unchanged nucleic acid sequence that is not normally expressed in a cell is a heterologous nucleic acid sequence. In a specific embodiment, the heterologous polynucleotide is a polypeptide of the disclosure having cytosine deaminase activity.

[00101] The phrase "non-dividing" cell refers to a cell that does not go through mitosis. Non-dividing cells may be blocked at any point in the cell cycle, (*e.g.*, G_0/G_1, $G_{1/S}$, $G_{2/M}$), as long as the cell is not actively dividing. For *ex vivo* infection, a dividing cell can be treated to block cell division by standard techniques used by those of skill in the art, including, irradiation, aphidocolin treatment, serum starvation, and contact inhibition. However, it should be understood that *ex vivo* infection is often performed without blocking the cells since many cells are already arrested (*e.g.*, stem cells). For example, a recombinant lentivirus vector is capable of infecting any non-dividing cell, regardless of the mechanism used to block cell division or the point in the cell cycle at which the cell is blocked. Examples of pre-existing non-dividing cells in the body include neuronal, muscle, liver, skin, heart, lung, and bone marrow cells, and their derivatives. For dividing cells onco-retroviral vectors can be used.

[00102] By "dividing" cell is meant a cell that undergoes active mitosis, or meiosis. Such dividing cells include stem cells, skin cells (*e.g.*, fibroblasts and keratinocytes), gametes, and other dividing cells known in the art. Of particular interest and encompassed by the term dividing cell are cells having cell proliferative disorders, such as neoplastic cells. The term "cell proliferative disorder" refers to a condition characterized by an abnormal number of cells. The condition can include both hypertrophic (the continual multiplication of cells resulting in an overgrowth of a cell population within a tissue) and hypotrophic (a

lack or deficiency of cells within a tissue) cell growth or an excessive influx or migration of cells into an area of a body. The cell populations are not necessarily transformed, tumorigenic or malignant cells, but can include normal cells as well. Cell proliferative disorders include disorders associated with an overgrowth of connective tissues, such as various fibrotic conditions, including scleroderma, arthritis and liver cirrhosis. Cell proliferative disorders include neoplastic disorders such as head and neck carcinomas. Head and neck carcinomas would include, for example, carcinoma of the mouth, esophagus, throat, larynx, thyroid gland, tongue, lips, salivary glands, nose, paranasal sinuses, nasopharynx, superior nasal vault and sinus tumors, esthesioneuroblastoma, squamous call cancer, malignant melanoma, sinonasal undifferentiated carcinoma (SNUC), brain (including glioblastomas) or blood neoplasia. Also included are carcinoma's of the regional lymph nodes including cervical lymph nodes, prelaryngeal lymph nodes, pulmonary juxtaesophageal lymph nodes and submandibular lymph nodes (Harrison's Principles of Internal Medicine (eds., Isselbacher, *et al.*, McGraw-Hill, Inc., 13th Edition, ppl850-1853, 1994). Other cancer types, include, but are not limited to, lung cancer, colon-rectum cancer, breast cancer, prostate cancer, urinary tract cancer, uterine cancer lymphoma, oral cancer, pancreatic cancer, leukemia, melanoma, stomach cancer, skin cancer and ovarian cancer.

[00103] In one embodiment, the heterologous polynucleotide within the vector comprises a cytosine deaminase that has been optimized for expression in a human cell. In a further embodiment, the cytosine deaminase comprises a sequence that has been human codon optimized and comprises mutations that increase the cytosine deaminase's stability (*e.g.*, reduced degradation or increased thermo-stability) compared to a wild-type cytosine deaminase. In yet another embodiment, the heterologous polynucleotide encodes a fusion construct comprising a cytosine deaminase (either human codon optimized or non-optimized, either mutated or non-mutated) operably linked to a polynucleotide encoding a polypeptide having UPRT or OPRT activity. In another embodiment, the heterologous

polynucleotide comprises a CD polynucleotide or fusion construct of the disclosure (*e.g.*, SEQ ID NO:3, 5, 11, 13, 15, or 17).

[00104] In another embodiment, replication competent retroviral vector can comprise a heterologus polynucleotide encoding a polypeptide comprising a cytosine deaminase (as described herein) and may further comprise a polynucleotide comprising a miRNA or siRNA molecule linked to a cell-type or tissue specific promoter.

[00105] The term "regulatory nucleic acid sequence" refers collectively to promoter sequences, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, enhancers and the like, which collectively provide for the replication, transcription and translation of a coding sequence in a recipient cell. Not all of these control sequences need always be present so long as the selected coding sequence is capable of being replicated, transcribed and translated in an appropriate host cell. One skilled in the art can readily identify regulatory nucleic acid sequence from public databases and materials. Furthermore, one skilled in the art can identify a regulatory sequence that is applicable for the intended use, for example, *in vivo*, *ex vivo*, or *in vitro*.

[00106] The term "promoter region" is used herein in its ordinary sense to refer to a nucleotide region comprising a DNA regulatory sequence, wherein the regulatory sequence is derived from a gene which is capable of binding RNA polymerase and initiating transcription of a downstream (3'-direction) coding sequence. The regulatory sequence may be homologous or heterologous to the desired gene sequence. For example, a wide range of promoters may be utilized, including viral or mammalian promoter as described above.

[00107] An internal ribosome entry sites ("IRES") refers to a segment of nucleic acid that promotes the entry or retention of a ribosome during translation of a coding sequence usually 3' to the IRES. In some embodiments the IRES may comprise a splice acceptor/donor site, however, preferred IRESs lack a splice acceptor/donor site. Normally, the entry of ribosomes into messenger RNA takes place via the cap located at the 5' end of all eukaryotic mRNAs. However, there are exceptions to this universal rule. The absence of a cap in some viral mRNAs suggests the

existence of alternative structures permitting the entry of ribosomes at an internal site of these RNAs. To date, a number of these structures, designated IRES on account of their function, have been identified in the 5' noncoding region of uncapped viral mRNAs, such as that, in particular, of picornaviruses such as the poliomyelitis virus (Pelletier *et al.*, 1988, Mol. Cell. Biol., 8, 1103-1112) and the EMCV virus (encephalo-myocarditis virus (Jang *et al.*, J. Virol., 1988, 62, 2636-2643). The disclosure provides the use of an IRES in the context of a replication-competent retroviral vector.

[00108] The heterologous nucleic acid sequence is typically under control of either the viral LTR promoter-enhancer signals or an internal promoter, and retained signals within the retroviral LTR can still bring about efficient integration of the vector into the host cell genome. Accordingly, the recombinant retroviral vectors of the disclosure, the desired sequences, genes and/or gene fragments can be inserted at several sites and under different regulatory sequences. For example, a site for insertion can be the viral enhancer/promoter proximal site (*i.e.*, 5' LTR-driven gene locus). Alternatively, the desired sequences can be inserted into a distal site (*e.g.*, the IRES sequence 3' to the env gene) or where two or more heterologous sequences are present one heterologous sequence may be under the control of a first regulatory region and a second heterologous sequence under the control of a second regulatory region. Other distal sites include viral promoter sequences, where the expression of the desired sequence or sequences is through splicing of the promoter proximal cistron, an internal heterologous promoter as SV40 or CMV, or an internal ribosome entry site (IRES) can be used.

[00109] In one embodiment, the retroviral genome of the disclosure contains an IRES comprising a cloning site for insertion of a desired polynucleotide sequence. In one embodiment, the IRES is located 3' to the env gene in the retroviral vector, but 5’ to the desired heterologous nucleic acid. Accordingly, a heterologous polynucleotide sequence encoding a desired polypeptide may be operably linked to the IRES.

[00110] In another embodiment a targeting polynucleotide sequence is included as part of the recombinant retroviral vector of the disclosure. The targeting polynucleotide sequence is a targeting ligand (*e.g.*, peptide hormones such as heregulin, a single-chain antibodies, a receptor or a ligand for a receptor), a tissue-specific or cell-type specific regulatory element (*e.g.*, a tissue-specific or cell-type specific promoter or enhancer), or a combination of a targeting ligand and a tissue-specific/cell-type specific regulatory element. Preferably, the targeting ligand is operably linked to the env protein of the retrovirus, creating a chimeric retroviral env protein. The viral GAG, viral POL and viral ENV proteins can be derived from any suitable retrovirus (*e.g.*, MLV or lentivirus-derived). In another embodiment, the viral ENV protein is non-retrovirus-derived (*e.g.*, CMV or VSV).

[00111] The recombinant retrovirus of the disclosure is therefore genetically modified in such a way that the virus is targeted to a particular cell type (*e.g.*, smooth muscle cells, hepatic cells, renal cells, fibroblasts, keratinocytes, mesenchymal stem cells, bone marrow cells, chondrocyte, epithelial cells, intestinal cells, neoplastic cells, glioma cells, neuronal cells and others known in the art) such that the nucleic acid genome is delivered to a target non-dividing, a target dividing cell, or a target cell having a cell proliferative disorder. Targeting can be achieved in two ways. The first way directs the retrovirus to a target cell by binding to cells having a molecule on the external surface of the cell. This method of targeting the retrovirus utilizes expression of a targeting ligand on the coat of the retrovirus to assist in targeting the virus to cells or tissues that have a receptor or binding molecule which interacts with the targeting ligand on the surface of the retrovirus. After infection of a cell by the virus, the virus injects its nucleic acid into the cell and the retrovirus genetic material can integrate into the host cell genome. The second method for targeting uses cell- or tissue-specific regulatory elements to promote expression and transcription of the viral genome in a targeted cell which actively utilizes the regulatory elements, as described more fully below. The transferred retrovirus genetic material is then transcribed and translated into proteins within the

host cell. The targeting regulatory element is typically linked to the 5' and/or 3' LTR, creating a chimeric LTR.

[00112] By inserting a heterologous nucleic acid sequence of interest into the viral vector of the disclosure, along with another gene which encodes, for example, the ligand for a receptor on a specific target cell, the vector is now target specific. Viral vectors can be made target specific by attaching, for example, a sugar, a glycolipid, or a protein. Targeting can be accomplished by using an antibody to target the viral vector. Those of skill in the art will know of, or can readily ascertain, specific polynucleotide sequences which can be inserted into the viral genome or proteins which can be attached to a viral envelope to allow target specific delivery of the viral vector containing the nucleic acid sequence of interest.

[00113] Thus, the disclosure includes in one embodiment, a chimeric env protein comprising a retroviral env protein operably linked to a targeting polypeptide. The targeting polypeptide can be a cell specific receptor molecule, a ligand for a cell specific receptor, an antibody or antibody fragment to a cell specific antigenic epitope or any other ligand easily identified in the art which is capable of binding or interacting with a target cell. Examples of targeting polypeptides or molecules include bivalent antibodies using biotin-streptavidin as linkers (Etienne-Julan *et al.*, J. Of General Virol., 73, 3251-3255 (1992); Roux *et al.*, Proc. Natl. Acad. Sci USA 86, 9079-9083 (1989)), recombinant virus containing in its envelope a sequence encoding a single-chain antibody variable region against a hapten (Russell *et al.*, Nucleic Acids Research, 21, 1081-1085 (1993)), cloning of peptide hormone ligands into the retrovirus envelope (Kasahara *et al.*, Science, 266, 1373-1376 (1994)), chimeric EPO/env constructs (Kasahara *et al.*, 1994), single-chain antibody against the low density lipoprotein (LDL) receptor in the ecotropic MLV envelope, resulting in specific infection of HeLa cells expressing LDL receptor (Somia *et al.*, Proc. Natl. Acad. Sci USA, 92, 7570-7574 (1995)), similarly the host range of ALV can be altered by incorporation of an integrin ligand, enabling the virus to now cross species to specifically infect rat glioblastoma cells (Valsesia-Wittmann *et al.*, J. Virol. 68, 4609-

4619 (1994)), and Dornberg and co-workers (Chu and Dornburg, J. Virol 69, 2659-2663 (1995)) have reported tissue-specific targeting of spleen necrosis virus (SNV), an avian retrovirus, using envelopes containing single-chain antibodies directed against tumor markers.

[00114] The disclosure provides a method of producing a recombinant retrovirus capable of infecting a target cell comprising transfecting a suitable host cell with the following: a vector comprising a polynucleotide sequence encoding a viral gag, a viral pol and a viral env, wherein the vector contains a cloning site for introduction of a heterologous gene, operably linked to a regulatory nucleic acid sequence, and recovering the recombinant virus.

[00115] The retrovirus and methods of the disclosure provide a replication competent retrovirus that does not require helper virus or additional nucleic acid sequence or proteins in order to propagate and produce virion. For example, the nucleic acid sequences of the retrovirus of the disclosure encode, for example, a group specific antigen and reverse transcriptase, (and integrase and protease-enzymes necessary for maturation and reverse transcription), respectively, as discussed above. The viral gag and pol can be derived from a lentivirus, such as HIV or an oncovirus such as MoMLV. In addition, the nucleic acid genome of the retrovirus of the disclosure includes a sequence encoding a viral envelope (ENV) protein. The env gene can be derived from any retroviruses. The env may be an amphotropic envelope protein which allows transduction of cells of human and other species, or may be an ecotropic envelope protein, which is able to transduce only mouse and rat cells. Further, it may be desirable to target the recombinant virus by linkage of the envelope protein with an antibody or a particular ligand for targeting to a receptor of a particular cell-type. As mentioned above, retroviral vectors can be made target specific by inserting, for example, a glycolipid, or a protein. Targeting is often accomplished by using an antibody to target the retroviral vector to an antigen on a particular cell-type (*e.g.*, a cell type found in a certain tissue, or a cancer cell type). Those of skill in the art will know of, or can readily ascertain without undue experimentation, specific methods to achieve delivery of a retroviral vector to a specific target. In one

embodiment, the env gene is derived from a non-retrovirus (*e.g.*, CMV or VSV). Examples of retroviral-derived env genes include, but are not limited to: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV), human immunodeficiency virus (HIV) and Rous Sarcoma Virus (RSV). Other env genes such as Vesicular stomatitis virus (VSV) (Protein G), cytomegalovirus envelope (CMV), or influenza virus hemagglutinin (HA) can also be used.

[00116] In one embodiment, the retroviral genome is derived from an onco-retrovirus or gammaretrovirus, and more particularly a mammalian onco-retrovirus or gamma retrovirus. By "derived" is meant that the parent polynucleotide sequence is an wild-type oncovirus which has been modified by insertion or removal of naturally occurring sequences (*e.g.*, insertion of an IRES, insertion of a heterologous polynucleotide encoding, for example, a polypeptide having cytosine deaminase activity of the disclosure and the like).

[00117] In another embodiment, the disclosure provides retroviral vectors that are targeted using regulatory sequences. Cell- or tissue-specific regulatory sequences (*e.g.*, promoters) can be utilized to target expression of gene sequences in specific cell populations. Suitable mammalian and viral promoters for the disclosure are described elsewhere herein. Accordingly, in one embodiment, the disclosure provides a retrovirus having tissue-specific promoter elements at the 5' end of the retroviral genome. Preferably, the tissue-specific regulatory elements/sequences are in the U3 region of the LTR of the retroviral genome, including for example cell- or tissue-specific promoters and enhancers to neoplastic cells (*e.g.*, tumor cell-specific enhancers and promoters), and inducible promoters (*e.g.*, tetracycline).

[00118] In some circumstances, it may be desirable to regulate expression. For example, different viral promoters with varying strengths of activity may be utilized depending on the level of expression desired. In mammalian cells, the CMV immediate early promoter if often used to provide strong transcriptional activation. Modified versions of the CMV promoter that are less potent have also been used when reduced levels of expression of the transgene are

desired. When expression of a transgene in hematopoietic cells is desired, retroviral promoters such as the LTRs from MLV or MMTV can be used. Other viral promoters that can be used include SV40, RSV LTR, HIV-1 and HIV-2 LTR, adenovirus promoters such as from the E1A, E2A, or MLP region, AAV LTR, cauliflower mosaic virus, HSV-TK, and avian sarcoma virus.

[00119] Similarly tissue specific or selective promoters may be used to effect transcription in specific tissues or cells so as to reduce potential toxicity or undesirable effects to non-targeted tissues. For example, promoters such as the PSA, probasin, prostatic acid phosphatase or prostate-specific glandular kallikrein (hK2) may be used to target gene expression in the prostate. Other promoters/regulatory domains that can be used are set forth in Table 1.

[00120] In certain indications, it may be desirable to activate transcription at specific times after administration of the gene therapy vector. This may be done with such promoters as those that are hormone or cytokine regulatable. For example in therapeutic applications where the indication is a gonadal tissue where specific steroids are produced or routed to, use of androgen or estrogen regulated promoters may be advantageous. Such promoters that are hormone regulatable include MMTV, MT-1, ecdysone and RuBisco. Other hormone regulated promoters such as those responsive to thyroid, pituitary and adrenal hormones may be used. Cytokine and inflammatory protein responsive promoters that could be used include K and T Kininogen (Kageyama *et al.*, 1987), c-fos, TNF-alpha, C-reactive protein (Arcone *et al.*, 1988), haptoglobin (Oliviero *et al.*, 1987), serum amyloid A2, C/EBP alpha, IL-1, IL-6 (Poli and Cortese, 1989), Complement C3 (Wilson *et al.*, 1990), IL-8, alpha-1 acid glycoprotein (Prowse and Baumann, 1988), alpha-1 antitypsin, lipoprotein lipase (Zechner *et al.*, 1988), angiotensinogen (Ron *et al.*, 1990), fibrinogen, c-jun (inducible by phorbol esters, TNF-alpha, UV radiation, retinoic acid, and hydrogen peroxide), collagenase (induced by phorbol esters and retinoic acid), metallothionein (heavy metal and glucocorticoid inducible), Stromelysin (inducible by phorbol ester, interleukin-1 and EGF), alpha-2 macroglobulin and alpha-1 antichymotrypsin. Tumor specific

promoters such as osteocalcin, hypoxia-responsive element (HRE), MAGE-4, CEA, alpha-fetoprotein, GRP78/BiP and tyrosinase may also be used to regulate gene expression in tumor cells.

[00121] In addition, this list of promoters should not be construed to be exhaustive or limiting, those of skill in the art will know of other promoters that may be used in conjunction with the promoters and methods disclosed herein.

TABLE 1 TISSUE SPECIFIC PROMOTERS

Tissue	Promoter
Pancreas	Insulin Elastin Amylase pdr-1 pdx-1 glucokinase
Liver	Albumin PEPCK HBV enhancer α fetoprotein apolipoprotein C α-1 antitrypsin vitellogenin, NF-AB Transthyretin
Skeletal muscle	Myosin H chain Muscle creatine kinase Dystrophin Calpain p94 Skeletal alpha-actin fast troponin 1
Skin	Keratin K6 Keratin K1
Lung	CFTR Human cytokeratin 18 (K18) Pulmonary surfactant proteins A, B and C CC-10 P1
Smooth muscle	sm22 α SM-alpha-actin
Endothelium	Endothelin-1 E-selectin von Willebrand factor TIE, KDR/flk-1 Melanocytes Tyrosinase
Adipose tissue	Lipoprotein lipase (Zechner *et al.*, 1988) Adipsin (Spiegelman *et al.*, 1989) acetyl-CoA carboxylase (Pape and Kim, 1989) glycerophosphate dehydrogenase (Dani *et al.*, 1989) adipocyte P2 (Hunt *et al.*, 1986)
Breast	Whey Acidic Protien (WAP)(Andres et al. PNAS 84:1299-1303 1987
Blood	β-globin

[00122] "Tissue-specific regulatory elements" are regulatory elements (*e.g.*, promoters) that are capable of driving transcription

of a gene in one tissue while remaining largely "silent" in other tissue types. It will be understood, however, that tissue-specific promoters may have a detectable amount of "background" or "base" activity in those tissues where they are silent. The degree to which a promoter is selectively activated in a target tissue can be expressed as a selectivity ratio (activity in a target tissue/activity in a control tissue). In this regard, a tissue specific promoter useful in the practice of the disclosure typically has a selectivity ratio of greater than about 5. Preferably, the selectivity ratio is greater than about 15.

[00123] It will be further understood that certain promoters, while not restricted in activity to a single tissue type, may nevertheless show selectivity in that they may be active in one group of tissues, and less active or silent in another group. Such promoters are also termed "tissue specific", and are contemplated for use with the disclosure. Accordingly, the tissue-specific regulatory elements used in the disclosure, have applicability to regulation of the heterologous proteins such as the polypeptides having cytosine deaminase activity of the disclosure as well as a applicability as a targeting polynucleotide sequence in retroviral vectors.

[00124] The retroviral vectors and CD polynucleotides and polypeptides of the disclosure can be used to treat a wide range of disease and disorders including a number of cell proliferative diseases and disorders (see, *e.g.*, U.S. Pat. Nos. 4,405,712 and 4,650,764; Friedmann, 1989, Science, 244:1275-1281; Mulligan, 1993, Science, 260:926-932, R. Crystal, 1995, Science 270:404-410, each of which are incorporated herein by reference in their entirety, see also, The Development of Human Gene Therapy, Theodore Friedmann, Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1999. ISBN 0-87969-528-5, which is incorporated herein by reference in its entirety).

[00125] The disclosure also provides gene therapy for the treatment of cell proliferative disorders. Such therapy would achieve its therapeutic effect by introduction of an appropriate therapeutic polynucleotide sequence (*e.g.*, a polypeptide of the disclosure having cytosine deaminase activity), into cells of

subject having the proliferative disorder. Delivery of polynucleotide constructs can be achieved using the recombinant retroviral vector of the disclosure.

[00126] In addition, the therapeutic methods (*e.g.*, the gene therapy or gene delivery methods) as described herein can be performed *in vivo* or *ex vivo*. For example, in the methods for treatment of cell proliferative diseases or disorders it may be useful to remove the majority of a tumor prior to gene therapy, for example surgically or by radiation. In some embodiments, the retroviral therapy may be preceded or followed by chemotherapy.

[00127] Thus, the disclosure provides a recombinant retrovirus capable of infecting a non-dividing cell, a dividing cell or a neoplastic cell, wherein the recombinant retrovirus comprises a viral GAG; a viral POL; a viral ENV; a heterologous nucleic acid (e.g., comprising a polypeptide of the disclosure having cytosine deaminase activity) operably linked to an IRES; and cis-acting nucleic acid sequences necessary for packaging, reverse transcription and integration. The recombinant retrovirus can be a lentivirus, such as HIV, or can be an oncovirus. As described above for the method of producing a recombinant retrovirus, the recombinant retrovirus of the disclosure may further include at least one of VPR, VIF, NEF, VPX, TAT, REV, and VPU protein. While not wanting to be bound by a particular theory, it is believed that one or more of these genes/protein products are important for increasing the viral titer of the recombinant retrovirus produced (*e.g.*, NEF) or may be necessary for infection and packaging of virion.

[00128] The disclosure also provides a method of nucleic acid transfer to a target cell to provide expression of a particular nucleic acid (*e.g.*, a heterologous sequence such as a polynucleotide encoding a polypeptide having cytosine deaminase activity). Therefore, in another embodiment, the disclosure provides a method for introduction and expression of a heterologous nucleic acid in a target cell comprising infecting the target cell with the recombinant virus of the disclosure and expressing the heterologous nucleic acid in the target cell. As mentioned above, the target cell can be any cell type including dividing, non-dividing, neoplastic,

immortalized, modified and other cell types recognized by those of skill in the art, so long as they are capable of infection by a retrovirus.

[00129] The disclosure also provides gene therapy for the treatment of cell proliferative or immunologic disorders. In one embodiment, a cell proliferative disorder is treated by introducing a CD polynucleotide of the disclosure, expressing the polynucleotide to produce a polypeptide comprising cytosine deaminase activity and contacting the cell with 5-fluorocytosine in an amount and for a period of time to produce a cytotoxic amount of 5-FU.

[00130] In addition, the disclosure provides polynucleotide sequence encoding a recombinant retroviral vector of the disclosure. The polynucleotide sequence can be incorporated into various viral particles. For example, various viral vectors which can be utilized for gene therapy include adenovirus, herpes virus, vaccinia, or, preferably, an RNA virus such as a retrovirus and more particularly a mammalian oncovirus. The retroviral vector can be a derivative of a murine, simian or human retrovirus. Examples of retroviral vectors in which a foreign gene (*e.g.*, a heterologous polynucleotide sequence) can be inserted include, but are not limited to: derivatives of Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), and Rous Sarcoma Virus (RSV). All of these vectors can transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated. In yet another embodiment, the disclosure provides plasmids comprising a recombinant retroviral derived construct. The plasmid can be directly introduced into a target cell or a cell culture such as NIH 3T3, HT1080 (human), CF2 (dog) or other tissue culture cells. The resulting cells release the retroviral vector into the culture medium.

[00131] The disclosure provides a polynucleotide construct comprising from 5′ to 3′: a promoter or regulatory region useful for initiating transcription; a psi packaging signal; a *gag* encoding nucleic acid sequence, a *pol* encoding nucleic acid sequence; an *env* encoding nucleic acid sequence; an internal ribosome entry site nucleic acid sequence; a heterologous polynucleotide encoding a marker, therapeutic (e.g., a polypeptide having cytosine deaminase

activity) or diagnostic polypeptide; and a LTR nucleic acid sequence. In specific embodiments, a heterologous polynucleotide encoding a polypeptide having cytosine deaminase activity may further comprise a domain encoding a polypeptide comprising UPRT or OPRT activity.

[00132] As described elsewhere herein and as follows the various segment of the polynucleotide construct of the disclosure (*e.g.*, a recombinant replication competent retoviral polynucleotide) are engineered depending in part upon the desired host cell, expression timing or amount, and the heterologous polynucleotide. A replication competent retroviral construct of the disclosure can be divided up into a number of domains that may be individually modified by those of skill in the art.

[00133] For example, the promoter can comprise a CMV promoter having a sequence as set forth in SEQ ID NO:19, 20, or 22 from nucleotide 1 to about nucleotide 582 and may include modification to one or more (*e.g.*, 2-5, 5-10, 10-20, 20-30, 30-50, 50-100 or more nucleic acid bases) so long as the modified promoter is capable of directing and initiating transcription. In one embodiment, the promoter or regulatory region comprises a CMV-R-U5 domain polynucleotide. The CMV-R-U5 domain comprise the immediately early promoter from human cytomegalovirus to the MLV R-U5 region. In one embodiment, the CMV-R-U5 domain polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22 from about nucleotide 1 to about nucleotide 1202 or sequences that are at least 95% identical to a sequence as set forth in SEQ ID NO:19, 20, or 22, wherein the polynucleotide promotes transcription of a nucleic acid molecule operably linked thereto. The gag domain of the polynucleotide may be derived from any number of retroviruses, but will typically be derived from an oncoretrovirus and more particularly from a mammalian oncoretrovirus. In one embodiment the gag domain comprises a sequence from about nucleotide number 1203 to about nucleotide 2819 or a sequence having at least 95%, 98%, 99% or 99.8% (rounded to the nearest 10^{th}) identity thereto. The pol domain of the polynucleotide may be derived from any number of retroviruses, but will typically be derived from an oncoretrovirus and more particularly from a mammalian oncoretrovirus. In one embodiment the

pol domain comprises a sequence from about nucleotide number 2820 to about nucleotide 6358 or a sequence having at least 95%, 98%, 99% or 99.9% (roundest to the nearest 10^{th}) identity thereto. The env domain of the polynucleotide may be derived from any number of retroviruses, but will typically be derived from an oncoretrovirus and more particularly from a mammalian oncoretrovirus. In some embodiments the env coding domain comprises an amphotropic env domain. In one embodiment the env domain comprises a sequence from about nucleotide number 6359 to about nucleotide 8323 or a sequence having at least 95%, 98%, 99% or 99.8% (roundest to the nearest 10^{th}) identity thereto. The IRES domain of the polynucleotide may be obtained from any number of internal ribosome entry sites. In one embodiment, IRES is derived from an encephalomyocarditis virus. In one embodiment the IRES domain comprises a sequence from about nucleotide number 8327 to about nucleotide 8876 or a sequence having at least 95%, 98%, or 99% (roundest to the nearest 10^{th}) identity thereto so long as the domain allows for entry of a ribosome. The heterologous domain can comprise a cytosine deaminase of the disclosure. In one embodiment, the CD polynucleotide comprises a human codon optimized sequence. In yet another embodiment, the CD polynucleotide encodes a mutant polypeptide having cytosine deaminase, wherein the mutations confer increased thermal stabilization that increase the melting temperature (Tm) by $10^{o}C$ allowing sustained kinetic activity over a broader temperature range and increased accumulated levels of protein. In one embodiment, the cytosine deaminase comprises a sequence as set forth in SEQ ID NO:19 from about nucleotide number 8877 to about 9353. The heterologous domain may be followed by a polypurine rich domain. The 3′ LTR can be derived from any number of retroviruses, typically an oncoretrovirus and preferably a mammalian oncoretrovirus. In one embodiment, the 3′ LTR comprises a U3-R-U5 domain. In yet another embodiment the LTR comprises a sequence as set forth in SEQ ID NO:19 from about nucleotide 9405 to about 9998 or a sequence that is at least 95%, 98% or 99.5% (rounded to the nearest 10^{th}) identical thereto.

[00134] The disclosure also provides a recombinant retroviral vector comprising from 5′ to 3′ a CMV-R-U5, fusion of the immediate

early promoter from human cytomegalovirus to the MLV R-U5 region; a PBS, primer binding site for reverse transcriptase; a 5′ splice site; a ψ packaging signal; a gag, ORF for MLV group specific antigen; a pol, ORF for MLV polymerase polyprotein; a 3′ splice site; a 4070A env, ORF for envelope protein of MLV strain 4070A; an IRES, internal ribosome entry site of encephalomyocarditis virus; a modified cytosine deaminase (thermostablized and codon optimized); a PPT, polypurine tract; and a U3-R-U5, MLV long terminal repeat. This structure is further depicted in Figure 1.

[00135] The disclosure also provides a retroviral vector comprising a sequence as set forth in SEQ ID NO:19, 20, or 22.

[00136] In another embodiment, the disclosure provides a method of treating a subject having a cell proliferative disorder. The subject can be any mammal, and is preferably a human. The subject is contacted with a recombinant replication competent retroviral vector of the disclosure. The contacting can be *in vivo* or *ex vivo*. Methods of administering the retroviral vector of the disclosure are known in the art and include, for example, systemic administration, topical administration, intraperitoneal administration, intra-muscular administration, intracranial, cerebrospinal, as well as administration directly at the site of a tumor or cell-proliferative disorder. Other routes of administration known in the art.

[00137] Thus, the disclosure includes various pharmaceutical compositions useful for treating a cell proliferative disorder. The pharmaceutical compositions according to the disclosure are prepared by bringing a retroviral vector containing a heterologous polynucleotide sequence useful in treating or modulating a cell proliferative disorder according to the disclosure into a form suitable for administration to a subject using carriers, excipients and additives or auxiliaries. Frequently used carriers or auxiliaries include magnesium carbonate, titanium dioxide, lactose, mannitol and other sugars, talc, milk protein, gelatin, starch, vitamins, cellulose and its derivatives, animal and vegetable oils, polyethylene glycols and solvents, such as sterile water, alcohols, glycerol and polyhydric alcohols. Intravenous vehicles include fluid and nutrient replenishers. Preservatives include antimicrobial, anti-oxidants, chelating agents and inert gases. Other

pharmaceutically acceptable carriers include aqueous solutions, non-toxic excipients, including salts, preservatives, buffers and the like, as described, for instance, in Remington's Pharmaceutical Sciences, 15th ed. Easton: Mack Publishing Co., 1405-1412, 1461-1487 (1975) and The National Formulary XIV., 14th ed. Washington: American Pharmaceutical Association (1975), the contents of which are hereby incorporated by reference. The pH and exact concentration of the various components of the pharmaceutical composition are adjusted according to routine skills in the art. See Goodman and Gilman's The Pharmacological Basis for Therapeutics (7th ed.).

[00138] For example, and not by way of limitation, a retroviral vector useful in treating a cell proliferative disorder will include an amphotropic ENV protein, GAG, and POL proteins, a promoter sequence in the U3 region retroviral genome, and all cis-acting sequence necessary for replication, packaging and integration of the retroviral genome into the target cell.

[00139] The following Examples are intended to illustrate, but not to limit the disclosure. While such Examples are typical of those that might be used, other procedures known to those skilled in the art may alternatively be utilized.

EXAMPLES

Example 1 Construction of modified CD genes and insertion into plasmid vectors.

[00140] Genetic enhancements to the wild type yeast cytosine deaminase gene have been made to include: (1) three positional mutations which change three amino acids (A23L, I140L and V108I) to increase thermal stability of the yeast cytosine deaminase protein and (2) additional gene sequence modifications to enhance human codon usage sequences to improve protein translation efficiency in human cells without further changes to the amino acid sequence.

[00141] Sequence design for CD included CDoptimized, CD-UPRT (+/- linker) and CD-OPRTase (+/- linker). The final cytosine deaminase coding sequence can comprise at the 5′ end a PSI1 site (full length) and 3′ end Not1 site plus poly A tail for PSI1/Not1 cassette based strategy.

[00142] The following sequence comprising a yeast cytosine deaminase was used for cloning, optimizing and mutation (the boxed nucleic acids comprise the restriction sites useed in subsequent methods for cloning:

AACACGATTATAAATGGTGACAGGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACA
TTGCCTATGAGGAGGCGGCCTTAGGTTACAAAGAGGGTGGTGTTCCTATTGGCGGATGTCTT
ATCAATAACAAAGACGGAAGTGTTCTCGGTCGTGGTCACAACATGAGATTTCAAAAGGGATC
CGCCACACTACATGGTGAGATCTCCACTTTGGAAAACTGTGGGAGATTAGAGGGCAAAGTGT
ACAAAGATACCACTTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATC
ATGTATGGTATTCCACGCTGTGTTGTCGGTGAGAACGTTAATTTCAAAAGTAAGGGCGAGAA
ATATTTACAAACTAGAGGTCACGAGGTTGTTGTTGTTGACGATGAGAGGTGTAAAAAGATCA
TGAAACAATTTATCGATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAGGCGGCCC
GCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:31)

The following Table summarizes the genes and resulting plasmid vectors that were made and their names.

Table: Vector constructs and names

Tocagen Code	Reference name	Original Name	5'LTR Prom	Envelope	Vector	IRES	Trans-gene	3'LTR
T5.0000	pACE-yCD	pACE-CD (Tai et al. 2005)	CMV	Ampho (4070A)	pACE	EMCV	Wt yeast CD	MLV U3
T5.0001	pAC3-yCD1	CDopt sequence	CMV	Ampho (4070A)	pAC3	EMCV	modified CD	MLV U3
T5.0002	pAC3-yCD2	CDopt+3pt	CMV	Ampho (4070A)	pAC3	EMCV	Modif-ied CD	MLV U3
T5.0003	pAC3-yCD2-U	Cdopt+3pt-UPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-UPRT	MLV U3
T5.0004	pAC3-yCD2-O	CDopt+3pt-OPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-OPRT	MLV U3
T5.0005	pAC3-yCD2-LO	CDopt+3pt-LINK-OPRT	CMV	Ampho (4070A)	pAC3	EMCV	CD2-L-OPRT	MLV U3
T5.0006	pAC3-eGFP	pAC3-emd, pAC3GFP	CMV	Ampho (4070A)	pAC3	EMCV	Emerald GFP	MLV U3
T5.0007	pAC3-yCD	pAC3-yCD	CMV	Ampho (4070A)	pAC3	EMCV	Wt yeast CD	MLV U3

[00143] The replication competent retroviral vector described by Kasahara *et al.* pACE-CD (U.S. Patent No. 6,899,871, the disclosure of which is incorporated herein) was used as a basis for additional modifications. A vector (pAC3-yCD)was modified to express a

modified yeast cytosine deaminase gene as described herein and was used in the contructs. See 1A below for a diagram of the vector construct for the initial transfected replication-competent retrovirus. CMV is the human CMV immediate early promoter, U3, R and U5 are the corresponding regions of the viral long terminal repeat (LTR). *Gag*, *pol* and *env* are the viral protein coding regions. 1B and 1D shows the plasmid structure and a sequence of the disclosure. After the genes were synthesized at a contractor (Bio Basic Inc., Markham, Ontario, Canada)they were inserted into the Psi1 –Not1 site of the pAC3 vector backbone (Figure 1). The plasmid backbone was normally generated by cutting the plasmid pAC3-eGFP with Psi1 and Not1 and purifying the large (about 11kb) fragment encoding the plasmid and retroviral backbone)

[00144] A. Humanized codon optimized CD gene (CDopt,aka CD1, T5.0001)A comparison of a human codon optimized cytosine deaminase of Conrad *et al.* and PCT WO 99/60008 indicates 91 total codons optimized in both, 36 codons identical, 47 codons had third base pair changes (all encode same amino acid) and 9 codons were different (however they encoded same amino acid). Of the 9 codons that differed:

AGC (Ser) to TCC (Ser)

CGT (Arg) to AGG (Arg)

CCA (Pro) to CCT (Pro)

All have equivalent GC content and encode the same amino acid.
The native yeast gene sequence above was separately codon optimized to give the following CD gene (CD1) and was called T5.0001 when when inserted into the plasmid vector pAC3 which encodes the replication competent retrovirus (RCR) with IRES.

TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTA**
CGAGGAGGCCGCCCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACA
ACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACC
CTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGA
CACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACG
GCATCCCTAGGTGTGTGGTGGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTG
CAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGATCATGAAGCA
GTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTGATAAGCGGCCGCAG
ATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG. (SEQ ID NO:32)

[00145] B. Heat stabilized CD gene Additional modifications were made to enhance the stability of the cytosine deaminase. Genetic enhancements to the wild type yeast cytosine deaminase gene were made to include three positional mutations which change three amino acids (A23L, I140L and V108I) to increase thermal stability of the yeast cytosine deaminase protein.

[00146] The following primer pairs were used in the generation of the gene for the cytosine deaminase polypeptide of the disclosure:

sense: 5'-tcgaggatatcggcgagtgaaacccgttattctttttggc-3' (SEQ ID NO:25)

antisense: 5'-gccaaaaagaataacgggtttcactcgccgatatcctcga-3'(SEQ ID NO:26)

sense: 5'tcggcgagtgatccggcggcggcgcctccggcggcggcgcctccggcggcggcgcc tccggcggcggcgccaacccgttatt-3'(SEQ ID NO:27)

antisense:5'-aataacgggttggcgccgccgccggaggcgccgccgccggaggcgccgccgc cggaggcgccgccgccggatcactcgccga-3'(SEQ ID NO:28)

[00147] To increase the stability of the native yeast CD, three amino acid substitutions were engineered into the protein. These substitutions were alone or in combination with human codon optimization.

[00148] The three amino acid substitutions are: A23L, V108I, I140L. A sequence encoding these substitutions is shown below.

ATGGTGACAGGGGGAATGGCAAGCAAGTGGGATCAGAAGGGTATGGACATTGCCTATGAGGA
GGCG**TTA**TTAGGTTACAAAGAGGGTGGTGTTCCTATTGGCGGATGTCTTATCAATAACAAAG
ACGGAAGTGTTCTCGGTCGTGGTCACAACATGAGATTTCAAAAGGGATCCGCCACACTACAT
GGTGAGATCTCCACTTTGGAAAACTGTGGGAGATTAGAGGGCAAAGTGTACAAAGATACCAC
TTTGTATACGACGCTGTCTCCATGCGACATGTGTACAGGTGCCATCATCATGTATGGTATTC
CACGCTGTGTC**ATC**GGTGAGAACGTTAATTTCAAAAGTAAGGGCGAGAAATATTTACAAACT
AGAGGTCACGAGGTTGTTGTTGTTGACGATGAGAGGTGTAAAAAG**TTA**ATGAAACAATTTAT
CGATGAAAGACCTCAGGATTGGTTTGAAGATATTGGTGAGTAGGCGGCCGCGCCATAGATAA
AATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:33)

[00149] The encoded polypeptide comprises the following sequence (substituted amino acids bold-underlinied):

```
  1 MVTGGMASKWDQKGMDIAYEEALLGYKEGGVPIGGCLINNKDGSVLGRGHNMRFQKGSAT
 61 LHGEISTLENCGRLEGKVYKDTTLYTTLSPCDMCTGAIIMYGIPRCVIGENVNFKSKGEK
121 YLQTRGHEVVVVDDERCKKLMKQFIDERPQDWFEDIGE-
```

[00150] Final construct design that integrates 3 amino acid substitutions A23L/V108I/I140L utilizing preferred codons and uses preferred human codon usage for entire sequence (this gene is called

CDopt+3pt [aka CD2] and T5.0002 when inserted into the plasmid vector pAC3 which encodes the RCR with IRES) (SEQ ID NO:34):

```
  1 ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTTACGAG
 61 GAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAAC
121 AAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACC
181 CTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAG
241 GACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATG
301 TACGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAAG
361 TACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTG
421 ATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTGATAA
```

Underlined codons denotes preferred codons for amino acid substitutions.

[00151] CDoptimized sequence design (human codon preference + 3 amino acid substitutions)

AACACGATTATAAATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATA
TCGCTTACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTG
ATCAACAACAAGGACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTC
CGCCACCCTGCACGGCGAGATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGT
ACAAGGACACCACCCTGTACACCACCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATC
ATGTACGGCATCCCTAGGTGTGTGATCGGCGAGAACGTGAACTTCAAGTCCAAGGGCGAGAA
GTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGTTGACGATGAGAGGTGTAAGAAGCTGA
TGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGGATATCGGCGAGTAAGCGGCC
GCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:35)

[00152] C. Construction of CD-UPRT fusion gene (CDopt+3pt-UPRT, [aka CDopt-UPRT and CD2-UPRT], T5.0003 in the pAC3 plasmid RCR vector)A fusion construct was also developed comprising a CD polypeptide as described above linked to a UPRT polypeptide to generate a CDoptimized-UPRT sequence using Scheme I as set forth in Figure 2A. The following primers were used to delete the stop-start between the CD and UPRT.

Primer sequences:

Primer Name	Primer Sequence (5' to 3') (SEQ ID NO:)
del118-123	5'-tcgaggatatcggcgagtgaaacccgttattctttttggc-3'(36)
del118-123-antisense	5'-gccaaaaagaataacgggtttcactcgccgatatcctcga-3'(37)

Primer Name	Length (nt.)	Tm	Duplex Energy at 68°C	Energy Cost of Mismatches
del118-123	40	79.06°C	-44.37 kcal/mole	21.1%

del118-123-antisense	40	79.06°C	-47.95 kcal/mole	20.3%

Primer Name	Primer-Template Duplex
del118-123	5'-tcgaggatatcggcgagtga------aacccgttattctttttggc-3' |||||||||||||||||||||||||||||||||||| ccaagctcctatagccgctcactatctacttgggcaataagaaaaaccgaag
del118-123-antisense	ggttcgaggatatcggcgagtgatagatgaacccgttattctttttggcttc |||||||||||||||||||||||||||||||||||| 3'-agctcctatagccgctcact------ttgggcaataagaaaaaccg

[00153] The resulting fusion polynucleotide comprises 1296 bp and the sequence set forth immediately below:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGAACCCGTTATTCTTTTTGGCTTCTCCATTCTTGTACCTTACATATCTTATATATTAT
CCAAACAAAGGGTCTTTCGTTAGCAAACCTAGAAATCTGCAAAAAATGTCTTCGGAACCATTTAAGAA
CGTCTACTTGCTACCTCAAACAAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAA
CTAGACCTGATTTCATTTTCTACTCCGATAGAATCATCAGATTGTTGGTTGAAGAAGGTTTGAACCAT
CTACCTGTGCAAAAGCAAATTGTGGAAACTGACACCAACGAAAACTTCGAAGGTGTCTCATTCATGGG
TAAAATCTGTGGTGTTTCCATTGTCAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTGTA
GGTCTGTGCGTATCGGTAAAATTTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTTATTCTAC
GAAAAATTACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCTGGCCACCGGTGG
TAGTGCTATCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGAGAGAATTTACTTCTTAA
ACCTAATCTGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTTCCCAGAGGTCAGAATTGTTACT
GGTGCCCTCGACAGAGGTCTAGATGAAAACAAGTATCTAGTTCCAGGGTTGGGTGACTTTGGTGACAG
ATACTACTGTGTTTAAGCGGCCGCGCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGG
GGGG (SEQ ID NO:38)

[00154] D. Construction of CD-linker UPRT fusion gene (CDopt+3pt-LINK-UPRT [aka CDopt-LINKER-UPRT and CD2-L-UPRT]). A fusion construct was also developed by cloning a linker (Ser-Gly-Gly-Gly-Gly)$_4$ domain between and in frame with the CD polyeptpide and the UPRT polyeptpide to generated a CDoptimized-linker-UPRT sequence using Scheme II as depicted in 2B. The following primers were used to insert the linker.

Primer Name	Primer Sequence (5' to 3') (SEQ ID NO:)
ins_60nt_after_477	5'-tcggcgagtgatccggcggcggcgcctccggcggcggcgcctccggcggcggcgcctccggcggcggcgccaacccgttatt-3' (39)
ins_60nt_after_477-antisense	5'-aataacgggttggcgccgccgccggaggcgccgccgccggaggcgcc

	gccgccggaggcgccgccgccggatcactcgccga-3' (40)			
Primer Name	**Length (nt.)**	**Tm**	**Duplex Energy at 68°C**	**Energy Cost of Mismatches**
ins_60nt_after_477	82	79.77°C	-30.19 kcal/mole	83.3%
ins_60nt_after_477-antisense	82	79.77°C	-32.31 kcal/mole	82.2%

[00155] The resulting contruct has size: 1356 bp and the sequence immediately below:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGC
GCCAACCCGTTATTCTTTTTGGCTTCTCCATTCTTGTACCTTACATATCTTATATATTATCCAAACAA
AGGGTCTTTCGTTAGCAAACCTAGAAATCTGCAAAAAATGTCTTCGGAACCATTTAAGAACGTCTACT
TGCTACCTCAAACAAACCAATTGCTGGGTTTGTACACCATCATCAGAAATAAGAATACAACTAGACCT
GATTTCATTTTCTACTCCGATAGAATCATCAGATTGTTGGTTGAAGAAGGTTTGAACCATCTACCTGT
GCAAAAGCAAATTGTGGAAACTGACACCAACGAAAACTTCGAAGGTGTCTCATTCATGGGTAAAATCT
GTGGTGTTTCCATTGTCAGAGCTGGTGAATCGATGGAGCAAGGATTAAGAGACTGTTGTAGGTCTGTG
CGTATCGGTAAAATTTTAATTCAAAGGGACGAGGAGACTGCTTTACCAAAGTTATTCTACGAAAAATT
ACCAGAGGATATATCTGAAAGGTATGTCTTCCTATTAGACCCAATGCTGGCCACCGGTGGTAGTGCTA
TCATGGCTACAGAAGTCTTGATTAAGAGAGGTGTTAAGCCAGAGAGAATTTACTTCTTAAACCTAATC
TGTAGTAAGGAAGGGATTGAAAAATACCATGCCGCCTTCCCAGAGGTCAGAATTGTTACTGGTGCCCT
CGACAGAGGTCTAGATGAAAACAAGTATCTAGTTCCAGGGTTGGGTGACTTTGGTGACAGATACTACT
GTGTTTAA GCGGCCGC GCCATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG
(SEQ ID NO:41)

[00156] E.Construction of CD-OPRT fusion gene (CDopt+3pt-OPRT [aka CDopt-OPRT and CD2-OPRT], T5.0004 when inserted into the pAC3 plasmid RCR vector) A fusion construct was also developed comprising a CD polypeptide as described above linked to an OPRT polypeptide to generated a CDoptimized-OPRTase (CD humanized+3ptmutation + OPRTase functional domain human) using Scheme III as shown in Figure 2C.

[00157] The resulting construct comprises a size of 1269 bp and the sequence immediately below:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAG**GCGGTCGCTCGTG**cagctttggggccattggtgacgggtctgtacgacgtgcaggct
ttcaagtttggggacttcgtgctgaagagcgggctttcctccccatctacatcgatctgcggggcat

cgtgtctcgaccgcgtcttctgagtcaggttgcagatattttattccaaactgcccaaaatgcaggca
tcagttttgacaccgtgtgtggagtgccttatacagctttgccattggctacagttatctgttcaacc
aatcaaattccaatgcttattagaaggaaagaaacaaaggattatggaactaagcgtcttgtagaagg
aactattaatccaggagaaacctgtttaatcattgaagatgttgtcaccagtggatctagtgttttgg
aaactgttgaggttcttcagaaggagggcttgaaggtcactgatgccatagtgctgttggacagagag
cagggaggcaaggacaagttgcaggcgcacgggatccgcctccactcagtgtgtacattgtccaaaat
gctggagattctcgagcagcagaaaaaagttgatgctgagacagttgggagagtgaagaggtttattc
aggagaatgtctttgtggcagcgaatcataatggttctccccttttctataaaggaagcacccaaagaa
ctcaGCTTCGGTGCACGTGCAGAGCTGCCCAGGATCCACCCAGTTGCATCGAAGTAAGCGGCCGCGCC
ATAGATAAAATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG (SEQ ID NO:42)

[00158] F.Construction of CD-linker-OPRT fusion gene (CDopt+3pt-LINK-OPRT, [aka CDopt-LINKER-OPRT and CD2-L-OPRT], T5.0005 in the pAC3 plasmid RCR vector)A fusion construct was also developed by cloning a linker (Ser-Gly-Gly-Gly-Gly)$_4$) domain between and in frame with the CD polyeptpide and the OPRT polyeptpide to generated a CDoptimized-linker-OPRT sequence using Scheme IV as shown in Figure 2D.

[00159] The resulting construct comprises a size of 1329 bp and the sequence immediately below:

AACACGA**TTATAA**ATGGTGACCGGCGGCATGGCCTCCAAGTGGGATCAAAAGGGCATGGATATCGCTT
ACGAGGAGGCCCTGCTGGGCTACAAGGAGGGCGGCGTGCCTATCGGCGGCTGTCTGATCAACAACAAG
GACGGCAGTGTGCTGGGCAGGGGCCACAACATGAGGTTCCAGAAGGGCTCCGCCACCCTGCACGGCGA
GATCTCCACCCTGGAGAACTGTGGCAGGCTGGAGGGCAAGGTGTACAAGGACACCACCCTGTACACCA
CCCTGTCCCCTTGTGACATGTGTACCGGCGCTATCATCATGTACGGCATCCCTAGGTGTGTGATCGGC
GAGAACGTGAACTTCAAGTCCAAGGGCGAGAAGTACCTGCAAACCAGGGGCCACGAGGTGGTGGTTGT
TGACGATGAGAGGTGTAAGAAGCTGATGAAGCAGTTCATCGACGAGAGGCCTCAGGACTGGTTCGAGG
ATATCGGCGAGTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGCGCCTCCGGCGGCGGC
GCCGCGGTCGCTCGTGcagctttggggccattggtgacgggtctgtacgacgtgcaggctttcaagtt
tggggacttcgtgctgaagagcgggctttcctccccatctacatcgatctgcggggcatcgtgtctc
gaccgcgtcttctgagtcaggttgcagatattttattccaaactgcccaaaatgcaggcatcagtttt
gacaccgtgtgtggagtgccttatacagctttgccattggctacagttatctgttcaaccaatcaaat
tccaatgcttattagaaggaaagaaacaaaggattatggaactaagcgtcttgtagaaggaactatta
atccaggagaaacctgtttaatcattgaagatgttgtcaccagtggatctagtgttttggaaactgtt
gaggttcttcagaaggagggcttgaaggtcactgatgccatagtgctgttggacagagagcagggagg
caaggacaagttgcaggcgcacgggatccgcctccactcagtgtgtacattgtccaaaatgctggaga
ttctcgagcagcagaaaaaagttgatgctgagacagttgggagagtgaagaggtttattcaggagaat
gtctttgtggcagcgaatcataatggttctccccttttctataaaggaagcacccaaagaactcaGCTT
CGGTGCACGTGCAGAGCTGCCCAGGATCCACCCAGTTGCATCGAAGTAAGCGGCCGCGCCATAGATAA
AATAAAAGATTTTATTTAGTCTCCAGAAAAAGGGGGG. (SEQ ID NO:43)

Example 2 Infectious Vector production

Vector can be produced in a number of ways, but the first step is to introduce the plasmid DNA vector into cells to allow production of infectious particles, that can then be harvested from the cell supernatant. Once infectious particles have been generated other methods of production can be implemented by those skilled in the art. Vector particles were generated by transient transfection of 293T cells (Pear *et al.* Proc Natl Acad Sci U S A. 90:8392-8396 1993). The 293T cells were thawed and put into culture, then passaged twice in T-75 flasks containing 15 mL of the DMEM medium that was prepared by mixing DMEM High Glucose medium (Hyclone# 30081, 500 mL) with FBS (Hyclone# SH30070, 50 mL), L-Glutamine (Cellgro# 25-005-CI, 5 mL), NEAA (Hyclone #SH30238, 5 mL), and Penicillin-strep (Cellgro# 30-002-CI, 5 mL). The flasks were incubated at 37°C and 5% CO2. After the 3rd passage cells were seeded in 6 T-25's, each containing 5 mL of the medium, at a cell density of 1.8 x 10^6 cells/T-25 (or 7.2 x 10^4 cells/cm^2). One day after seeding the T-25's, the cells were transfected with the T5.0002 plasmid that expressed the viral vector using the Calcium Phosphate Transfection Kit from Promega (Cat# E1200). Eighteen hours following transfection, the media in one set of the flasks (3 flasks each set) were replaced with fresh medium containing 10 mM NaB. The media in the 2nd set of the flasks were not replaced, which served as a control (zero NaB). Eight hours post NaB treatment, the media in all flasks were replaced with the fresh medium containing no NaB. The expression was allowed to continue for both sets of flasks until the next day (22 hours duration). The supernatants from both sets of flasks were harvested and assayed for their titers by qPCR expressed in Transducing Units (TU)/ml (see example 3).

The titer results are shown in the following table.

Condition	First titer	Second titer (after storing at -80oC for 68 days)
Without NaB	1.5 (±0.05) x 106 TU/mL	1.2 (±0.2) x 106 TU/mL
10 mM NaB	1.4 (±0.3) x 106 TU/mL	7.0 (±0.14) x 105 TU/mL

Subsequent vector preparations were produced in this manner, without sodium butyrate. The other vector plasmids have been used in the same way to generate vector preparations with titers between 10e5 TU/ml and 10e7 TU/ml. Such material can be further purified and concentrated, if desired, as described below and see also: US 5792643; T. Rodriguez et al. J Gene Med 9:233 2007; US Application 61218063. In certain embodiments of the disclosure the dosing was calculated by grams of brain weight. In such embodiments, the dosing of a replication competent retroviral vector of the disclosure useful in the methods for treatment can range from 10^3 to 10^7 TU per gram brain weight.

Example 3 Quantitative PCR titering assay

[00247] The functional vector concentration, or titer, is determined using a quantitative PCR-based (qPCR) method. In this method, vector is titered by infecting a transducible host cell line (*e.g.* PC-3 human prostatic carcinoma cells, ATCC Cat# CRL-1435) with a standard volume of vector and measuring the resulting amount of provirus present within the host cells after transduction. The cells and vector are incubated under standard culturing condition (37 °C, 5% CO2) for 24 hr to allow for complete infection prior to the addition of the anti-retroviral AZT to stop vector replication.
Next, the cells are harvested from the culture dish and the genomic DNA (gDNA) is purified using an Invitrogen Purelink gDNA purification kit and eluted from the purification column with sterile RNase-/DNase-free water. The A260/A280 absorbance ratio is measured on a spectrophotometer to determine the concentration and relative purity of the sample. The gDNA concentrations are normalized with additional RNase-/DNase-free water to the lowest concentration of any given set of gDNA preparations such that the input DNA for the qPCR is constant for all samples analyzed. Genomic DNA purity is further assessed by electrophoresis of an aliquot of each sample on an ethidium bromide stained 0.8% agarose gel. If the sample passes an A260/A280 absorbance range of 1.8-2.0 and shows a single band of gDNA, then the sample is ready for qPCR

analysis of provirus copy number of the vector. Using primers that interrogate the LTR region of the provirus (reverse-transcribed vector DNA and vector DNA that is integrated into the host gDNA), qPCR is performed to estimate the total number of transduction events that occurred when the known volume of vector was used to transduce the known number of cells. The number of transduction events per reaction is calculated from a standard curve that utilizes a target-carrying plasmid of known copy-number that is serial diluted from 1E7 to 1E1 copies and measured under identical qPCR conditions as the samples. Knowing how many genomic equivalents were used for each qPCR reaction (from the concentration previously determined) and how many transduction events that occurred per reaction, we determine the total number of transduction events that occurred based on the total number of cells that were present at the time of transduction. This value is the titer of the vector after dilution into the medium containing the cells during the initial transduction. To calculate the corrected titer value, the dilution is corrected for by multiplying through by the volume of culture and the volume of titer divided by the volume of titer. These experiments are performed in replicate cultures and analyzed by qPCR using triplicate measurements for each condition to determine an average titer and with its associated standard deviation and coefficient of variance.

Example 4 Expression levels measured by Western Blot

[00160] Figure 3 demonstrates that higher levels of the human codon optimized with the three mutations for higher stability are observed compared to wild type yCD protein in a Western blot analysis of U-87 cells infected with virus encoding either the wild type (ACE-yCD) or fully optimized (AC3-yCD2) cytosine deaminase genes.

Example 5 Genetic Stability of viral vectors.

[00161] It is recognized that after reverse transcription and the first integration event into treated cells, the DNA provirus and any subsequent progeny retrovirus has a conventional LTR structure from MLV on either end. This configuration has been shown to be stable after multiple cycles of infection (See Figure 4 below).

[00162] Approximately 10^6 naïve U-87 cells were initially infected with the viral vector at an MOI of 0.01, and grown until fully infected to complete a single cycle of infection. Supernatant is then repassed onto uninfected cells and the cycle repeated. Genomic stability of the yCD2 sequence was assessed by PCR amplification of the integrated provirus from the infected cells using MLV specific primers flanking the transgene insertion site. For each set of infections, amplification of the vector plasmid (pAC3-yCD2 and the Kasahara *et al.* vector pACE-CD) was also performed to track full-length amplicon sizing on the gel. The appearance of any bands smaller than full-length amplicon would be an indicator of vector instability. Such experiments demonstrated that a vector of the disclosure (T5.0002 – comprising the modified vector and CDopt+3pt (CD2) heterologous polynucleotide maintained stability for more passages than pACE-CD or T5.007 both of which carry the wild type yeast.

Example 6 Cell Killing Experiments

[00163] In *in-vitro* cell culture experiments, the experiments demonstrate that the cytosine deaminase in cells expressing the yCD2 protein is at least as active as that from cells expressing the wild type yCD protein, by performing 5-FC titrations on RG2 rat cells (Figure 5A) or U-87 cells (Figure 5B) infected either with virus made from pAC3-yCD2/T5.0002) [AC3-yCD2(V)] from pAC3-yCD/T5.0007 [AC3-yCD(V)] or the other vectors. Briefly, for U-87 cells, 5 days post infection at a multiplicity of infection of 0.1 (*i.e.* 100% infected) with either AC3-yCD (wild type CD) vector or AC3-yCD2 (thermostabilized & codon optimized) vector were subject to increasing amounts of 5-FC or 0.1 mM of 5-FU as a positive control for 8 days. On day 8 of 5-FC treatment, cell cultures were assessed for viability using an MTS assay (Promega CellTiter 96 AQUEOUS One Solution Proliferation Assay). Data shows comparable killing between the two retroviral vectors at increasing doses of 5-FC treatment. The RG2 cultures were treated similarly and also show if anything a slight shift in the killing curve towards lowerconcentration of 5-FC for virus from T5.0002CD expression assay. U87 cells were transduced at a multiplicity of infection (MOI) of 0.1, cultivated for 5 days to allow viral spread and and

cells from day 5 post transduction were harvested. The cells were then collected by centrifugation at 800 x g for 5 min. The supernatant was aspirated away from the cell pellet and washed with 5 mL of phosphate buffered saline (PBS) and again centrifuged at 800 x g for 5 min. The resulting cell pellet was taken up in 1.5 mL of PBS, resuspended by passage through a pipette tip and placed in a freezer at -20C. Cells were lysed by a freeze/thaw method. Previously resuspended cells were allowed to thaw at room temperature, passed through a pipette tip, mixed with protease inhibitor cocktail and again refrozen at -20C. Previous to the enzyme assay, the sample was again thawed at room temperature and passed through a pipette tip. The suspension was then centrifuged at 14,000 rpm in a tabletop centrifuge for 5 min. The supernatant was decanted away from the pellet and placed in a fresh eppendorf tube and placed on ice. yCD enzyme activity was assessed by using an HPLC assay.

[00164] The HPLC assay was performed on a Shimadzu LC20AT unit connected in series with a photoarray detector and autoinjector. The solid phase was a Hypersil BDS C18 HPLC column with a 5 um sphere size and 4.0 x 250 mm column dimensions. The mobile phase was 50 mM ammonium phosphate, pH 2.1, containing 0.01% tert-butylammonium perchlorate and 5% methanol; the system was equilibrated at 22 C. All reagents were ACS grade and solvents were HPLC grade. A reaction mix was made consisting of 800 µL with a final concentration of 0.125 mg/mL 5FC (1 mM) in PBS and placed in a 1.5 mL autosampler vial. The reaction was then initiated by adding 200uL of each cell lysate. The reaction / autosampler vials were placed in the auto sampler and 5uL of the reaction mixture was injected. Time points were taken periodically by retrieving a 5 uL aliquot from each reaction vial and analyzing on the HPLC column. The conversion rates of 5FC to 5FU were calculated by comparing the peak areas with known amounts from a previously generated standard curve of 5FU. The rate of 5FC conversion to 5FU was derived by plotting the amount of 5FU (in nmol) generated against its corresponding time interval. Protein concentration for the cell sample was derived and the Specific Activity of the cell lysate samples were calculated by dividing the conversion rate (nmol/min) by the amount of protein used in the

assay in mg. Figure 6 shows the specific activity of various vectors after 5 days on transduction at an MOI of 0.1. The data demonstrate that pACE-yCD (T5.0000) < pAC3-yCD1(T5.0001) <pAC3-CD2 (T5.0002)in terms of the specific activity of cytosine deaminase in tissue culture cells.

Example 7. Tumors treated with the fully modified CD gene (yCD2) are eliminated more efficiently than tumor treated with the unmodified yeast CD gene, and do not recur (PR-01-08-001).

[00165] To determine which vector construct gives rise to the most effective vector in a subcutaneous mouse/human xenograft model. Three different constructs were evaluated: T5.0001 (partially modified CD); T5.0007 (unmodified yeast CD gene); T5.0002 (fully modified yCD2 gene). Tumor growth, survival and tumor regression post 5-FC prodrug administration was evaluated in a subcutaneous model of human glioma (U87) in immunodeficient mice. Two different 5-FC concentrations were evaluated to determine a dose response seen between 5-FC and the tumor-vector constructs.

[00166] A total of 12 groups consisting of 9-11 female mice per group were studied. All mice underwent right dorsal flank implantation on Day 0 with either: a 98% mixture of an uninfected U-87 tumor cell line and a 2% mixture of a U-87 cell line that has been infected with one of three TOCA 511 vector constructs, T5.0002, T5.0001, T5.0007; or an uninfected control U-87 cell line (100%). Mice were inoculated with 2x10e6 cells/mouse. Mice in groups 1-3 had uninfected U87 cells, groups 4-6 had U87 cell mixture containing transduced T5.0002 U87, groups 7-9 had U87 cell mixture containing transduced with T5.0001, and groups 10-12 had U87 cell mixture containing transduced with T5.0007. The tumors were allowed to grow for 6 days, until tumor size was approximately 100mm^3. Each vector dose group of mice was randomized to receive one of two doses of 5-FC (200 or 500 mg/kg/day), administered as a single IP injection, beginning on Day 6, or no 5-FC. 5-FC administration continued daily for 28 consecutive days. Surviving mice after day 29 were evaluated for tumor size and then sacrificed on varying days with tumors, if present, extracted for analysis. At day 29, mice from group 4-12 were re-randomized and the mice were subdivided into

2 subgroups (either continued on 5-FC to monitor tumor regression or discontinued 5-FC treatment to monitor tumor regrowth).

Results: Tumors treated with T5.0002, T5.0000 and T5.0007 all demonstrated tumor regression to undetectable at 200 and 500 mg/kg dose at day 29. All T5.0002 treated animals did not reform tumors when 5FC treatment was discontinued, out to 39 days. In contrast some tumors treated with T5.0001 and T5.0007 did recur by the 39 day time point. The experiment was terminated at day 39.

Conclusion: T5.0002 plus 5-FC is a more effective antitumor therapy than T5.0001 or T5.0007 plu 5-FC.

[00167] A number of embodiments of the disclosure been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the disclosure. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising:

a human codon optimized polynucleotide encoding a polypeptide of SEQ ID NO:4, wherein the polypeptide comprises cytosine deaminase (CD) activity.

2. The isolated polynucleotide of claim 1, wherein the polynucleotide comprises a sequence as set forth in SEQ ID NO:3 and which comprises 1-50 silent mutations.

3. The isolated polynucleotide of claim 1 or 2, wherein the polynucleotide comprises SEQ ID NO:3.

4. The isolated polynucleotide of claim 1, further comprising a UPRT or OPRT polynucleotide encoding a polypeptide having UPRT or OPRT activity operably linked to the polypeptide having CD activity.

5. The isolated polynucleotide of claim 4, wherein the UPRT or OPRT polynucleotide is codon optimized.

6. The isolated polynucleotide of claim 4, wherein the UPRT or OPRT polynucleotide is separated from the polynucleotide encoding the polypeptide having CD activity by a linker.

7. An isolated polynucleotide comprising:

a first coding sequence encoding a polypeptide having cytosine deaminase activity; and

a second coding sequence encoding a polypeptide having URPT or OPRT activity,

wherein the first coding sequence is operably linked to the second coding sequence.

8. The isolated polynucleotide of claim 7, wherein the first and second coding sequences are separated by a sequence encoding a peptide linker.

9. The isolated polynucleotide of claim 7, wherein the polynucleotide encodes a polypeptide comprising a sequence as set foth in SEQ ID NO:12, 14, or 16.

10. A substantially purified polypeptide produced by expression of the polynucleotide of claim 1, 2, 4 or 7.

11. A vector comprising a polynucleotide of claim 1, 2, 4 or 7.

12. The vector of claim 11, wherein the vector is a plasmid.

13. The vector of claim 11, wherein the vector is an expression vector.

14. The vector of claim 11, wherein the vector comprises a viral vector.

15. The vector of claim 14, wherein the viral vector comprises a replication competent retrovirus.

16. A host cell containing the vector of claim 12, 13, 14, or 15.

17. A host cell comprising a polynucleotide of claim 1, 4 or 7.

18. The host cell of claim 17, wherein the host cell comprises a human cell.

19. The host cell of claim 18, wherein the host cell comprises a cell comprising a cell proliferative disorder.

20. The host cell of claim 17, wherein the host cell comprises a stably transformed human cell.

21. The host cell of claim 20, wherein the host cell comprises a cell proliferative disorder.

22. The host cell of claim 19 or 21, wherein the host cell comprises glioblastoma multiforme.

23. The host cell of claim 17, wherein the host cell is transformed *in vivo*.

24. A recombinant replication competent retrovirus (RCR) comprising:

a retroviral GAG protein;

a retroviral POL protein;

a retroviral envelope;

a retroviral polynucleotide comprising Long-Terminal Repeat (LTR) sequences at the 3' end of the retroviral polynucleotide sequence, a promoter sequence at the 5' end of the retroviral polynucleotide, said promoter being suitable for expression in a mammalian cell, a gag nucleic acid domain, a pol nucleic acid domain and an env nucleic acid domain;

a cassette comprising an internal ribosome entry site (IRES) operably linked to a heterologous nucleic acid comprising a polynucleotide of claim 1, wherein the cassette is positioned 5' to the 3' LTR and 3' to the env nucleic acid domain encoding the retroviral envelope; and

cis-acting sequences necessary for reverse transcription, packaging and integration in a target cell.

25. The retrovirus of claim 24, wherein the retroviral polynucleotide sequence is derived from murine leukemia virus (MLV), Moloney murine leukemia virus (MoMLV), Feline leukemia Virus or Gibbon ape leukemia virus (GALV).

26. The retrovirus of claim 24 or 25, wherein the MLV is an amphotropic MLV.

27. The retrovirus of claim 24, wherein the retrovirus is a gammaretrovirus.

28. The retrovirus of claim 24, wherein the target cell is a cell having a cell proliferative disorder.

29. The retrovirus of claim 24, wherein the target cell is a neoplastic cell.

30. The retrovirus of claim 28, wherein the cell proliferative disorder is selected from the group consisting of lung cancer, colon-rectum cancer, breast cancer, prostate cancer, urinary tract cancer, uterine cancer, brain cancer, head and neck cancer, pancreatic cancer, melanoma, stomach cancer and ovarian cancer, rheumatoid arthritis or other auto-immune disease.

31. The retrovirus of claim 24, wherein the promoter sequence is associated with a growth regulatory gene.

32. The retrovirus according to claim 24, wherein the promoter sequence comprises a tissue-specific promoter sequence.

33. The retrovirus according to claim 32, wherein the tissue-specific promoter sequence comprises at least one androgen response element (ARE).

34. The retrovirus according to claim 33, wherein the androgen response element is derived from the probasin promoter.

35. The retrovirus according to claim 32, wherein the tissue-specific promoter sequence comprises a probasin promoter.

36. The retrovirus according to claim 24, wherein the promoter comprises a CMV promoter having a sequence as set forth in SEQ ID NO:19, 20, or 22 from nucleotide 1 to about nucleotide 582 and may include modification to one or more nucleic acid bases and which is capable of directing and initiating transcription.

37. The retrovirus of claim 24, wherein the promoter comprises a sequence as set forth in SEQ ID NO:19 or 20 from nucleotide 1 to about nucleotide 582.

38. The retrovirus of claim 24, wherein the promoter comprises a CMV-R-U5 domain polynucleotide.

39. The retrovirus of claim 38, wherein the CMV-R-U5 domain comprise the immediately early promoter from human cytomegalovirus linked to an MLV R-U5 region.

40. The retrovirus of claim 39, wherein the CMV-R-U5 domain polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22 from about nucleotide 1 to about nucleotide 1202 or sequences that are at least 95% identical to a sequence as set forth in SEQ ID NO:19, 20, or 22, wherein the polynucleotide promotes transcription of a nucleic acid molecule operably linked thereto.

41. The retrovirus of claim 24, wherein the gag of the polynucleotide is derived from a gammaretrovirus.

42. The retrovirus of claim 41, wherein the gag nucleic acid domain comprises a sequence from about nucleotide number 1203 to about nucleotide 2819 of SEQ ID NO: 19 or a sequence having at least 95%, 98%, 99% or 99.8% identity thereto.

43. The retrovirus of claim 24, wherein the pol domain of the polynucleotide is derived from a gammaretrovirus.

44. The retrovirus of claim 43, wherein the pol domain comprises a sequence from about nucleotide number 2820 to about nucleotide 6358 of SEQ ID NO:19 or a sequence having at least 95%, 98%, 99% or 99.9% identity thereto.

45. The retrovirus of claim 24, wherein the env domain comprises a sequence from about nucleotide number 6359 to about nucleotide 8323

of SEQ ID NO:19 or a sequence having at least 95%, 98%, 99% or 99.8% identity thereto.

46. The retrovirus of claim 24, wherein the IRES is derived from an encephalomyocarditis virus.

47. The retrovirus of claim 46, wherein the IRES comprises a sequence from about nucleotide number 8327 to about nucleotide 8876 of SEQ ID NO:19 or a sequence having at least 95%, 98%, or 99% identity thereto.

48. The retrovirus of claim 24, wherein the heterologous nucleic acid comprises a polynucleotide having a sequence as set forth in SEQ ID NO:3, 5, 11, 13, 15 or 17.

49. The retrovirus of claim 24, wherein the heterologous nucleic acid encodes a polypeptide comprising a sequence as set forth in SEQ ID NO:4.

50. The retrovirus of claim 24, wherein the heterologous nucleic acid is human codon optimized and encodes a polypeptide as set forth in SEQ ID NO:4.

51. The retrovirus of claim 24, wherein the heterologous nuecleic acid comprises a sequence as set forth in SEQ ID NO: 19 from about nucleotide number 8877 to about 9353.

52. The retrovirus of claim 51, wherein the 3' LTR is derived from a gammaretrovirus.

53. The retrovirus of claim 52, wherein the 3' LTR comprises a U3-R-U5 domain.

54. The retrovirus of claim 52, wherein the 3' LTR comprises a sequence as set forth in SEQ ID NO:19 from about nucleotide 9405 to about 9998 or a sequence that is at least 95%, 98% or 99.5% identical thereto.

55. The retrovirus of claim 24, wherein the retroviral polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22.

56. An isolated polynucleotide comprising from 5′ to 3′:

a CMV-R-U5 fusion of the immediate early promoter from human cytomegalovirus to an MLV R-U5 region;

a PBS, primer binding site for reverse transcriptase;

a 5′ splice site;

ψ packaging signal;

a gag coding sequence for MLV group specific antigen;

a pol coding sequence for MLV polymerase polyprotein;

a 3′ splice site;

a 4070A env coding sequence for envelope protein of MLV strain 4070A;

an internal ribosome entry site (IRES) from encephalomyocarditis virus;

a modified cytosine deaminase coding sequence;

a polypurine tract; and

a U3-R-U5 MLV long terminal repeat.

57. The polynucleotide of claim 56, wherein the retroviral polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22.

58. The polynucleotide of claim 56, wherein the retroviral polynucleotide sequence is derived from murine leukemia virus (MLV), Moloney murine leukemia virus (MoMLV), Feline leukemia virus or Gibbon ape leukemia virus (GALV).

59. The polynucleotide of claim 58, wherein the MoMLV is an amphotropic MoMLV.

60. A method of treating a subject with a cell proliferative disorder comprising contacting the subject with a polynucleotide of

claim 1, 7 or 56 under conditions such that the polynucleotide is expressed and contacting the subject with 5-fluorocytosine.

61. The method of claim 60, wherein the polynucleotide is integrated into cell of the subject.

62. The method of claim 60, wherein the polynucleotide is delivered by a retroviral vector.

63. The method of claim 62, wherein the retroviral vector comprises a sequence as set forth in SEQ ID NO:19, 20, or 22.

64. The method of claim 60, wherein the cell proliferative disorder is glioblasoma multiforme.

65. The method of claim 60, wherein the cell proliferative disorder is selected from the group consisting of lung cancer, colon-rectum cancer, breast cancer, prostate cancer, urinary tract cancer, uterine cancer, brain cancer, head and neck cancer, pancreatic cancer, melanoma, stomach cancer and ovarian cancer, rheumatoid arthritis or other auto-immune disease.

66. A method of treating a cell proliferative disorder in a subject comprising contacting the subject with a retrovirus of claim 24, and contacting the subject with 5-fluorocytosine.

67. The method of claim 66, wherein the heterologous nucleic acid encodes a polypeptide comprising a sequence as set forth in SEQ ID NO:4, 12, 14, 16, or 18.

68. The method of claim 66, wherein the retroviral polynucleotide comprises a sequence as set forth in SEQ ID NO:19, 20, or 22.