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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 April 2019 (25.04.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/075562 A1

(51) International Patent Classification:

C07K 14/15 (2006.01)	*A61P 31/12* (2006.01)
A61K 31/35 (2006.01)	*A61P 35/00* (2006.01)
A61K 31/575 (2006.01)	*C07D 493/04* (2006.01)
A61K 31/7088 (2006.01)	*C07K 16/10* (2006.01)
A61K 39/395 (2006.01)	*C12N 15/48* (2006.01)
A61P 25/00 (2006.01)	*C12N 5/10* (2006.01)
A61P 31/00 (2006.01)	*C12P 21/02* (2006.01)

(21) International Application Number:
PCT/CA2018/051306

(22) International Filing Date:
17 October 2018 (17.10.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/573,290 17 October 2017 (17.10.2017) US

(71) Applicant: **UNIVERSITY OF MANITOBA** [CA/CA]; Room 202 Administration Building, Winnipeg, Manitoba R3T 2N2 (CA).

(72) Inventors: **DOUVILLE, Renée**; 121 Brookhaven Bay, Winnipeg, Manitoba R2J 2S5 (CA). **GURM, Mamneet**; 38 Lakebourne Drive, Winnipeg, Manitoba R2P 1K9 (CA).

(74) Agent: **VICKERS, Mark F.** et al.; Borden Ladner Gervais LLP, 100 Queen Street, Suite 1300, Ottawa, Ontario K1P 1J9 (CA).

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

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

(54) Title: ENDOGENOUS RETROVIRUS-K (ERVK) ENCODES AN ALTERNATE ENVELOPE PROTEIN

FIG 63

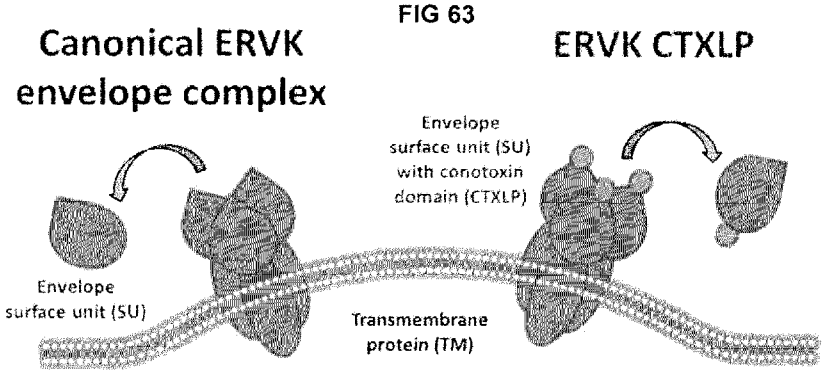


Figure 63: The ERVK Env and CTXLP proteins. Env is composed of the SU and TM subunits and is the prototypical gene product of the *env* gene. CTXLP is composed of the SU subunit and a C-terminal omega conotoxin domain and is encoded in an alternate open reading frame and may be produced due to ribosomal frameshifting.

(57) Abstract: The present disclosure relates to an endogenous Retrovirus K protein (ERVK) with an alternative envelope protein titled CTXLP. Said CTXLP peptide is represented by the sequences set forth in SEQ ID NO: 1. Additionally, antibodies that specifically recognize the epitope(s) set forth in SEQ ID NO:1 are and methods of use thereof and kits comprising the peptide set forth in SEQ ID NO:1 are also included in the present disclosure.

WO 2019/075562 A1

[Continued on next page]

Published:

— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

ENDOGENOUS RETROVIRUS-K (ERVK) ENCODES AN ALTERNATE ENVELOPE PROTEIN

CROSS REFERENCE TO RELATED APPLIATION

This application claims priority to United States Patent Application U.S. 62/573,290, filed October 17, 2017, the entire contents of which is hereby incorporated by reference.

FIELD

[0001] The present disclosure relates generally an endogenous Retrovirus-K (ERVK) alternate envelope protein.

BACKGROUND

[0002] Conotoxins are neurotoxic peptides found in the *Conus* genus of marine snails used to immobilize prey[22]. Conus species are distinct in their ability to produce hundreds of different toxic peptides[23]. Conotoxins are disulfide-rich and are usually 10-30 amino acids in length[22]. Conotoxins act as antagonists to specific voltage and ligand-gated ion channels[22]. In humans, symptoms of conotoxin exposure include poor coordination, blurred vision, speech difficulties, and nausea[23]. Conotoxins have also been associated with episodes of delirium and psychosis[24].

[0003] The O-superfamily of conotoxins exhibits an ICK fold. Members of the O-superfamily include μ-conotoxins, which inhibit voltage-gated sodium channels, and d-conotoxins, which delay sodium channel inactivation[25]. K-Conotoxins are inhibitors of voltage-gated potassium channels; ω -conotoxins inhibit N-type voltage-gated calcium channels (VGCCs)[25]. N-type VGCCs are located in presynaptic nerve terminals and are involved in neurotransmitter release[26]. ω -Conotoxin's selectivity for N-type VGCCs has allowed for their development as therapeutic agents. The ω -conotoxin MVIIA has been developed into a drug for relief of chronic and inflammatory pain[27].

[0004] Genes encoding an ω-conotoxin-like protein (CTXLP) have also been identified in certain viruses. Nuclear polyhedrosis viruses (NPV) have been shown to secrete a small conotoxin-like peptide[28]. NPVs are insect pathogens belonging to the family baculoviridae[28]. Although NPV-CTXLP's function has not been elucidated, its structure was found to have a nearly identical structure to the conserved ω-conotoxin's cysteine motif[28].

SUMMARY

[0005] In one aspect there is described an isolated polypeptide that comprises or consists of: an amino acid sequence having at least about 90% identity with the amino acid sequence set forth in SEQ ID NO:1 (CSDYGINCSHSYGCCSRSCIALFC).

[0006] In one example the isolated polypeptide comprises or consists of an amino acid sequence having at least about 95% identity with the amino acid sequence set forth in SEQ ID NO:1.

[0007] In one example the isolated polypeptide comprises or consists of the amino acid sequence set forth in SEQ ID NO:1.

[0008] In one aspect there is described an isolated nucleic acid molecule comprising a nucleotide sequence encoding a peptide comprising or consisting of an amino acid sequence having at least about 90% identity with the amino acid sequence set forth in SEQ ID NO:1.

[0009] In one example the isolated nucleic acid molecule comprises or consists of a nucleotide sequence having at least about 90% identity with the nucleotide acid sequence encoding the polypeptide of SEQ ID NO: 1.

[0010] In one example the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 95% identity with the nucleotide acid sequence encoding the polypeptide of SEQ ID NO: 1.

[0011] In one aspect there is described a vector comprising the nucleic acid molecule according to any one of claims 4 to 6.

[0012] In one aspect there is described a mammalian cell comprising the nucleic acid molecule of any one of claims 4 to 6.

[0013] In one example said mammalian cell is a human cell or non-human primate cell.

[0014] In one aspect there is described a host cell comprising the nucleic acid molecule of any one of claims 4 to 6.

[0015] In one example said host cell is a mammalian cell, an insect cell (such as Drosophila melanogaster), a bacteria cell, or a fungal cell.

[0016] In one aspect there is described a method for producing the peptide according to any one of claim 1 to 3, comprising: culturing a mammalian cell according to claim 8 or 9, or a host cell of claim 10 or 11, in a culture medium; and isolating the peptide from the mammalian cell of claims 8 or 9, or host cell of claim 10 or 11, or culture medium thereof.

[0017] In one aspect there is described an antibody that specifically recognizes the peptide of any one claims 1 to 3.

[0018] In one example said antibody is a monoclonal antibody or a polyclonal antibody.

[0019] In one aspect there is described a method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

[0020] In one aspect there is described a method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology.

[0021] In one example said condition or disorder is an infectious disease.

[0022] In one example said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

[0023] In one example said condition or disorder is a neurological disease.

[0024] In one example said neurological disease is amyotrophic lateral sclerosis (ALS), bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

[0025] In one example said condition or disorder is a cancer.

[0026] In one example said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

[0027] In one example said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease.

[0028] In one example said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

[0029] In one example said active agent comprises a Michael acceptor electrophile (MAE).

[0030] In one example said active agent comprises gambogic acid.

[0031] In one example said active agent comprises celastrol.

[0032] In one example said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

[0033] In one example said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

[0034] In one example further comprising administering a human anti-Nogo-A antibody.

[0035] In one example said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

[0036] In one aspect there is described a use of a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology for treating or preventing conditions or disorders associated with CTXLP in a subject.

[0037] In one aspect there is described a use of a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology in the manufacture of a medicament for treating or preventing conditions or disorders associated with CTXLP in a subject.

[0038] In one aspect there is described a use of a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology for treating or preventing conditions or disorders associated with ERVK in a subject.

[0039] In one aspect there is described a use of a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology in the manufacture of a medicament for treating or preventing conditions or disorders associated with ERVK in a subject.

[0040] In one example said condition or disorder is an infectious disease.

[0041] In one example said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

[0042] In one example said condition or disorder is a neurological disease.

[0043] In one example said neurological disease is amyotrophic lateral sclerosis (ALS), bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

[0044] In one example said condition or disorder is a cancer.

[0045] In one example said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

[0046] In one example said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease

[0047] In one example said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

[0048] In one example said active agent comprises a Michael acceptor electrophile (MAE).

[0049] In one example said active agent comprises gambogic acid.

[0050] In one example said active agent comprises celastrol.

[0051] In one example said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

[0052] In one example said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

[0053] In one example further comprising the use of a human anti-Nogo-A antibody.

[0054] In one example said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

[0055] In one aspect there is described a method for transcriptional activation, comprising contacting a DNA molecule comprising a gene with a peptide of any one of claims 1 to 3.

[0056] In one aspect there is described a diagnostic reagent for use in the detection of CTXLP protein in a subject, comprising an antibody of claims 11 or 12.

[0057] In one aspect there is described a diagnostic reagent for use in the detection of CTXLP mRNA in a subject, comprising an isolated nucleic acid according to any one of claims 4 to 6.

[0058] In one aspect there is described a diagnostic reagent for use in the detection CTXLP activity in a subject, comprising a peptide of any one of claims 1 to 3.

[0059] In one aspect there is described a method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier or a pharmaceutically acceptable salt thereof when the amount of

CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

[0060] In one aspect there is described a method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier or a pharmaceutically acceptable salt thereof when the amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

[0061] In one example said condition or disorder is an infectious disease.

[0062] In one example said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

[0063] In one example said condition or disorder is a neurological disease.

[0064] In one example said neurological disease is amyotrophic lateral sclerosis, bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

[0065] In one example said condition or disorder is a cancer.

[0066] In one example said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

[0067] In one example said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease

[0068] In one example said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

[0069] In one example said active agent comprises a Michael acceptor electrophile (MAE).

[0070] In one example said active agent comprises gambogic acid.

[0071] In one example said active agent comprises celastrol.

[0072] In one example said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

[0073] In one example said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

[0074] In one example, further comprising administering a human anti-Nogo-A antibody.

[0075] In one example said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

[0076] In one example the amount of CTXLP polypeptide is determined using an antibody of claim 13 or 14.

[0077] In one aspect there is described a kit comprising:(a) a container comprising a pharmaceutical composition containing the peptide of any one of claims 1 to 3, and/or a nucleic acid according to any one of claim 4 to 6, and/or a vector of claim 7, a mammalian cell of claims 8 or 9, a host cell of claims 10 or 11, and/or an antibody of claim 13 or 14, in solution or in lyophilized form;(b) optionally, a second container containing a diluent or reconstituting solution for the lyophilized formulation; and(c) optionally, instructions for use.

[0078] In one example further comprising one or more of (iii) a buffer, (iv) a diluent, (v) a filter, (vi) a needle, or (v) a syringe.

[0079] In one aspect there is described a method identifying CTXLP inhibitors, comprising: contacting a mammalian cell or a host cell, such as an insect cell (such as

Drosophila melanogaster), a bacteria cell, or a fungal cell, with a test compound or test composition, and measuring an amount of CTXLP protein, CTXLP-mRNA, CTXLP-regulated gene, or CTXLP-associate biomarker.

[0080] In one aspect there is described a method identifying CTXLP inhibitors, comprising: contacting a organoid, with a test compound or test composition, and measuring an amount of CTXLP protein, CTXLP-mRNA, CTXLP-regulated gene, or CTXLP-associate biomarker

BRIEF DESCRIPTION OF THE DRAWINGS

[0081] Embodiments of the present disclosure will now be described, by way of example only, with reference to the attached Figures.

[0082] Figure 1 depicts two types of ERVK genomes. The ERVK genome consists of four main Retroviridae genes, which are from 5' to 3': gag, pro, pol, and env. These viral genes are flanked by long terminal repeats (LTRs) containing U3, R and U5 regions. Two types of ERVK genomes can be distinguished based on a 292 bp deletion in the env gene.

[0083] Figure 2 depicts open reading frames on both strands of the endogenous retrovirus K-113 genome. ORFs (yellow) on both the sense and antisense strands were predicted using CLCbio software. Any amino acid-encoding codon was accepted as an ORF start, although each ended with a stop codon. Note the overlapping ORFs within known ERVK genes, such as gag, protease, polymerase and envelope.

[0084] Figure 3 depicts alignment of cone snail and viral omega conotoxin domain sequences. Sequences from 3 Conus species (black), 1 conotoxin-like protein domain sequence from Autographa Californica Nuclear Polyhedrosis Virus (blue), the consensus sequence generated from the aforementioned sequences (red) and the sequence of the putative Endogenous Retrovirus K-113 conotoxin-like protein domain. Modified from[28]. Note the characteristic C-C-CC-C-C knottin folding motif[29]

[0085] Figure 4 depicts an alignment and sequence logo of the putative endogenous retrovirus K-113 conotoxin-like protein domain and 10 Nuclear Polyhedrosis Virus conotoxin-like protein domain sequences. Sequences were aligned and sequence logo was assessed using Geneious v5 software[30]. Note the conserved C-G-NC-Y-CCS-C-A-FC sequence logo in these viral conotoxin-like proteins.

[0086] Figure 5 depicts an alignment of the ERVK CTXLP cysteine-rich motif which has strong similarity to both nuclear polyhedrosis virus (NPV, 46.2%) and Conus (45.8%) conotoxin proteins.

[0087] Figure 6 depicts the amino acid logo of the knottin domain from cluster representative CTXLP sequences.

[0088] Figure 7 depicts modeled 3-dimensional structure of the putative Endogenous Retrovirus-K113 conotoxin-like protein domain. Protein tertiary structure was predicted using Knotter1D3D software (gray sticks = carbon, green sticks = hydrogen, red sticks = oxygen, blue sticks = nitrogen and yellow spheres = sulfur). Note the interactions of the yellow cysteine residues, as they form disulfide bonds.

[0089] Figure 8 depicts aligned overlap of the predicted structures of viral conotoxin-like proteins from ERVK-113 and Ecotropis obliqua NPV. Knotter1D3D was used to predict the structures of putative ERVK-113 CTXLP domain (blue) and Ecotropis obliqua NPV CTXLP domain (red). Structure alignment is based on sequence alignment and was prepared using UCSF Chimera software31.

[0090] Figure 9 depicts an aligned overlap of the predicted structures of viral conotoxin-like protein backbones from ERVK-113 and Ecotropis obliqua NPV. Knotter1D3D was used to predict the structures of putative ERVK-113 CTXLP domain (blue) and Ecotropis obliqua NPV CTXLP domain (red). Structure alignment is based on sequence alignment and was prepared using UCSF Chimera software[31].

[0091] Figure 10 depicts a predicted inhibitor cysteine knot fold of ERVK CTXLP cysteine-rich peptide. Disulfide bonds connect cysteine 1 to cysteine 4, cysteine 2 to cysteine 5, and cysteine 3 to cysteine 6, resulting in an inhibitor cysteine knot fold.

[0092] Figure 11 depicts an alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and 12 spider toxin ICK peptides. Sequences were aligned and sequence logo generated using Geneious Software. A conserved C-C-CC-C-C motif is observed in all sequences. CTXLP, all Hainantoxin and one Guanxitoxin contained a conserved G between the first and second cysteine, as indicated by the star. Grey callouts indicate the cysteine motif spacing of each toxin.

[0093] Figure 12 depicts alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and agouti-related peptide and agouti signalling protein. Sequences were aligned and sequence logo generated using Geneious Software. A conserved C-C¬CC-C-C-C is observed in all sequences.

[0094] Figure 13 depicts alignment and sequence logo of ERVK CTXLP cysteine-rich peptide to 7 VEGF Proteins. Sequences were aligned and sequence logo generated using Geneious Software. No significant conservation was identified between the VEGF proteins and CTXLP, due to the differences in spacing and total number of cysteine residues, as indicated by the grey bar cysteine spacing motif.

[0095] Figure 14 depicts alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and HIV-1 and HIV-2 Tat proteins. Sequences were aligned and sequence logo generated using Geneious Software. Conservation of 6 of the 7 CTXLP cysteine residues are found in HIV Tat, as well as 1 lysine and 1 leucine residue in the C terminus, between amino acids 75 and 80.

[0096] Figure 15 depicts example Alignment of Chromosome 1 ERVK HML-2 insertions with the DNA sequences for Rec exon 1 and CTXLP. Sequences were aligned and sequence logo generated using Geneious software. Rec exon 1 aligned with bp 1 to 261 of env. CTXLP aligned with bp 1413 to 1505 of env.

[0097] Figure 16 depicts alignment of conotoxin-like peptides of 25 human ERVK HML-2 insertions. Alignment and sequence logo generated using Geneious Software. CTXLP-peptides showed variability in the amino acid sequence. Three distinct alleles were identified, as well as several unique sequences.

[0098] Figure 17 depicts CTXLP variants in the humans, based on genome build GRCh38.

[0099] Figure 18 depicts schematic representation of CTXLP and amino acid sequence similarities found using NCBI-CDD and Pfam databases. The SU subunit of CTXLP is red and the omega conotoxin domain is in green. The wider portions of the diagram represent the ordered regions of the proteins and the narrow region represent the disordered region as predicted by ELM resource56.

[00100] Figure 19 depicts analysis of the ERVK envelope transcript reveals prototypic RNA secondary structures. Shown is the predicted IRES-like RNA hairpin structures in the ERVK env transcript. The first 350 bp of ERVK Env-encoding RNA contains numerous AUG (methionine) translational start sites. Two distinct IRES-like hairpins are identified at nucleotides 84-187 and 213-318. tRNA can potentially bind at the AUG start site identified in IRES-like hairpin to produce a smaller isoform of ERVK Env or CTXLP. Alos shown is the predicted RNA secondary structure for ERVK-4 env transcript upstream and including the CTXLP ORF. Directly upstream of the CTXLP ORF translational start is a conserved -1

programmed ribosomal frameshifting sequence, which contains three elements i) a slippery site containing an X-XXY-YYZ motif which after frameshifting by - 1 results in XXX-YYY reading, ii) a 5 to 10 nucleotide spacer sequence, and iii) a downstream hairpin-type pseudoknot. The ERVK env transcript slippery site is encoded by a U-UUA-AAU sequence, followed by a 5 nucleotide spacer. Structure prediction formed with RNAfold Software shows a strong probability of hairpin-loops forming within the CTXLP-coding region, likely providing a downstream hairpin-type pseudoknot.

[00101] Figure 20 depicts the predicted CTXLP isoforms derived from the ERVK envelope transcript. Ribosomal frameshifting event resulting in formation of CTXLP peptide fused to surface unit protein. At the RNA slippery site within the ERVK Env transcript, the ribosome translating the RNA may bounce back by 4 nucleotides and begin reading in an alternate frame. This introduces a canonical KRQK nuclear localization sequence before proceeding into the CTXLP peptide. The resulting protein is a modified SU-CTXLP fusion protein. We show that ERVK can also produce SU-CTXLP fusion proteins. These CTXLP isoforms contain a nuclear localization sequence (NLS) and an additional N-linked glycosylation site at position 480.

[00102] Figure 21 depicts bioinformatic identification of CTXLP in the genome of endogenous retrovirus-K. ERVK113 was used as a template for the CTXLP domain in the ERVK envelope gene. The ERVK envelope polyprotein is cleaved by the cellular protease furin downstream of the R-X-R/K-R site. This splits the ERVK Env polyprotein into the surface unit (SU) and transmembrane (TM) proteins which interact to form the viral spike protein on the surface of virions. A -1 programmed ribosomal frameshift (-1 PRF) allows for the translation of the CTXLP cysteine-rich motif at the C-terminal end of the SU protein. Post-translational modification of ERVK SU protein includes glycosylation. N-linked N-X-S/T glycosylation sites are identified in red boxes.

[00103] Figure 22 depicts predicted post-translational modifications and protein interactions of CTXLP. (A) Schematic diagram of predicted glycosylation sites. (B)Schematic diagram of predicted phosphorylation and SUMOylation sites. (C) Schematic diagram of predicted protein cleavage sites. (D) Schematic diagram of predicted protein interaction sites.

[00104] Figure 23 depicts antigenic profile of the ERVK CTXLP domain, and predicted epitopes.

[00105] Figure 24 depicts rabbit immunization protocol for generation of a polyclonal antibody against the ERVK CTXLP domain.

[00106] Figure 25 depicts Western blot of ERVK-expressing NCCIT whole cell extract and immunoprecipitated CTXLP-enriched fraction. The far-right lane in the image is an image of the left α-CTXLP lane that was over-exposed to bring out the details in the bands.

[00107] Figure 26 depicts PNGase treatment of IP-purified CTXLP protein results in a decrease in western blot band size associated with removal of N-linked glycosylation moieties.

[00108] Figure 27 depicts ERVK Env, and CTXLP expression in SVGA cells treated with 0.1 ng/mL TNFα, 1 ng/mL TNFα, and 1 ng/mL LIGHT. NCCIT cells were used as a positive control and β-actin as a loading control. n = 3.

[00109] Figure 28 depicts ERVK CTXLP is inducible in human neurons. ReNcell-derived neurons were treated with increasing doses of pro-inflammatory cytokines TNFα or LIGHT for 24 hours. CTXLP expression (90kDa form) was enhanced optimally with 1ng/ml TNFα and 10ng/ml LIGHT treatment, n=1.

[00110] Figure 29 depicts CTXLP protein expression predominantly localizes with chromatin in NCCIT and SVGA cells. CTXLP expression was also identified in the cytosolic and nuclear fraction and soluble and insoluble whole cell lysates of NCCIT cells. Moreover, CTXLP appeared dispersed throughout NCCIT cells in confocal imaging. In contrast, SVGA cells exhibited expression of the small (32kDa) and large (90-110kDa) isoforms of CTXLP in association with the chromatin (A). This was supported by nuclear localization of CTXLP in SVGA cells as shown by confocal.

[00111] Figure 30 depicts ERVK CTXLP is inducible in astrocytes with pro-inflammatory cytokines. Confocal analysis of protein expression for ERVK CTXLP (red) and ERVK reverse transcriptase (RT, green) in cells treated with or without TNFα and LIGHT, n=2. Enhanced CTXLP expression precedes increases in RT expression. DAPI stain indicates nuclei (blue).

[00112] Figure 31 depicts the cellular localization of ERVK CTXLP and SU proteins in human astrocytes. Increased CTXLP protein and Env expression were identified (but not co-localized) in TNFα-treated cells compared to controls. Increased CTXLP expression indicated by arrows. Distinct puncta were identified within the 1) cytoplasm, 2) nucleus, and 3) surface membrane of the cell.

[00113] Figure 32 depicts CTXLP overexpression in soluble fractions of SVGA cells enhances the 32 kDa and 90 kDa forms of endogenous CTXLP. SVGA cells were

transfected using Lipofectamine LTX with 5 µg Empty Vector, or 0.5, 2 and 5 µg CTXLP cysteine-rich construct for 48 hr.

[00114] Figure 33 depicts that CTXLP binds interferon response elements (ISREs) within the ERVK promoter (5' LTR). CTXLP may regulate ERVK gene expression, as well as other genes containing ISREs. Chromatin immunoprecipitation (ChIP) following 8 hours of 10 ng/ml TNFα or LIGHT treatment in human ReNcell-derived neurons (n=2) and human astrocytic cell line (SVGA) (n=3). Notable increase in CTXLP chromatin binding in neurons upon pro-inflammatory stimulation with the cytokine TNFα.

[00115] Figure 34 depicts ERVK Expression Multidimensional Scaling. Two dimensional plots were produced by reduction of the high dimensional RNA-seq expression data derived from the Sequence Read Archive (SRA) using the R package EdgeR. The plot labels correspond to SRA studies as follows: ALS (SRP064478), Bipolar Disorder (SRP074904), Breast Cancer (SRP058722), HIV/HCV (SRP068424), Multiple Sclerosis (SRP110016), Prostate Cancer (ERP000550), Rheumatoid Arthritis (SRP102685), and Schizophrenia (SRP090259). In these plots each axis represents the leading log-2-scaled fold-change at one particular ERVK locus; the two loci chosen are those with the most extreme values in the majority of the clinical group's samples. Since these represent the biggest difference between samples, if no separation is apparent in these plots, there is no clear difference in transcript profiles. Each sample is represented by its SRA accession number, coloured red for disease-associated samples and black are the control samples.

[00116] Figure 35 depicts ERVK Multidimensional Scaling by CTXLP status in Human Disease. Each plot was produced as described in Figure 34, except with the modification that ERVK loci were subsetted into three states: "CTXLP+" which could produce CTXLP, "Disrupted" which cannot produce CTXLP but may have had an ancestral loci that produced CTXLP, and finally "CTXLP-" loci which do not and likely never did produce CTXLP from the ERVK env gene. In all cases, samples from the ALS, Bipolar Disorder, Breast Cancer, HIV-1/HCV co-infection, Multiple Sclerosis and Rheumatoid Arthritis cohorts expressed all three types of ERVK env transcripts.

[00117] Figure 36 depicts Per-Locus Differential ERVK Expression. Each panel shows in detail the expression of individual ERVK loci encoding envelope in each human disease condition. ERVK loci (black) are plotted against CTXLP+ (red), CTXLP- (blue) and disrupted (grey) loci. The plot labels correspond to SRA studies as follows: ALS (ALS: SRP064478), Bipolar Disorder (BP: SRP074904), Breast Cancer (BC: SRP058722), HIV/HCV (HV;

HIV/HCV + interferon (HI): SRP068424), Multiple Sclerosis (MS: SRP110016), Prostate Cancer (PC: ERP000550), Rheumatoid Arthritis (RA: SRP102685), and Schizophrenia (SZ: SRP090259). For each study, controls (C) and indicated to the right of cases. Panel A exclude ERVK loci with very low expression; only loci with a median expression greater than 0 and a mean expression greater than 0.1 are plotted. Panel B shows only loci which were highly expressed; only ERVK loci which had a maximum expression higher than 2 are plotted.

[00118] Figure 37 depicts ERVK CTXLP encoding transcripts and CTXLP protein are present in Amyotrophic Lateral Sclerosis (ALS). Re-analysis of RNAseq data6 in sporadic ALS and control spinal cords for expression of disrupted non-coding (black), Env+/CTXLP- (blue) and Env+/CTXLP+ (red) env transcripts. Principle component analysis (PCA) reveals ALS patient clustering in terms of CTXLP+ transcript expression, with most frequently expressed CTXLP encoding loci indicated.

[00119] Figure 38 depicts ERVK CTXLP levels are enhanced in autopsy spinal cord and brain tissues of patients with ALS, as measured by western blot analysis. Bar graph represents total CTXLP (A) and CX3CL1 (B) quantification in NN (n=9) and ALS (n=15) motor cortex specimens, as measured by western blot. Bar graph represents total CTXLP (C), CX3CL1 (D), ERVK Env SU (E) and CaV2.2 (F) quantification in NN (n=6) and ALS (n=13) cervical spinal cord specimens, as measured by western blot.

[00120]

[00121] Figure 39 depicts confocal micrographs of ERVK CTXLP levels being enhanced in autopsy spinal cord and brain tissues of patients with ALS. (A) Representative 10x confocal micrographs of ERVK CTXLP expression in ex vivo cervical spinal cord of a neuronormal control (NN, n=3) and patient with ALS (n=3). DAPI stain depicts nuclei. High magnification reveals staining in cells surrounding MAP2+ axons, suggesting CTXLP+ oligodendrocytes. CTXLP+ rings ranged from 6-16μM in diameter. (B) Representative 40x confocal micrographs of ERVK CTXLP, voltage-gated calcium channel CaV2.2 (C-terminal antibody) and neuronal MAP2 expression in Brodmann area 6 (BA6) motor cortex tissue of a NN control (n=3) and patient with ALS (n=3). Note the translocation of nuclear CTXLP to cytoplasmic aggregates in neurons from an ALS patient, as well as an overall decrease in CaV2.2 expression. DAPI stain depicts nuclei.

[00122] Figure 40 depicts micrographs showing that ERVK CTXLP levels are enhanced in autopsy cervical and lumbar spinal cord tissues from patients with ALS, as

measured by light and confocal microscopy. Representative 10x confocal micrographs of ERVK CTXLP expression in ex vivo cervical (CC) and lumbar (LC) spinal cord of a neuronormal control (NN, n=3) and patients with ALS (n=3). Solochrome cyanine (SC) stain (purple) with eosin counterstain (pink) depicts tissue myelination; pale lesions appear in ALS tissues. These lesioned areas exhibit increased CTXLP expression is in red. Oligodendrocyte precursor marker TCF4 is in green. DAPI stain depicts cellular nuclei. Note that CTXLP expression occurs in either lateral and/or anterior cortical spinal tracts.

[00123] Figure 41 depicts confocal micrographs of ERVK CTXLP levels are associated with demyelination and CTXLP is enhanced in TCF4+Olig1+ oligodendrocyte precursors in cervical spinal cord tissues from patients with ALS. Representative 20x confocal micrographs of ERVK CTXLP expression in ex vivo cervical (CC) spinal cord of a neuronormal control (NN, n=3) and a patient with ALS (n=3). Notably decreased myelin stain as measure my MAG protein expression (green) is evident in ALS tissue as compared to control. In ALS tissue, CTXLP expression (red) co-localizes with TCF4 (green) or Olig1 (green) markers indicative of oligodendrocyte precursor cells. DAPI stain depicts cellular nuclei.

[00124] Figure 42 depicts confocal micrographs showing ERVK CTXLP+ oligodendrocyte precursors express myelin inhibitory protein Nogo-A, or are in close proximity to Nogo-A positive cells in spinal cord tissues of patients with ALS. Human ex vivo cervical spinal cord tissues were stained for ERVK CTXLP (red), TCF4 (green), Nogo-A (grey) and nuclei (blue) in neuro-normal controls (n=3) and patients with ALS (n=3). Image merging for CTXLP and TCF4 indicate that oligodendrocyte precursors express CTXLP in ALS. Image merging for CTXLP and Nogo-A indicate that oligodendrocyte precursors can express myelin inhibitor protein Nogo-A (left panel) or alternately are in proximity to Nogo-A expressing cells in ALS (right panel). White stars indicate areas that are magnified to depict overlapping protein expression in CTXLP+ rings.

[00125] Figure 43 depicts cancer cells express greater levels of CTXLP as compared to non-cancer cells. Prototypic cell lines for teratocarcinoma (NCCIT) and breast cancer (T47D) were examined for CTXLP expression as compared to astrocytic SVGA cells using confocal microscopy. No antibody negative control is to show that specificity of CTXLP (red) staining requires an antibody targeting ERVK CTXLP. Nuclei are shown in blue using a DAPI stain.

[00126] Figure 44 depicts CTXLP expression in G-Bioscience Ready-to-screen cancer tissue and cell line blots. TB56-I (A), TB55 (B) and TB56-II (C) blots were screen for CTXLP expression (blue bars, top blot) normalized to β-actin loading control (lower blot). Enhanced CTXLP expression in noted is several cancer types, including T cell lymphoma, neuroepithelioma, prostate, ovary, testis and skin cancers.Figure 45 depicts changes in the gene expression of pro-inflammatory NF-κB p65 and anti-viral IRF7 in response to CTXLP and SU expression. 293T cells were transfected with plasmids encoding empty vector, ERVK CTXLP or ERVK SU for 24 hours. Cell pellets were collected, RNA extracted and cDNA produced for use in Q-PCR experiments to evaluate relative gene expression of RELA and IRF7. Analysis performed using ΔΔCt method and 18S RNA as a calibrator.

[00127] Figure 46 depicts confocal images of control, CTXLP-expressing and SU-expressing 293T cells. Cells were stained with DAPI nuclear stain, and antibodies against Env SU (green) fluorescent dye, NF-KB p65 (red) fluorescent dye and CTXLP (grey / white) fluorescent dye. Only CTXLP expressing cells exhibit upregulated NF-KB p65 expression. Representative micrographs of n=3 experiments.

[00128] Figure 47 depict western blot and confocal micrographs of ERVK CTXLP, but not ERVK Env SU, depleting CaV2.2 calcium channel-associated transcription regulator (CCAT). CaV2.2 expression in SVGA cells treated with 5µl of immunoprecipitation (IP) products from NCCIT cell lysates extracted using rabbit pre-immune serum, custom rabbit anti-CTXLP antibody or custom rabbit anti-ERVK Env SU antibody for 24 hours. Quantification of CaV2.2 depletion in IP-product treated astrocytes (2hrs), based on confocal quantification (****$p<0.0001$, 80-100 cells per condition quantified). Confocal imaging of CaV2.2 (N-terminal antibody) and CaV2.2 CCAT (C-terminal antibody) illustrates that CTXLP depletes nuclear CaV2.2 CCAT within 2 hours, but not the membrane-associated CaV2.2 channel (n=2).

[00129] Figure 48 depicts Caspase-3 expression in SVGA cells in control and CTXLP-treated conditions after 24 hours. (A) Images of live cells after 24 hours taken using EVOS imager, top image depicts untreated control cells and bottom image depicts cells treated with 5µl of CTXLP immunoprecipitated solution. (B) Graphical depiction of caspase-3 expression in control and CTXLP-treated cell after 24 hours. .

[00130] Figure 49 depicts CTXLP induces caspase-3 activation and apoptosis, which can be blocked by excess extracellular calcium. Cell survival 1 hour and 24 hours post treatment with 5µl of buffer, calcium chloride, CTXLP or CTXLP and calcium chloride. Cells

were examined using an EVOS microscopy for caspase-3 activation (green) or nuclei (blue), n=2). Excess calcium chloride is known to block the cellular effects of conotoxin proteins[81].

[00131] Figure 50 depicts cell survival and confluency 24 hours post treatment with 5µl of pre-immune serum, pre-immune serum and calcium chloride, CTXLP, CTXLP and calcium chloride, SU, or SU and calcium chloride. (A) Graphical depiction of caspase-3 expression in SVGA cells 24 hours post treatment. (B) Graphical depiction of cell confluency of SVGA cells 24 hours post-treatment. Note that pre-immune serum was used as a negative control as this is the component of immunoprecipitation product.

[00132] Figure 51 depicts Live cell images of SVGA cells in control, SU-treated and CTXLP-treated conditions stained for caspase-3 using EVOS live cell imaging. Cells in each condition were imaged after 5 days. Control cells were imaged at 7 days and SU and CTXLP-treated cells were imaged after 8 days. Controls cells express greater amounts of the apoptotic marker caspase-3, and have not proliferated to the extent observed in CTXLP and Env treatments.

[00133] Figure 52 depicts percentage of control and CTXLP-transfected SVGA cells expressing caspase-3 after 24 hours. Graphical depiction of apoptosis marker caspase-3 in control and CTXLP-transfected SVGA cell after 24 hours.

[00134] Figure 53 depicts abnormal cellular morphology in cells exposed to CTXLP. (A) Fluorescent image of CTXLP-treated SVGA cell expressing caspase-3 (green) and stained with DAPI nuclear stain (blue). (B) Confocal image of Env SU expression (green) in a CTXLP-expressing 293T cell.

[00135] Figure 54 depicts CTXLP-limiting drug screen in human NCCIT teratocarcinoma cells. The ERVK-expressing NCCIT teratocarcinoma cell line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 (data not shown) and 48 hours with known IC50 concentrations of drugs (NCCIT cells alone, 1% DMSO as drug carrier control, 100µM Curcumin, 50µM Rosmarinic acid, 0.25µM Gambogic acid, 0.25µM Celastrol, 200µM D-Penicillamine and 50µM Tetramethyl Nordihydroguaiaretic acid/TMNGA). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, as compared to β-actin loading control. Results demonstrate that select MAEs can reduce CTXLP expression in NCCIT cells.

[00136] Figure 55 depicts that MAE drug Celastrol abrogates CTXLP expression in human NCCIT teratocarcinoma cells. The ERVK-expressing NCCIT teratocarcinoma cell

line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 hours with increasing doses of celastrol (Cel, 0.1, 0.25, 1 and 2.5μM). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, as compared to β-actin loading control. Results demonstrate that Cel dose-dependently reduces CTXLP expression in NCCIT cells.

[00137] Figure 56 depicts that MAE drug Gambogic acid abrogates CTXLP expression in human NCCIT teratocarcinoma cells. The ERVK-expressing NCCIT teratocarcinoma cell line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 hours with increasing doses of gambogic acid (GA, 0.1, 0.25, 1 and 2.5μM). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, CaV2.2 C-terminal calcium channel-associated transcriptional regulator (CaV2.2 CCAT) and NOGO-A, as compared to β-actin loading control. Results demonstrate that GA dose-dependently reduces CTXLP and NOGO-A expression in NCCIT cells. CaV2.2 CCAT is inversely correlated with the expression of ERVK CTXLP in this culture system.

[00138] Figure 57 depicts that endogenous levels of CTXLP in human astrocytes can be depleted in the presence of gambogic acid. Human astrocytic cell line SVGA were treated with increasing doses of gambogic acid in the low micromolar range (0.25 and 0.5μM).

[00139] Figure 58 depicts that Gambogic acid blocks TNFα-induced CTXLP expression in human neurospheres. The ReNcell CX neuroprogenitor cell line was grown in suspension to produce human neurospheres (approximately 0.5mm diameter). Neurospheres were treated for 24 hours in neurobasal media alone, or with or without 1ng/ml TNFα and/or 0.5μM gambogic acid (GA). Cells were collected, protein extracted, and western blot performed to measure the expression of soluble (standard lysis) and insoluble (RIPA lysis) proteins for ERVK CTXLP, CaV2.2 C-terminal calcium channel-associated transcriptional regulator (CaV2.2 CCAT) and NOGO-A, as compared to β-actin loading control. Results demonstrate that TNFα enhances CTXLP and NOGO-A expression in neurons, whereas in the presence of GA this effect is blocked. CaV2.2 CCAT is inversely correlated with the expression of ERVK CTXLP in this neuronal culture system.

[00140] Figure 59 depicts alignments of orthologous and non-orthologous ERVK loci with at least one Toxin_18+ ORF. Examination of CTXLP encoding loci in three non-human primate genomes, Pan troglodytes (Common chimpanzee), Gorilla gorilla gorilla (Western

lowland gorilla), and Cercocebus atys (Sooty Mangabey), as well as humans reveals orthologous loci and conservation of the Toxin_18 cysteine motif (yellow). Orthology was determined by pairwise best BLAST matches of whole Retroexplorer retroelement entries and their flanking 1000bp, which mostly correspond to entire ERVs, but which sometimes were fragments. Three and four-way orthology was determined from pairwise orthology. Some loci do not have any species-specific sequence in the alignment, as the orthologous region in that species did not return a tBLASTx result. Non-orthologous sequences represent entries for which no orthologue was identified (possible paralogues or unique insertions). Representative sequences from each from each primate species highlights the degree of conservation in the Toxin_18 cysteine motif (yellow).

[00141] Figure 60 depicts different mutational patterns between orthologues and paralogues of ERVK env genes. Represented is a combined set of heatmaps generated by superheat from frames 0 (CTXLP) and 1 (Envelope) of the human and gorilla orthologues ORFs, which where both are positive for Toxin_18 positive (CTXLP). The sequences which are orthologues are indicated by black squares in the center of each space (paralogues do not have black squares). Blue is an ω (dN/dS ratio) less than 1, indicating purifying selection and similarity between the sequences. Yellow is an ω more than 1, indicating diversifying selection and dissimilarity between the sequences. Grey indicates that ω could not be computed (in all cases dS = 0, indicating no synonymous differences between the two sequences). Sequences which are identical along the diagonal are blacked out. The cytological bands are based on the human genome, with Gorilla designations indicating their respective human orthologue/paralogue coordinate. It is notable that orthologous sequences have a low ω when it is defined (when undefined it still has a low dN value). This suggests that there is more conservation between orthologues in different species than between paralogues from the same species. This pattern is much more apparent for CTXLP than for Env reading frame, where differences are smaller if present at all.

[00142] Figure 61 depicts that the transgene employed in the generation of ERVK envelope transgenic mice encodes CTXLP. The viral gene insert for the vector used in the generation of ERVK envelope transgenic mice8, was analysed for the potential to encode and produce ERVK CTXLP. The red annotation indicates the location of CTXLP in the transgene insert.

[00143] Figure 62 depicts an alignment of ERVK113 CTXLP sequence and the ERVK Consensus sequence found in the ERVK envelope transgenic mice. Note the similarity and retention of key cysteine motif.

[00144] Figure 63 depicts the ERVK Env and CTXLP proteins. Env is composed of the SU and TM subunits and is the prototypical gene product of the env gene. CTXLP is composed of the SU subunit and a C-terminal omega conotoxin domain and is encoded in an alternate open reading frame and may be produced due to ribosomal frameshifting.

[00145] Figure 64 depicts illustration of the disruption of voltage-gated calcium channel CaV2.2 by ERVK CTXLP. Both canonical endogenous retrovirus-K (ERVK) envelope protein and conotoxin-like protein (CTXLP) can be produced from the ERVK env gene. ERVK CTXLP disrupts CaV2.2 on multiple levels, by decreasing CaV2.2 channel expression, as well as depleting the CaV2.2 calcium channel-associated transcription regulator (CCAT) in the nucleus. CTXLP inhibition of voltage gated calcium ion channel CaV2.2 may preventing neurotransmitter release and the continuation of signal transduction in the postsynaptic neurons.

[00146] Figure 65 depicts the putative pathological implications of CTXLP expression in oligodendrocyte precursor cells. CTXLP expression is enhanced in the presence of pro-inflammatory cytokines, including TNFα. Increased CTXLP protein levels in ex vivo human spinal cord tissue coincides with an increase in oligodendrocyte precursor cell (OPC) markers, transcription factor 4 (TCF4) and Olig1, along with elevated neurite outgrowth inhibitor A (Nogo-A) levels in adjacent cells and decreased myelin associated glycoprotein (MAG) axonal/myelin expression, which all suggest oligodendrocyte (OL) pathology. Nogo-A is a regulator of OPC differentiation, inhibitor of OL myelination, and axonal growth cone collapse during axon regeneration in the CNS. MAG is a marker for differentiated OLs and highly expressed in myelinating OLs. Based on expression patterns of OPC and OL markers, this suggests that heightened CTXLP expression in ALS is associated with OPCs being arrested in an immature state and inhibition of OL myelination. Treatment with CTXLP inhibitors, including gambogic acid, may reduce inflammatory signalling and decrease OPC and OL pathology by inhibiting increased Nogo-A expression.

DETAILED DESCRIPTION

[00147] In one aspect, there is described herein the identification of a region in the ERVK provirus DNA which encodes a conotoxin-like polypeptide, and which may have

significance in ERVK pathogenesis. In a specific example, the polypeptide is CTXLP (CSDYGINCSHSYGCCSRSCIALFC) (SEQ ID NO: 1).

[00148] In one example, there is described an isolated polypeptide that comprises or consists of: an amino acid sequence having at least about 70% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 75% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 80% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 85% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 90% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 95% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 99% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 100% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example, the isolated polypeptide comprises or consists of an amino acid sequence having at least about 70% identity to about 100% identify with the amino acid sequence set forth in SEQ ID NO:1.

[00149] In one example, there is described an isolated nucleic acid molecule comprising a nucleotide sequence encoding a peptide comprising or consisting of an amino acid sequence having at least about 70% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 75% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 80% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 85% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 90% identity with the amino acid sequence set

forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 95% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 99% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 100% identity with the amino acid sequence set forth in SEQ ID NO:1. In another example the isolated nucleic acid encoding a peptide comprising or consisting of an amino acid sequence having at least about 70% to about 100% identity with the amino acid sequence set forth in SEQ ID NO:1.

[00150] The term “isolated”, as used herein, refers to altered or removed from the natural state. For example, a polypeptide or nucleic acid naturally present in a living animal is not “isolated,” but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is “isolated.” An isolated nucleic acid or polypeptide can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

[00151] Unless otherwise specified, a “nucleotide sequence encoding a polypeptide” (and the like) includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. The phrase nucleotide sequence that encodes a polypeptide protein or an RNA may also include introns to the extent that the nucleotide sequence encoding the protein may in some version contain an intron(s).

[00152] The terms “peptide,” “polypeptide,” and “protein”, as used herein are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide must contain at least two amino acids, and no limitation is placed on the maximum number of amino acids that can comprise a protein's or peptide's sequence. Polypeptides include any peptide or protein comprising two or more amino acids joined to each other by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. “Polypeptides” include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion

proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides, or a combination thereof.

[00153] In some examples, there is described a vector comprising the nucleic acid molecule described above and herein.

[00154] The term "vector" or "expression vector" as used herein refers to a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an in vitro expression system. Expression vectors include all those known in the art, such as cosmids, plasmids (e.g., naked or contained in liposomes) and viruses (e.g., lentiviruses, retroviruses, adenoviruses, and adeno-associated viruses) that incorporate the recombinant polynucleotide. In on example, the vector is a pcDNA3.1 vector.

[00155] The term "homologous" as used herein refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared×100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

[00156] "Similarity", for example between two peptides, may be determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to a sequence of a second polypeptide. Variants are defined to include peptide sequences different from the original sequence, for example, different from the original sequence in less than 40% of residues per segment of interest, different from the original sequence in less than 25% of residues per segment of interest, different by less than 10% of residues per segment of interest, or different from the original protein sequence in just a few residues per segment of interest and at the same time sufficiently homologous to the original sequence to preserve the functionality of the original sequence.

[00157] The term “sequence identity” of a polypeptide or polynucleotide as used herein refers to a degree of sameness in an amino acid residue or a base in a specific region of two sequences that are aligned to best match each other for comparison. The sequence identity is a value obtained via alignment and comparison of the two sequences in the specific region for comparison, in which a partial sequence in the specific region for comparison may be added or deleted with respect to a reference sequence. The sequence identity represented in a percentage may be calculated by, for example, comparing two sequences that are aligned to best match each other in the specific region for comparison, determining matched sites with the same amino acid or base in the two sequences to obtain the number of the matched sites, dividing the number of the matched sites in the two sequences by a total number of sites in the compared specific regions (i.e., a size of the compared region), and multiplying a result of the division by 100 to obtain a sequence identity as a percentage. The sequence identity as a percentage may be determined using a known sequence comparison program, for example, BLASTP or BLASTN (NCBI), CLC Main Workbench (CLC bio), or MegAlign™ (DNASTAR Inc).

[00158] A polypeptide of may be synthesized by conventional techniques. For example, the peptides may be synthesized by chemical synthesis using solid phase peptide synthesis. These methods employ either solid or solution phase synthesis methods. Automated synthesis may be used.

[00159] In some example, a polypeptide may be produced by culturing a cell comprising a nucleic acid which encoded the polypeptide, and isolating the polypeptide from the host cell or culture medium thereof.

[00160] The peptides of the invention can be post-translationally modified. For example, post-translational modifications that fall within the scope of the present invention include signal peptide cleavage, glycosylation, acetylation, isoprenylation, proteolysis, myristoylation, protein folding and proteolytic processing, etc. Some modifications or processing events require introduction of additional biological machinery. For example, processing events, such as signal peptide cleavage and core glycosylation, are examined by adding canine microsomal membranes or Xenopus egg extracts to a standard translation reaction.

[00161] In some examples, the polypeptides described herein may include unnatural amino acids formed by post-translational modification or by introducing unnatural amino

acids during translation. A variety of approaches are available for introducing unnatural amino acids during protein translation.

[00162] A “cell” or “host cell” refers to an individual cell or cell culture that can be or has been a recipient of any recombinant vector(s), isolated polynucleotide, or polypeptide. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology or in total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. A host cell includes cells transfected or infected in vivo or in vitro with a recombinant vector or a polynucleotide of the invention. A host cell which comprises a recombinant vector of the invention is a recombinant host cell.

[00163] In one example, the host cell is a cell obtained or derived from a subject.

[00164] The term “subject” or “patient” as used herein, refers to an animal, and can include, for example, domesticated animals, such as cats, dogs, etc., livestock (e.g., cattle, horses, pigs, sheep, goats, etc.), laboratory animals (e.g., mouse, rabbit, rat, guinea pig, etc.), mammals, non-human mammals, primates, non-human primates, rodents, birds, reptiles, amphibians, fish, and any other animal. The subject may be an infant, a child, an adult, or elderly. In a specific example, the subject is a human.

[00165] In one example the cell host is a human cell.

[00166] In one example, the cell is SVGA (astrocytes), RenCell CX (neuroprogenitor cells), or NCCIT (teratocarcinoma).

[00167] In some examples, there is described an antibody that specifically binds to a polypeptide as described herein. In one example, the polypeptide comprises or consists of the sequence of SEQ ID NO: 1.

[00168] The term “antibody” or “antibodies” is used herein refers to both polyclonal and monoclonal antibodies. In addition to intact or “full” immunoglobulin molecules, also included in the term “antibodies” are fragments (e.g., CDRs, Fv, Fab and Fc fragments) or polymers of those immunoglobulin molecules and humanized versions of immunoglobulin molecules, as long as they exhibit any of the desired properties according to the description.

[00169] Antibodies of the description may also be generated using well-known methods.

[00170] In some examples, a polypeptide may be used for generating an antibody of the description may be partially or fully purified from a natural source, or may be produced using recombinant DNA techniques.

[00171] In some examples, the antibodies may be purchased commercially.

[00172] In some examples, the generation of two or more different sets of monoclonal or polyclonal antibodies may maximize or increase the likelihood of obtaining an antibody with the specificity and affinity required for its intended use.

[00173] The antibodies produced may tested for their desired activity by known methods, in accordance with the purpose for which the antibodies are to be used (e.g., Immunoblooting, ELISA, immunohistochemistry, immunotherapy, etc).

[00174] For example, antibodies may be tested in ELISA assays or, Western blots, immunohistochemical staining of formalin-fixed or frozen tissue sections. After their initial in vitro characterization, antibodies intended for therapeutic or in vivo diagnostic use are tested according to known clinical testing methods.

[00175] The term “monoclonal antibody” as used herein refers to an antibody obtained from a substantially homogeneous population of antibodies, i.e.; the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. The monoclonal antibodies herein specifically include “chimeric” antibodies in which a portion of the heavy and/or light chain is identical with or homologous to corresponding sequences in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is identical with or homologous to corresponding sequences in antibodies derived from another species or belonging to another antibody class or subclass, as well as fragments of such antibodies, so long as they exhibit the desired antagonistic activity.

[00176] Monoclonal antibodies may be prepared using hybridoma methods. In a hybridoma method, a mouse or other appropriate host animal is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes may be immunized in vitro.

[00177] The monoclonal antibodies may also be made by recombinant DNA methods.

[00178] In some example, the antibodies are humanized antibodies. Methods for humanizing non-human antibodies are well known in the art.

[00179] In some examples, antibodies may be labeled with probes suitable for detection by various imaging methods. Methods for detection of probes include, but are not limited to, fluorescence, light, confocal and electron microscopy; magnetic resonance

imaging and spectroscopy; fluoroscopy, computed tomography and positron emission tomography.

[00180] Examples of probes may include, but are not limited to, fluorescein, rhodamine, eosin and other fluorophores, radioisotopes, gold, gadolinium and other lanthanides, paramagnetic iron, fluorine-18 and other positron-emitting radionuclides. Antibodies may be directly or indirectly labeled with said probes. Attachment of probes to the antibodies includes covalent attachment of the probe, incorporation of the probe into the antibody, and the covalent attachment of a chelating compound for binding of probe, amongst others well recognized in the art.

[00181] In one example, there is described a method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of an active agent or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity.

[00182] In one example, there is described a method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of an active agent or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity.

[00183] In one example, the active agent is a CTXLP inhibitors.

[00184] In one example, a CTXLP inhibitors inhibits or reduces the activity of CTXLP polypeptide.

[00185] In one example, a CTXLP inhibitors inhibits or reduces the level or amount of CTXLP polypeptide.

[00186] In one example, a CTXLP inhibitors inhibits or reduces the level or amount of of CTXLP mRNA.

[00187] In some example, a CTXLP inhibitor may be, without being limiting thereto, a small molecule, an antibody, a nucleic acid, an aptamer, a peptide.

[00188] The term “small molecule” as used herein refers to a molecule of less than about 1,000 daltons, in particular organic or inorganic compounds.

[00189] In one example, the small molecule may be a small molecule inhibitor of HIV Tat. In one example, the small molecule inhibitor of HIV Tat is a Michael acceptor electrophile (MAE). In one example, the MAE is curcumin, rosmarinic acid, gambogic acid,

celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine). In one example, the small molecule inhibitor of HIV Tat is a sulfhydryl compound with chelating properties. In one example, the sulfhydryl compound with chelating properties is N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides. In one example the small molecule inhibitor of HIV Tat is a Thioredoxin reductase 1 (TRR1) inhibitor. In one example, the Thioredoxin reductase 1 (TRR1) inhibitor is B5 (curcumin analog).

[00190] In one example, the CTXLP inhibitor is a nucleic acid molecule interfering specifically with CTXLP expression. In some example, the nucleic acid CTXLP inhibitor may be an antisense against CTXLP, a siRNA against CTXLP, a shRNA against CTXLP, or a ribozyme.

[00191] The term “RNAi” or “interfering RNA” refers an RNA, which is capable of down-regulating the expression of the targeted polypeptide, such as CTXLP. It encompasses small interfering RNA (siRNA), double-stranded RNA (dsRNA), single-stranded RNA (ssRNA), micro-RNA (miRNA), and short hairpin RNA (shRNA) molecules. RNA interference, designates a phenomenon by which dsRNA specifically suppresses expression of a target gene at post-translational level. In normal conditions, RNA interference is initiated by double-stranded RNA molecules (dsRNA) of several thousand base pairs in length. In vivo, dsRNA introduced into a cell is cleaved into a mixture of short dsRNA molecules called siRNA. The enzyme that catalyzes the cleavage, Dicer, is an endo-RNase that contains RNase III domains

[00192] siRNA are usually designed against a region 50-100 nucleotides downstream the translation initiator codon, whereas 5′UTR (untranslated region) and 3′UTR are usually avoided. The chosen siRNA target sequence should be subjected to a BLAST search against EST database to ensure that the only desired gene is targeted. Various products are commercially available to aid in the preparation and use of siRNA. In a preferred embodiment, the RNAi molecule is a siRNA of at least about 15-50 nucleotides in length, preferably about 20-30 base nucleotides.

[00193] RNAi can comprise naturally occurring RNA, synthetic RNA, or recombinantly produced RNA, as well as altered RNA that differs from naturally-occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations

can include addition of non-nucleotide material, such as to the end of the molecule or to one or more internal nucleotides of the RNAi, including modifications that make the RNAi resistant to nuclease digestion.

[00194] RNAi may be administered in free (naked) form or by the use of delivery systems that enhance stability and/or targeting, e.g., liposomes, or incorporated into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, bioadhesive microspheres, or proteinaceous vectors. They may also be administered in the form of their precursors or encoding DNAs.

[00195] Antisense nucleic acid can also be used to down-regulate the expression of CTXLP. The antisense nucleic acid can be complementary to all or part of a sense nucleic acid encoding CTXLP e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence, and is thought to interfere with the translation of the target mRNA.

[00196] An antisense nucleic acid can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. Particularly, antisense RNA can be chemically synthesized, produced by in vitro transcription from linear (e.g. PCR products) or circular templates (e.g., viral or non-viral vectors), or produced by in vivo transcription from viral or non-viral vectors.

[00197] Antisense nucleic acid may be modified to have enhanced stability, nuclease resistance, target specificity and improved pharmacological properties. For example, antisense nucleic acid may include modified nucleotides designed to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides.

[00198] Ribozyme molecules can also be used to block the expression of CTXLP. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. Ribozyme molecules specific for CTXLP can be designed, produced, and administered by methods commonly known to the art.

[00199] The term “aptamer” refers to a molecule of nucleic acid or a peptide able to bind specifically to CTXLP polypeptide.

[00200] The term “administering” as used herein includes all means of introducing the compounds and compositions described herein to the subject, including, but are not limited to, oral (po), intravenous (iv), intramuscular (im), subcutaneous (sc), transdermal, inhalation, and the like. The compounds and compositions described herein may be administered in unit dosage forms and/or formulations containing conventional nontoxic pharmaceutically-acceptable carriers, adjuvants, and vehicles.

[00201] Non limiting examples of oral administration include tablets, capsules, elixirs, syrups, and the like.

[00202] Non limiting examples of parenteral administration include intravenous, intraarterial, intraperitoneal, epidurial, intraurethral, intrasternal, intramuscular and subcutaneous, as well as any other art recognized route of parenteral administration.

[00203] Non limiting examples of means of parenteral administration include needle (including microneedle) injectors, needle-free injectors and infusion techniques, as well as any other means of parenteral administration recognized in the art. Parenteral formulations are typically aqueous solutions which may contain excipients such as salts, carbohydrates and buffering agents (preferably at a pH in the range from about 3 to about 9), but, for some applications, they may be more suitably formulated as a sterile non-aqueous solution or as a dried form to be used in conjunction with a suitable vehicle such as sterile, pyrogen-free water. The preparation of parenteral formulations under sterile conditions, for example, by lyophilization, may readily be accomplished using standard pharmaceutical techniques well known to those skilled in the art. Parenteral administration of a compound is illustratively performed in the form of saline solutions or with the compound incorporated into liposomes. In cases where the compound in itself is not sufficiently soluble to be dissolved, a solubilizer such as ethanol can be applied.

[00204] The dosage of each compound(s) depends on several factors, including: the administration method, the condition to be treated, the severity of the condition, whether the condition is to be treated or prevented, and the age, weight, and health of the person to be treated. Additionally, pharmacogenomic (the effect of genotype on the pharmacokinetic, pharmacodynamic or efficacy profile of a therapeutic) information about a particular patient may affect the dosage used.

[00205] In one example, examples of ERVK associated diseases may include but are not limited to infectious diseases, autoimmune diseases, neurological diseases, cancer, and other conditions such as idiopathic nephrotic syndrome.

[00206] In one example, examples of CTXLP associated diseases may include but are not limited to infectious diseases, autoimmune diseases, neurological diseases, cancer, and other conditions such as idiopathic nephrotic syndrome.

[00207] Infectious disease, including but not limited to HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

[00208] Autoimmune disease, including but not limited to, Insulin dependent diabetes mellitus, morphea, Psoriasis, rheumathoid arthritis, systemic lupus erythematosus, Type 2 diabetes mellitus (in people with SCZ).

[00209] Neurological disease, including but not limited to, amyotrophic lateral sclerosis, bipolar disorder, Kennedy's disease, multiple sclerosis, Schizophrenia

[00210] Cancer, including but not limited to, Breast cancer, Chronic myelogenous leukemia, Colon cancer, Gastric cancer, Germ cell tumours, Germinogenic tongue tumours, Gonadoblastomas, Hepatocellular carcinoma, Leukemia, Lymphoma, Melanoma, Myelodysplastic syndrome, Nasopharyngeal carcinoma, Ovarian cancer, Pancreatic cancer, Prostate cancer, Trophoblastic tumours, Tumorigenesis (via AR interaction), Thyroid adenoma, ERVK in cancerous tissues.

[00211] Other, including but not limited to, Idiopathic nephrotic syndrome.

[00212] In one example, there is described a method for transcriptional activation, comprising contacting a DNA molecule comprising a gene with a polypeptide as described herein.

[00213] In one example, there is described a diagnostic reagent for use in the detection of CTXLP polypeptide in a subject, comprising an antibody specific for CTXLP polypeptide.

[00214] In one example, there is described a diagnostic reagent for use in the detection of CTXLP mRNA in a subject, comprising an isolated nucleic acid specific for CTXLP.

[00215] In one example, there is described a diagnostic reagent for use in the detection CTXLP activity in a subject, comprising a polypeptide as described herein.

[00216] In one example, there is described a method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent or a pharmaceutically acceptable salt thereof when the amount of CTXLP polypeptide, or CTXLP

activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity.

[00217] In one example, there is described a method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent or a pharmaceutically acceptable salt thereof when the amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity.

[00218] Method are conveniently practiced by providing the compounds and/or compositions used in such method in the form of a kit. Such kit preferably contains the composition. Such a kit preferably contains instructions for the use thereof.

[00219] In one example, there is described a kit comprising:(a) a container comprising a pharmaceutical composition containing a polypeptide as described herein, and/or a nucleic acid as described herein, and/or an expression vector, and/or a host cell, and/or an antibody as described herein, in solution or in lyophilized form;(b) optionally, a second container containing a diluent or reconstituting solution for the lyophilized formulation; and(c) optionally, instructions for use.

[00220] In one example, the kit further comprising one or more of (iii) a buffer, (iv) a diluent, (v) a filter, (vi) a needle, or (v) a syringe.

[00221] To gain a better understanding of the invention described herein, the following examples are set forth. It should be understood that these examples are for illustrative purposes only. Therefore, they should not limit the scope of this invention in anyway.

[00222] **EXAMPLES**

[00223] Endogenous retroviruses (ERVs) are host genetic elements originating from prior infection of host germ-line cells that are subsequently inherited through the germline. ERVs represent approximately 8% of human genomic DNA. ERVs can benefit their host, or in other contexts are proposed to be involved in pathogenesis and disease. Notably, our interest in ERVK CTXLP lies in its association to motor neuron conditions such as Amyotrophic Lateral Sclerosis (ALS), as well in cancers.

[00224] **ERVK CTXLP Bioinformatics:** Endogenous retrovirus-K (ERVK) conotoxin-like protein (CTXLP) is produced following a ribosomal frameshifting event and is subject to post-translational modifications (PTMs). PTMs and alternative start sites allow for a variety of

CTXLP isoforms which may drive distinct pathogenic mechanisms. The prevalence and polymorphic variability of ERVK CTXLP-encoding insertions suggests that CTXLP is a pervasive and conserved ERVK protein. The molecular characterization of CTXLP revealed a conotoxin domain which predicts that it acts as antagonist to specific voltage-gated calcium channels. CTXLP also contains a cysteine motif that aligned to multiple cone snail, spider and viral toxins, which are known to function as antagonists to voltage-gated ion channels. This intrinsic capacity to interfere with calcium channels through these motifs suggests a putative mechanism by which ERVK can act in the pathogenesis of motor neuron diseases such as ALS.

[00225] **CTXLP biological characterization:** CTXLP protein isoform expression in NCCIT and SVGA cells was elucidated by Western blots which indicated presumed isoform sizes of 32 kDa, 51 kDa, and 90/110 kDa. In NCCIT cells, endogenous CTXLP is ubiquitously expressed in the nucleus, and also identified in the cytoplasm and cell membrane, based on cell fractionation and confocal experiments. In contrast, in SVGA cells basal CTXLP levels are limited, but highly inducible by pro-inflammatory stimuli. In addition, CTXP expression is almost exclusively in the chromatin fraction and demonstrates a prominence in the nucleus upon confocal imaging. The notable exception is that after pro-inflammatory activation for 24 hours CTXLP puncta appear in the cytoplasm and on cellular membranes reminiscent of pathogenic protein aggregates. Moreover, the localization pattern in response to pro-inflammatory activators resulting in a prominence in the nucleus ability to bind chromatin suggests that CTXLP may be involved in viral transcription. A primary candidate as a viral transcription factor is the 32 kDa CTXLP isoform, as small cysteine-rich proteins have previously been identified as transcriptional activators, as per HIV-1 Tat (15 kDa) and HTLV Tax (40 kDa) role as viral transcription co-activators.

[00226] **CTXLP Expression in disease states:** ERVK CTXLP localized to the motor cortex in spinal cord sections from autopsy samples of patients with ALS, but not neuro-normal controls. Concomitantly, CTXLP expression was substantially enhanced in diseased ALS tissues, aligning with oligodendrocytes, Nogo-A expression and demyelinated lesions. In addition, cancer cell lines and tissue expressed greater levels of CTXLP relative to normal controls. Together, these findings provide significant evidence for the activity of CTXLP in ALS and certain cancers.

[00227] **Pathological consequences of CTXLP expression**: ERVK CTXLP has the capacity to enhance NF-κB p65 and p50 proteins that play a critical role in ALS pathogenesis.

In addition, CTXLP administration or transfection induced significant levels of capase-3. The induction of caspase-3 activation and apoptosis by CTXLP was inhibited by excess extracellular calcium pointing to a calcium channel mediated activation of toxicity. Remarkably, despite the initial die off of cells, cells remaining in the cultures appeared to demonstrate appreciable cellular proliferation relative to control suggesting the induction of a carcinogenic process. CTXLP also had a notable effect on the depletion of CaV2.2 voltage-gated calcium channel-associated transcriptional regulator (CaV2.2 CCAT) from the nucleus.

[00228] **ERVK CTXLP can be targeted by small molecule therapeutics:** A drug screen revealed that celastrol and gambogic acid have the capacity to inhibit endogenous CTXLP expression in NCCIT cancer cell line. Moreover, gambogic acid was able to reduce inducible CTXLP expression the presence of TNFα and ameliorate the concomitant expression of pathogenic marker Nogo-A. This strongly suggests that therapeutic targeting of CTXLP in human disease could be an agent in the efforts to ameliorate the devastation of ALS.

[00229] **Development of cell and animal models to investigate CTXLP pathogenesis:** Human tissue and animal models for the study of CTXLP in ALS and cancer are needed. We are actively working to further develop our human tissue culture models. In addition, together with Dr. Alberto Civetta, we are in the process of developing a model in *Drosophilia* at the University of Winnipeg. Importantly, we will continue to pursue mammalian models with our collaborators which offer an opportunity to explore multiple features of pathogenesis as we continue to elucidate the processes involved in CTXLP pathogenesis.

[00230] **ERVK CTXLP is a novel pathological target for the development of therapeutics for inflammatory, neurological and oncogenic diseases.**

ABBREVIATIONS

[00231] Abbreviations used in text.

AGRP	Agouti-related peptide
ALS	Amyotrophic lateral sclerosis
ASIP	Agouti-signalling protein
BA6	Brodmann area 6
BLAST	Basic local alignment search tool
BMAA	Beta-N-methylamino-L-alanine
CC	Cervical spinal cord
CCAT	Calcium channel-associated transcription regulator
cDNA	Complimentary deoxyribonucleic acid
CEL	Celastrol
ChIP	Chromatin immunoprecipitation

CNS	Central nervous system
CTXLP	Conotoxin-like protein
CX3CL1	Chemokine (C-X3-C motif) ligand 1
DAPI	4′,6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
Env	Envelope
ERV	Endogenous retrovirus
ERVH	Endogenous retrovirus-H
ERVK	Endogenous retrovirus-K
ERVW	Endogenous retrovirus-W
GA	Gambogic acid
HAART	Highly active antiretroviral therapy
HAUSP/USP7	Herpesvirus-associated ubiquitin-specific protease/Ubiquitin-specific-processing protease 7
HCV	Hepatitis C virus
HIV	Human Immunodeficiency virus
HTLV	Human T-lymphotrophic virus
HML	Human Mouse mammary tumour virus-like
ICK	Inhibitor cysteine knot
IP	Immunoprecipitation
IRES	Internal ribosomal entry site
IRF7	Interferon regulatory factor 7
ISRE	Interferon response element
LATS	Large tumor suppressor kinase
LC	Lumbar spinal cord
LIGHT	Homologous to lymphotoxin, exhibits inducible expression and competes with HSV glycoprotein D for binding to herpesvirus entry mediator, a receptor expressed on T lymphocytes
LTR	Long terminal repeat
MAE	Michael acceptor electrophile
MAG	Myelin-associated glycoprotein
MAP2	Microtubule-associated protein 2
MAPK	Mitogen-activated protein kinases
MC1R	Melanocortin receptor 1
MMTV	Mouse mammary tumour virus
MOG	Myelin oligodendrocyte glycoprotein
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis
MUSCLE	MUltiple Sequence Comparison by Log-Expectation
NCCIT	National Cancer Center Institute Tokyo, teratocarcinoma cell line
NF-κB	Nuclear factor κB
NCBI	National Centre for Biotechnology Information
NEC-1/2	Necrostatin-1/2
NgR1	Nogo-A receptor
NN	Neuronormal
NLS	Nuclear localization signal
NPV	Nuclear polyhedrosis virus
Olig1/2	Oligodendrocyte transcription factor 1/2
OPC	Oligodendrocyte precursor cell
ORF	Open reading frame

PCA	Principle component analysis
PLP	Proteolipid protein
PRF	Programmed ribosomal frameshift
PTM	Post-translational modification
Q-PCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RelA	REL-associated protein
RNA	Ribonucleic acid
RT	Reverse transcriptase
RTN4R	Reticulon 4 receptor
SC	Solochrome cyanine
SRA	Sequence Read Archive
SU	Surface unit
SVGA	SV40 T antigen glial astrocytes
Tat	Trans-activator of transcription
Tax	Transactivator from the X-gene region
TCF4	Transcription factor 4
TDP-43	TAR DNA-binding protein 43
TM	Transmembrane
TMEV	Theiler's Murine Encephalomyelitis Virus
TNFα	Tumour necrosis factor α
TRAF-2/6	TNF receptor associated factor
VEGF	Vascular endothelial growth factors
VGCC	Voltage gated calcium channel
WCE	Whole cell extract

[00232] **Endogenous retroviruses**

[00233] Retroviruses are single-stranded RNA viruses that replicate through reverse transcription[1]. Retroviruses use the enzyme reverse transcriptase to convert their genomic RNA to DNA, and then use a viral integrase to insert itself into a host genome[2]. Retroviruses are categorized as being either exogenous or endogenous[3]. Examples of exogenous retroviruses include Human-Immunodeficiency virus (HIV) and Human T-lymphotropic virus (HTLV). Alternatively, endogenous retroviruses (ERVs) are genetic elements originating from prior infection of host germ-line cells, allowing them to be inherited through Mendelian genetics[3]. ERVs represent approximately 8% of human genomic DNA[4]. ERVs can benefit their host[5], or in other contexts are proposed to be involved in pathogenesis and disease[6].

[00234] Endogenous Retrovirus-K (ERVK) is the most recently endogenated retrovirus in the human genome[1]. ERVK is a group of similar viruses that are categorized into 10 clades (sub-groups). ERVK (HML-2 clade) first entered the human genome approximately 28 million years ago, occurring before the divergence of hominids and old-world monkeys[7]. More recent insertions of ERVK occurred up to 200,000 years ago, and are specific to the

human lineage. This has resulted in several human-specific ERVK insertions[8]. Approximately 1000 ERVK loci have been identified in the human genome[9]. Although the majority of ERVK insertions have been silenced through mutations and negative selection, there are an estimated 24 fixed loci capable of producing viral proteins[3,6]. ERVs are also found to be highly polymorphic between individuals and different ethnic groups[7]. ERVK expression has been detected in several tissues throughout the body at varying levels between individuals[3,10].

[00235] **ERVK genome**

[00236] The ERVK genome consists of the essential retroviral genes gag-pro-pol-env, along with its own accessory genes[1] (Figure 1). The group specific antigen (gag) gene encodes structural proteins including the viral capsid[8,11]. The protease (pro) gene encodes a protease which cleaves newly synthesized viral proteins[1]. The polymerase (pol) gene encodes for proteins including reverse transcriptase and integrase[2,11]. The envelope (env) gene encodes the glycoproteins of the viral envelope[11]. The ERVK genome is flanked by long terminal repeats (LTRs), which were assistive in retroviral DNA insertion into the host[12]. Once inserted into the host genome, the virus is considered a provirus. LTRs contain elements of enhancers and promoters, including transcription factor binding-sites and interferon-stimulated response elements that regulate both retroviral and host gene expression[6,11].

[00237] ERVK can be organized into two types based on their genome. Type 1 proviruses contain a 292 base-pair deletion near the 5' end of env not found in type 2 proviruses[13]. The presence or absence of this deletion affects the accessory proteins the provirus produces[13,14].

[00238] **ERVK envelope protein**

[00239] The ERVK envelope (Env) protein is initially translated as a large, inactive polyprotein[15,16]. The polyproteins dimerizes or trimerizes and are then cleaved by the cellular protease furin, forming a surface unit (SU) and transmembrane (TM) subunit[15]. Like other retroviral envelope proteins, the assembled Env trimer is heavily glycosylated and is expressed on the viral capsid membrane, as well as infected host cell membranes, allowing for incorporation of the virus into host cells[15,17].

[00240] **Cysteine knot proteins**

[00241] Cysteine knots are protein structural motifs found throughout animals, fungi and plants[18]. Cysteine knot proteins are known for their stability, attributed to their 3 disulfide bonds; two of the disulfide bonds and their peptide backbone form a ring that the third bond

goes through, thus forming a "knot" structure[18]. Cysteine knot proteins are categorized as cyclic cysteine knots, growth factor cysteine knots, or inhibitor cysteine knots (ICK). Cyclic cysteine knots are found in plants and often have defense functions as bactericides and insecticides[18]. Growth factor cysteine knots are found in extracellular signaling molecules and are involved in various functions including cell-cell communication and embryonic development[19]. Examples include the vascular endothelial growth factors (VEGFs), and nerve growth factor[19]. ICK proteins are found in fungi, plants and animals and act as antagonists to a variety of receptors and ion channels[18].

[00242] ICK proteins include a vast array of peptides found in various living organisms. The ICK structure consists of six conserved (connected as Cys1-CysIV, CysII-CysV, and CysIII-CysVI) cysteine residues and an otherwise variable peptide backbone[18] (Figure 3). Within the animal kingdom, ICK peptides are found in the venoms of spiders, scorpions, and marine snails, and function either as pore-blockers or gate-modifiers of ion channels[18]. Mammalian ICK peptides have also been identified, including agouti-signalling protein (ASIP) and agouti-related peptide (AGRP)[20].

[00243] Animal ICKs are proposed to be a result of divergent evolution[21]. Functional constraints during evolution have resulted in spider, snail, and scorpion ICKs maintaining a similar gene structure, protein fold, and target receptor, which are all evidence for a common ancestor[21]. Alternatively, plant and fungi ICK do not have these similarities to animal ICKs, suggesting they are a product of convergent evolution[21]. In certain baculoviruses, a cysteine-rich ORF has been detected, that potentially translates into an ICK fold[21]. The putative ICK motif resembles the animal ICKs, suggesting that viruses may have obtained this genetic sequence by a gene transfer event after infecting an ICK-carrying host[21].

[00244] **Conotoxins**

[00245] Conotoxins are neurotoxic peptides found in the Conus genus of marine snails used to immobilize prey[22]. Conus species are distinct in their ability to produce hundreds of different toxic peptides[23]. Conotoxins are disulfide-rich and are usually 10-30 amino acids in length[22]. Conotoxins act as antagonists to specific voltage and ligand-gated ion channels[22]. In humans, symptoms of conotoxin exposure include poor coordination, blurred vision, speech difficulties, and nausea[23]. Conotoxins have also been associated with episodes of delirium and psychosis[24].

[00246] The O-superfamily of conotoxins exhibits an ICK fold. Members of the O-superfamily include μ-conotoxins, which inhibit voltage-gated sodium channels, and δ-

conotoxins, which delay sodium channel inactivation[25]. K-Conotoxins are inhibitors of voltage-gated potassium channels; ω -conotoxins inhibit N-type voltage-gated calcium channels (VGCCs)[25]. N-type VGCCs are located in presynaptic nerve terminals and are involved in neurotransmitter release[26]. ω -Conotoxin's selectivity for N-type VGCCs has allowed for their development as therapeutic agents. The ω -conotoxin MVIIA has been developed into a drug for relief of chronic and inflammatory pain[27].

[00247] Genes encoding an ω-conotoxin-like protein (CTXLP) have also been identified in certain viruses. Nuclear polyhedrosis viruses (NPV) have been shown to secrete a small conotoxin-like peptide[28]. NPVs are insect pathogens belonging to the family baculoviridae[28]. Although NPV-CTXLP's function has not been elucidated, its cysteine bridges were found to have a nearly identical structure to the conserved w-conotoxin's cysteine motif[28]. We have discovered a novel CTXLP ORF in the envelope gene of ERVK. The full pathogenic potential of ERVK CTXLP domain remains unknown.

[00248] ## ERVK CTXLP Bioinformatics

[00249] **Identification of a conotoxin-like domain in the ERVK genome**

[00250] *Splicing and Conserved Domains in the ERVK Genome (Start Codon-Biased Analysis)*

[00251] NetGene2 splice site prediction yielded a large number of predicted splice junctions (105-119 per ERVK sequence). However, after exhaustive analysis, none of these splice junctions resulted in the creation of domains that could be identified using the Conserved Domains Database. However, the predicted splicing patterns resulted in the identification of between 27 and 46 newly created ORFs per ERVK sequence.

[00252] *Conserved Domains (Start Codon-Unbiased Analysis)*

[00253] After finding no conserved domains in the initial analysis, the requirement for a start codon (ATG, CTG, TTG, GTG or ATT) at the beginning of each ORF was removed, because a start codon could be introduced through splicing and thus was not strictly necessary. The removal of this requirement resulted in slightly different ORFs, which can be seen in Figure 2. Analysis of these ORFs identified a previously undescribed region that would generate a peptide with significant homology to known proteins. This ORF occurred in both type 1 and type 2 genomes (that is, it was not affected by the 292-base pair deletion). DNA with the potential to encode a peptide containing a domain with homology to the O-conotoxin superfamily was identified in a region of env, but in a different reading frame, from

nucleotide 7863 to nucleotide 7934 in the 5'-3' direction. This ORF did not contain the typical methionine codon that is often used as a start codon for translation.

[00254] *Conotoxin-like Domain*

[00255] The putative conotoxin-like domain contained six characteristic cysteine residues and one characteristic glycine residue, indicating that it is most similar to the ω-conotoxin family. Another group of viruses, Nuclear Polyhedrosis Viruses, which are insect-infecting Baculoviruses, produce a similar conotoxin-like protein (NPV CTXLP). The putative ERVK CTXLP showed the greatest similarity to these viral proteins. Figure 3 shows the sequences of several ω-conotoxins produced by 3 cone snail species, as well as the sequence of a Nuclear Polyhedrosis Virus conotoxin-like domain. Although these sequences differ from each other in notable ways, 7 conserved residues (6 cysteines and 1 glycine) are found in all of them. These residues are also observed in the ERVK CTXLP domain.

[00256] The ERVK CTXLP sequence showed the greatest homology to NPV CTXLP sequences (E-value = 1.09×10^{-5}). Figure 4 shows the sequence logo for ERVK CTXLP and 10 NPV sequences, in which several more amino-acid residues (in addition to the 7 described above) are conserved.

[00257] Figure 5 summarizes the Conus and NPV sequences with greatest similarity and centrality to ERVK CTXLP. Figure 6 shows the logo of the knotin domain of CTXLP and its amino acid composition.

[00258] *Three-Dimensional Modeling of the ERVK Conotoxin-like Protein*

[00259] Conotoxins adopt a knot-like conformation, called a knottin domain, which is important for their action. Omega-conotoxin and NPV CTXLP knottins include 3 disulfide bonds. Tertiary structure prediction of the ERVK-113 CTXLP protein using Knotter 1D3D software resulted in the conclusion that it too could form these characteristic features. The predicted 3-dimensional structure of the ERVK-113 CTXLP domain is shown in Figure 7.

[00260] This predicted structure was then superimposed on the predicted structure of an NPV CTXLP domain to examine the similarity between the two. This structure alignment (Figure 8) is based on sequence alignment and was prepared using UCSF Chimera software[31].

[00261] The root mean square deviation between 24 atom pairs in this alignment is 0.426 angstroms. However, it can be difficult to see how similar the predicted structure of these two protein domains are from this image (Figure 8). As such, the structures were reduced to their respective peptide backbones. The resultant image can be seen in Figure 9.

[00262] *Conotoxin-Like Proteins are Not Encoded by Other Retroviruses*

[00263] After identifying that these two distantly related groups of viruses (ERVK and NPVs) both contain conotoxin-like protein coding capacity, we also searched for conotoxin-like domains within translations of all three reading frames of the env region of several other retroviral genomes (HIV-1, HTLV-1, MMTV, ERVW, ERVH). No conotoxin-like domains were identified in any of these retroviruses from our analysis.

[00264] *Alignments of ERVK CTXLP and other cysteine-rich proteins*

[00265] The ERVK CTXLP ORF is 39 amino acids long, with the cysteine-rich motif accounting for 30/39 amino acids (CSDYGINCSHSYGCCSRSCIALFCSVSKLC). The CTXLP cysteine-rich sequence was aligned to inhibitor cysteine knot (ICK) proteins and other cysteine-rich proteins using Geneious software (Version R8)[30]. A sequence logo was generated from the alignment to assess amino acid conservation between CTXLP and known cysteine-rich proteins (Figure 10). Geneious alignment software was used to compare ERVK CTXLP's cysteine-rich amino acid sequence to other cysteine-rich proteins (Table 1) to further understand CTXLP's structure and potential function.

Table 1: Proteins and peptides from various organisms and their respective accession numbers compared to CTXLP cysteine-rich motif found using Geneious software.

Protein/Peptide	Organism	Accession #
Guanxitoxin-2	Spider	P84837.1
Guanxitoxin-1D	Spider	P84836.1
Hainantoxin-I	Spider	D2Y1X6.1
Hainantoxin-III	Spider	D2Y1X9.1
Hainantoxin-IV	Spider	1NIY_A
Hainantoxin-V	Spider	P60975.1
Hanatoxin-1	Spider	P56852.1
Hanatoxin-2	Spider	P56853.1
Sgtx	Spider	1LA4_A
Grammotoxin	Spider	P60590.2
Huwentoxin-I	Spider	P56676.2
Huwentoxin-X	Spider	P68424.2
agouti-related peptide	Human	O00253.1
agouti-signalling protein	Human	1Y7K_A
VEGF-A	Human	P15692.2
VEGF-B	Human	P49765.2
VEGF-C	Human	CAA63907.1
VEGF-D	Human	BAA24264.1
VEGF-E	Human	ABA00650.1
VEGF-F	Snake	1WQ8_A
Placental Growth Factor	Human	AAH07789.1
Tat	HIV-1	CCD30501.1
Tat	HIV-2	AAA76845.1
Tax	HTLV-1	BAD95659.1
Tax	HTLV-2	AFC76143.1
Tax	HTLV-3	Q0R5R1.1
Envelope	HTLV-4	CAA29690.1
Envelope	Jaagsiekte Sheep Retrovirus	AAK38688.1

[00266] Several spiders are known to utilize ICK peptides in their toxins[32]. The putative ERVK CTXLP cysteine motif was aligned to 12 spider toxins from various species of spiders. **Figure 11** shows ERVK CTXLP peptide had significant similarity to the Hainantoxin-I and Hainantoxin-IV ICK motifs, with a pairwise identity above 25% suggesting conserved protein function (NCBI).

[00267] The sequence logo generated showed that the cysteine knot (C-C-CC-C-C) motif is conserved in ERVK CTXLP and the spider toxins examined. Although there was significant sequence diversity in other amino acids, each sequence contained the essential 6

cysteine residues for an ICK. Five of the 12 spider toxins also contained a glycine residue in an identical position to ERVK CTXLP's characteristic glycine. The overall cysteine spacing of the CTXLP motif was unique when compared with spider toxins, suggesting that despite forming an ICK fold, the overall protein conformations are likely divergent. This could explain receptor binding specificity of each toxin species. Spider toxins and CTXLP were then examined for conserved motifs (**Table 2**). Overall, there was little conservation outside of the ICK motif, with the most significant conservation found in Hainantoxin-I, a voltage gated sodium channel inhibitor.

Table 2: Spider toxins, host species and target receptors show similarity and conserved motifs with ERVK CTXLP.

Toxin	Species	Target Receptors	Identity with CTXLP (%)	Conserved Amino Acid Motifs
Grammotoxin	*Grammostola spatulata*	P/Q, N-type VGCC	16.7	G (56), S (59), SK(73)
Huwentoxin-I	*Selenocosmia huwena*	Presynaptic N-type VGCC	12.5	D (56), K (76)
Huwentoxin-X	*Selenocosmia huwena*	N-type VGCC (Dorsal Root Ganglion)	10.3	G (55), K (77)
Hainantoxin-I	*Selenocosmia hainana*	Voltage Gated Sodium Channel	26.5	G (55), K (78)
Hainantoxin-III	*Selenocosmia hainana*	Neuronal tetrodotoxin-sensitive Voltage Gated Sodium Channel	23.5	G (55), K (78), SK (75)
Hainantoxin-IV	*Selenocosmia hainana*	Voltage Gated Sodium Channel	25.8	G (55), S (61), S (68)
Hainantoxin-V	*Selenocosmia hainana*	Voltage Gated Sodium Channel	23.5	G (55), S (61)
Guanxitoxin-I	*Plesiophrictus guangxiensis*	Voltage Gated Potassium Channel (Kv2.1 subtype)	17.1	None
Guanxitoxin-II	*Plesiophrictus guangxiensis*	Voltage Gated Potassium Channels	17.1	None
Hanatoxin-I	*Grammostola rosea*	Voltage Gated Potassium Channel (Kv2.1 subtype)	16.7	K (78)
Hanatoxin-II	Grammostola rosea	Voltage Gated Potassium Channels	16.7	K (78)
Sgtx-I	*Scodra griseipes*	Voltage Gated Potassium Channel (Kv2.1 subtype)	13.9	K (78)

[00268] When ERVK CTXLP peptide was aligned to the human ICK proteins agouti-signalling protein (ASIP) and agouti-related peptide (AGRP), significant similarity was found in the conserved cysteine domain (both with a pairwise identity of 21.9% with CTXLP), suggesting structural similarity and possible functional overlap (Figure 12).

[00269] Seven of ten cysteine residues found in the agouti family peptides aligned with ERVK CTXLP. Agouti proteins use 8 cysteines to form an ICK structure[33], whereas CTXLP only has 7 cysteines and is likely to take on a simpler ICK fold. Agouti-like proteins are the only known ICK domain containing protein in humans; however, these findings suggest that CTXLP may also be a human-derived ICK protein.

[00270] ERVK CTXLP was also aligned to 7 VEGF proteins, which utilize a growth factor cysteine knot. Although there was some alignment between the cysteine residues and some similar motifs (ex. GCC) identified, the large gaps in spacing and the different spacing of cysteine residues in the VEGF proteins suggests that there is no significant similarity to ERVK CTXLP (all with identity ≤ 7.7 %) (Figure 13). The dissimilarity suggests that CTXLP does not take on a growth factor cysteine knot structure.

[00271] When ERVK CTXLP was aligned to the retroviral accessory protein Tat from HIV-1 and HIV-2, some degree of similarity was detected (identity 19.4% and 16.1%, respectively) (Figure 14). Tat (HIV-1 and HIV-2) contains a conserved C-C-CC-C-C motif that has a much tighter spacing than ERVK CTXLP. In contrast to HIV-1 Tat and CTXLP, HIV-2 Tat did not contain a central “CC” pair of cysteine residues. ERVK CTXLP was also aligned with Tax proteins from HTLV-1, HTLV-2, and HTLV-3 along with envelope from HTLV-4 and envelope from Jaagsiekte retrovirus, and no homology was detected (data not shown). The conservation of a cysteine motif and central “CC” cysteine pair in HIV-1 Tat and CTXLP is a potential basis for conserved structure and function of these viral proteins.

[00272] Aligning ERVK CTXLP to several cysteine-rich peptides provided insight into the potential function of the CTXLP protein domain. ERVK CTXLP showed the greatest similarity to ICK peptides. The cysteine knot motif was conserved in all of the spider toxins examined, and Hainaintoxin-I showed the greatest similarity to CTXLP with an identity of 26.5%, suggesting similarity in function (NCBI). All other amino acid residues were highly variable, suggesting that the conservation of the cysteine residues and the tertiary structure are more important for peptide function rather than the primary amino acid sequence. The spider toxins function as antagonists to voltage gated ion channels, suggesting CTXLP may have a similar function[18]. Hainantoxin-I is a voltage-gated sodium channel inhibitor[34]; thus, CTXLP may function as a voltage-gated sodium channel inhibitor. Although the ERVK CTXLP had significant similarity to Hainantoxin-I, ERVK CTXLP still had the greatest similarity (25.9-33.3%) to the cone snail ω-conotoxins, suggesting CTXLP functions as a VGCC inhibitor. Previous studies have also shown that ω-conotoxin’s amino acid residues threonine 11,

tyrosine 13, lysine 2, lysine 4, and arginine 22 are important for calcium channel receptor binding[35]. CTXLP has some similar conserved residues including a tyrosine in position 12 and an arginine in position 17. CTXLP may alternatively utilize different amino acid residues to bind to cognate VGCC targets.

[00273] ERVK CTXLP also showed significant identity to the human agouti-family proteins, specifically ASIP and AGRP peptides. ASIP and AGRP are mammalian ICK peptides that both function as antagonists to melanocortin receptors 1, 3 and 4 (MC1R, MC3R, and MC4R)[36]. ASIP is produced in the skin to promote pigment production, while AGRP is involved in metabolism[36]. The agouti-family of peptides contains a unique ICK pattern[33]. CTXLP is only capable of forming 3 of the 4 cysteine bridges identified in agouti, suggesting that CTXLP takes on the basic ICK fold. The similarity between ERVK CTXLP to the agouti family of peptides provides further support for CTXLP's structure as an ICK peptide, along with first evidence for the presence of viral ICK peptides in humans.

[00274] Vascular endothelial growth factors (VEGFs) contain a growth factor cysteine knot motif, and are signalling molecules involved in angiogenesis[19]. ERVK CTXLP did not show significant similarity (≤ 7.7%) to the VEGF proteins. A dissimilar cysteine motif with a different spacing of cysteine residues and a significant difference in overall protein size (12-47 kDa for VEGF versus 32 and 51 kDa for CTXLP), suggests that ERVK CTXLP does not function in a growth factor or cytokine manner[19]. CTXLP's similarity to the ICK peptides and dissimilarity to VEGF suggests that ERVK CTXLP likely functions as a receptor antagonist via an ICK motif.

[00275] Cysteine-rich peptides have also been identified in exogenous retroviruses. ERVK CTXLP has some sequence similarity with the Tat accessory protein of HIV-1 and HIV-2. Although Tat is not an ICK peptide and has a slightly different cysteine spacing pattern to ERVK CTXLP, a similar cysteine rich motif was identified in both proteins (19.6%). The cysteine-rich motif of Tat endows this protein with neurotoxic properties[37]. Tat expression in the brains of HIV-1 infected patients has been associated with neuronal apoptosis via caspase activation and calcium accumulation[38]. The structural similarities between Tat and ERVK CTXLP may suggest that they both use similar mechanisms for pathogenicity. HIV-2 is known to be a less pathogenic than HIV-1[39]. A partial explanation to this decreased pathogenicity may lie in the structural differences between their respective Tat proteins. HIV-2 Tat has a deletion of one cysteine, losing the "CC" motif. Interestingly, all ERVK CTXLP domains examined contained a "CC" motif. The mechanisms surrounding HIV Tat neurotoxicity are diverse and manifold[38,40,41], suggesting that substantial research may be required to address potential

ERVK CTXLP cellular toxicity in the CNS. The cysteine motif in HIV Tat has also been associated with increased HIV transactivation, by translocating to the nucleus and interacting with transcriptional machinery[38]. HIV Tat can also transactivate ERVK[42]. Thus, the multiple functions of HIV-1 and HIV-2 Tat suggest that CTXLP's conserved cysteine motif may also contribute to neurotoxicity and retroviral transcription. Other retroviral proteins examined (HTLV Tax) did not show any homology to ERVK CTXLP, suggesting that this pathogenic mechanism is not conserved among all retroviruses.

[00276] *Identification of CTXLP-encoding ERVK loci in the human genome*

[00277] Nomenclature for each ERVK loci is based on their common names, as well as their chromosome location. Geneious was used to align both ERVK *rec* exon 1 and the predicted CTXLP DNA sequence with 95 ERVK HML-2 insertions identified in the human genome. After the ERVK insertions were aligned, many insertions (33) were excluded from further analysis due to the absence of an intact env. The ERVK insertions with an intact env were then aligned to both the *rec* gene and CTXLP cysteine motif nucleotide sequences. Figure 15 shows an example of the *rec* exon 1 and CTXLP nucleotide alignments against intact ERVK envelope genes. The CTXLP sequence was found at bp 1413 of env, just at the 3' end of the coding region for the envelope SU protein. This location is the cleavage site of the envelope polyprotein, where there is a junction between the SU and TM proteins. The aligned nucleotide sequences were then translated into an amino acid sequence and examined for intact Rec and CTXLP coding sequences. Table 3 shows that there are 25 ERVK insertions (out of the 62 examined) capable of producing a CTXLP protein, with 5 proviruses also capable of producing Rec.

Table 3: Sixty-two ERVK HML-2 human insertions and their chromosomal location examined for an intact Rec and intact CTXLP ORF along with any known disease associations with Multiple Sclerosis, Cancer or Schizophrenia.

Accession number	Genomic location	ERVK insertion	Intact Rec?	Intact Conotoxin?
JN675007	1p31.1	**ERVK-1[a]_HML-2_1p31.1_75842771**	No	**Yes**
JN675010	1p36.21b	ERVK-76[b]_HML-2_1p36.21b_ 13458305	No	No
JN675011	1p36.21c	ERVK-76[b]_HML-2_1p36.21c_13678850	No	No
JN675013	1q23.3	**ERVK-18[a]_HML-2_1q23.3_ 160660575**	No	**Yes**
JN675014	1q22	**ERVK-7[a]_HML-2_1q22_ 155596457**	No	**Yes**
JN675015	1q24.1	ERVK-12[b]_HML-2_1q24.1_ 166574603	No	No
JN675016	1q32.2	ERVK_ HML-2_1q32.2_ 207808457	No	No
JN675018	2q21.1	**ERVK_HML-2_2q21.1_ 130719538**	No	**Yes**
JN675019	3p12.3	ERVK_HML-2_3p12.3_ 75600465	No	No
JN675020	3p25.3	ERVK-2[a,b]_HML-2_3p25.3_9889346	No	No
JN675021	3q12.3	**ERVK-5[a]_HML-2_3q12.3_ 101410737**	No	**Yes**
JN675022	3q13.2	**ERVK-3[a]_HML-2_3q13.2_112743479**	No	**Yes**
JN675023	3q21.2	**ERVK-4[a]_HML-2_3q21.2_ 125609302**	No	**Yes**
JN675025	3q27.2	**ERVK-11[a]_HML-2_3q27.2_185280336**	No	**Yes**
JN675026	4p16.1a	ERVK-17[b]_HML-2_4p16.1a_9123515	No	No
JN675027	4p16.1b	ERVK-50c[b]_HML-2_4p16.1b_9659588	No	No
JN675029	4p16.3b	ERVK-7[b]_HML-2_4p16.3b_3980069	No	No
JN675030	4q13.2	ERVK_HML-2_4q13.2_463709	No	No
JN675032	4q32.3	ERVK-13[a]_HML-2_4q32.3_5916840	No	No
JN675034	5p12	ERVK_HML-2_5p12_46000159	No	No
JN675035	5p13.3	**ERVK-104[b]_HML-2_5p13.3_30487114**	No	**Yes**
JN675036	5q33.2	ERVK-18b[b]_HML-2_5q33.2_154016502	No	No
JN675037	5q33.3	**ERVK-10 [a]_HML-2_5q33.3_156084717**	No	**Yes**
JN675039	6p21.1	ERVK-OLD35587b_HML-2_6p22.1_42861409	No	No
JN675040	6p22.1	ERVK-69_HML-2_6p22.128650367	No	No
JN675041	6q14.1	**ERVK-9[a]_HML-2_6q14.1_78427019**	**Yes**	**Yes**
JN675043	7p22.1a	**ERVK-14[a]_HML-2_4622057**	**Yes**	**Yes**
JN675044	7p22.1b	**ERVK-14[a]_HML-2_4630561**	**Yes**	**Yes**
JN675049	8p23.1a	**ERVK-8[a]_HML-2_8p23.1a _7355397**	No	**Yes**
JN675050	8p23.1b	ERVK-27[b]_HML-2_8p23.1b_8054700	No	No
JN675051	8p23.1c	ERVK_HML-2_8p23.1_12073970	No	No
JN675052	8p23.1d	ERVKOLD130352[b]_HML-2_8p23.1d_12316492 1	No	No
JN675053	8q11.1	ERVK-70[b]_HML-2_8q11.1_47175650	No	No
JN675057	9q34.11	ERVK-31[b]_HML-2_9q34.11_131612515	No	No
JN675058	10p12.1	**ERVK-103[b]_HML-2_10p12.1_27182399**	No	**Yes**
JN675059	10p14	ERVK-16[a]_HML-2_10p14_6867109	No	No
JN675060	10q24.2	ERVK-17[a]_HML-2_10q24.2_101580569	No	No
JN675061	11p15.4	ERVK3-4[a]_HML-2_11p15.4_ 3468656	No	No
JN675062	11q12.1	ERVK_HML-2_ 11q12.1_58767448	No	No
JN675063	11q12.3	ERVK-OLDAC004127[b]_HML-2_11q12.3_ 62135963	No	No
JN675064	11q22.1	**ERVK-25[a]_HML-2_11q22.1_ 101565794**	**Yes**	**Yes**
JN675065	11q23.3	**ERVK-20[a]_HML-2_ 11q23.3_118591724**	No	**Yes**
JN675066	12p11.1	ERVK-50E[b]_HML-2_12p11.1_ 34772555	No	No
JN675067	12q13.2	**ERVK_HML-2_12q13.2_ 55727215**	No	**Yes**
JN675068	12q14.1	**ERVK-21[b]_HML-2_12q14.1_ 58721242**	No	**Yes**

JN675073	15q25.2	ERVK_HML-2_15q25.2_84829020	No	No
JN675074	16p11.2	**ERVK_HML-2_16p11.2_ 34231474**	No	**Yes**
JN675075	17p13.1	ERVK_HML-2_17p13.1_7960357	No	No
JN675076	19p12a	ERVK52[b]_HML-2_ 19p12a _20387400	No	No
JN675077	19p12b	**ERVK113[b]_HML-2_19p12b_21841536**	**Yes**	**Yes**
JN675078	19p12c	**ERVK51[b]_HML-2_19p12c_22757824**	No	**Yes**
JN675080	19q11	ERVK-19[b]_HML-2_19q11_228128498	Yes	No
JN675081	19q13.12a	ERVK_HML-2_19q13.12a_36063207	No	No
JN675082	19q13.12b	ERVKOLD12309_HML-2_19q13.12b_37597549	No	No
JN675083	19q13.41	ERVK3-6[a]_HML-2_19q13.41_53248274	No	No
JN675084	19q13.42	LTR13[b]_HML-2__19q13.42_53862348	No	No
JN675085	20q11.22	**ERVK59[b]_HML-2_20q11.22_32714750**	No	**Yes**
JN675086	21q21.1	**ERVK-23[a]_HML-2_21q21.1_19933916**	No	**Yes**
JN675087	22q11.21	**ERVK-24[a]_HML-2_22q11.21_18926187**	No	**Yes**
JN675088	22q11.23	ERVK-KOLD345b_HML-2_22q11.23_23879930	No	No
JN675090	Xq11.1	ERVK_HML-2_Xq11.1_61959549	No	No
JN675094	Yp11.2	ERVK_HML-2_ Yp11.2_6 826441	No	No

Disease associations: MS (yellow,) MS (No CTXLP; pale yellow), Cancer (Green), Cancer (No CTXLP; dark green), Schizophrenia (blue).

a) Mayer, J., Blomberg, J., & Seal, R. L. (2011). A revised nomenclature for transcribed human endogenous retroviral loci. Mobile DNA, 2(1), 7.

b) Subramanian, R. P., Wildschutte, J. H., Russo, C., & Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology, 8(1), 90.

[00278] Of the 95 ERVK DNA sequences, 33 were excluded due to an incomplete env sequence. The remaining 62 sequences were then translated and examined for an intact CTXLP in the appropriate reading frame. The resulting CTXLP peptide sequences were aligned and a sequence logo and consensus sequence were generated to assess amino acid conservation and detect polymorphisms (**Figure 16**).

[00279] In total, 25 ERVK insertions containing the CTXLP cysteine motif were analysed for overall conservation of the peptide sequence and to identify specific variants. **Figure 16** shows that although each CTXLP peptide has a nearly identical amino acid sequence (identity 95.6%), there are distinct CTXLP polymorphisms identified.

[00280] Ten distinct CTXLP polymorphisms were detected. The most prevalent polymorphism is found in ERVK-3, ERVK-104, ERVK-10, ERVK-9, ERVK-14, ERVK-14(b), ERVK-8, ERVK-103, ERVK-25, ERVK-7, ERVK-21, ERVK-16p11.2, and ERVK-113 (Allele 1). The second most prevalent substitutes glycine for serine, relative to the consensus, at two alignment positions (Ser16Gly, Ser27Gly) and is found in ERVK-5 and ERVK-20 (Allele 2). ERVK-HML-2_2q21.2 differs only at the latter Serine (Ser27Gly). ERVK-18 differs only at the former Serine (Ser16Gly). ERVK-51 has the former variation as well as a valine in position 20

(Ser16Gly, Ile20Val). ERVK-1 has phenylalanine at position 22 (Leu22Phe). ERVK-4 contains arginine at position 5 (Gly5Arg). ERVK-HML-2_12q13.2 contained an asparagine at position 16 (Ser16Asn). ERVK-23 shows three polymorphisms at positions 5, 17, and 27 (Gly5Arg, Arg17Lys, Val26Glu). The prevalence and polymorphic variability of ERVK CTXLP-encoding insertions suggests that CTXLP is a pervasive and conserved ERVK protein.

[00281] Out of the identified CTXLP encoding proviruses, 20 of 25 ERVK insertions were human-specific[43]. ERVK-18, ERVK-5, ERVK-69, ERVK-20, ERVK-HML-2_16p21, and ERVK-51 are found in other primates including orangutan, chimpanzee, and rhesus monkey, demonstrating that ERVK CTXLP is an evolutionarily conserved protein, which either entered the genome of a common primate ancestor or through cross-species infection with a specific CTXLP-encoding ERVK virus[16] (See **Figure 59** below in "primate models" section). The evolutionary age and clade of ERVK retroviruses that encode CTXLP suggests that CTXLP is unique to primates[43,44], and that the human genome has an enrichment of this type of ERVK proviruses.

[00282] A re-analysis of CTXLP variants in the human genome using a different methodology resulted in similar conclusion regarding the polymorphic nature of CTXLP+ ERVK genomes (**Figure 17**).

[00283] *ERVK CTXLP domain and disease associations*

[00284] CTXLP was identified in both type 1 and type 2 ERVK (Figure 1). The prevalence of CTXLP in both types of ERVK suggests that CTXLP originated early on in ERVK evolution, being present before the divergence of ERVK Δ *env* genomes[45]. Select CTXLP+ loci had known disease associations (**Table 3**). Out of the 25 CTXLP-encoding ERVK insertions examined, no insertions were associated with ALS – most likely due to a lack of research on this topic. Two of the 25 insertions (ERVK-18 and ERVK-10) were associated with the psychiatric condition schizophrenia. Two out of 3 insertions associated with MS contained CTXLP-encoding insertions. ERVK expression has previously been associated with neurological disease[6], suggesting that CTXLP may be one mechanism for neurotoxicity. Surprisingly, 9 CTXLP-encoding insertions were associated with cancer. ERVK *env* expression has previously been identified in several cancers[46]. As we demonstrate below (see "CTXLP and disease" section, **Figures 25, 36, 43 and 44**), there is a notable association between CTXLP and cancer, suggesting that currently identified (**Table 3**) and other CTXLP-encoding loci with no known disease association may serve as future biomarkers for cancer. The

different polymorphisms of CTXLP identified did not associate with any specific disease conditions. Any effects these polymorphisms have on CTXLP function remains unknown.

[00285] Identification of ERVK CTXLP – an alternate form of the ERVK envelope protein

[00286] *Predicted full ERVK CTXLP protein sequence*

[00287] The results of both the Pfam and NCBI-CDD databases indicated that the predicted CTXLP amino acid sequence (used to produce the CTXLP plasmids described below) shares similarities with both ERVK Rec, an oncogenic alternate splice product of the *env* gene, and ERVK Env. These results support the prediction that CTXLP is partially composed of the SU unit of the Env glycoprotein. The NCBI-CDD database also indicated the presence of a surface glycoprotein signal peptide domain. Lastly, the C-terminal portion of the CTXLP sequence was found to share similarities with the O-conotoxin superfamily, which ω-conotoxins are a part of. Lastly, the DUF4408 domain corresponds to a domain of unknown function which is primarily found in plants. Together, these results suggest that CTXLP is composed of the ERVK Env SU unit with a C-terminal ω-conotoxin domain (**Figure 18**).

[00288] *Programmed frameshifting and internal ribosomal entry site*

[00289] Since the reading frames of ERVK *env* (frame +1) and CTXLP (frame +3) differed by -1, the ERVK *env* transcript (**Figure 19**) was examined for evidence of -1 programmed ribosomal frameshifting (-1 PRF), using ERVK-1, ERVK-18, ERVK-7, ERVK_HML-2_2q21.1, ERVK-5, ERVK-3, ERVK-4, ERKV-104, and ERVK-10 insertions.

[00290] *RNAfold analysis of RNA structures in the ERVK env transcript*

[00291] Our biomedical experiments suggested that there were CTXLP isoforms of different sizes, therefore we examined whether the conventional and alternative methionine start sites could be used to make both long and short CTXLP proteins. The first 350 bp of the ERVK Env-encoding RNA was inserted into RNAfold software to predict RNA secondary structure (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi). **Figure 19B** suggests that the CTLXP proteins can originate from the conventional Env methionine start or potentially use IRES-like RNA hairpins to initiate translation at a downstream methionine.

[00292] This alternative mechanism for CTXLP expression is the use of internal ribosomal entry site (IRES), using an alternative translational start site (**Figure 19B**). The 51 and 32 kDa isoforms of CTXLP likely form using the start of the *env* ORF (nucleotide position 1) or an IRES in the *env* reading frame, starting from one of its many methionine start codons (specifically amino acid position 200). Translating *env* from alternative start codons would result in different

sized isoforms of CTXLP. Certain viruses, including HIV, use IRES to allow for mRNA translation to begin in the middle of the transcript[48,49]. The mRNA forms a hairpin structure that allows the binding of the 40S RNA subunit followed by protein translation[50]. Using RNAfold software to predict mRNA secondary structure upstream of CTXLP, an RNA fold structure similar to HIV IRES was predicted (nucleotide 84 to 187, and 213 to 318 in *env* transcript), that could take advantage of alternate methionine start codon.

[00293] ERVK *env* nucleotide sequences were also inserted into RNAfold starting from 150 base pairs upstream of the CTXLP ORF to predict RNA secondary structure. RNA secondary structure was examined for evidence of -1 programmed ribosomal frameshifting motifs, including a slippery site with the X-XXY-YYZ form, a 5-10 nucleotide spacer and a downstream pseudoknot or hairpin structure (**Figure 19C**). The 1-350 bp region of the ERVK *env* gene was also examined for RNA secondary structure motifs, as it contains numerous alternative methionine start sites. RNA secondary structure was then examined for potential internal ribosomal entry site (IRES) binding sites, which take the form of complex RNA hairpins[48]. The likelihood that these ERVK RNA secondary structures represent an IRES was determined using IRES prediction software called IRESite (http://iresite.org/IRESite_web.php), and reported as a similarity with known cellular and viral IRES 2D structures.

[00294] PRF can occur when three elements are combined: i) a slippery site containing an X-XXY-YYZ motif which after frameshifting by -1 results in XXX-YYY reading, ii) a 5 to 10 nucleotide spacer sequence, and iii) a downstream hairpin-type pseudoknot. ERVK CTXLP-encoding insertions contained an appropriate U-UUA-AAU slippery site to allow for -1 frameshifting to UUU-AAA. After the slippery site there was a 5 nucleotide spacer sequence before the CTXLP ORF. All sequences examined showed a strong probability of forming a hairpin-type pseudoknot within the RNA sequence encoding the CTXLP cysteine-rich motif (**Figure 19C**). Paradoxically, if the envelope protein translates past the CTXLP ORF start and frameshifts by -4, this introduces a conserved KRQK nuclear localization sequence (NLS) into the hypothetical protein (**Figure 20**). Another transcription factor that contains a KRQK NLS is NF-κB p50[51]. There are no other known NLS in either the ERVK envelope protein or the predicted ERVK CTXLP proteins.

[00295] If CTXLP encoding originates from the conventional start site in the *env* transcript, followed by a -1 PRF then it may produce a 51 kDa CTXLP protein. Alternatively, if CTXLP encoding originates from an IRES site using an alternate methionine in the *env* transcript (**Figure 20**), followed by a -1 PRF then it may produce a 32 kDa CTXLP protein. The insertions that

showed the highest probability of forming a potential IRES binding site upstream of the CTXLP ORF were ERVK-113 and ERVK-4. These sequences analysed by IRESite software predict that the IRES-like RNA hairpins most resemble those found in HIV, Theiler's Murine Encephalomyelitis Virus (TMEV), and Hepatitis C virus (HCV).

[00296] We had previously hypothesized that ERVK CTXLP be produced via alternative splicing of the Rec transcript. Although alternative splicing is a common mechanism in retroviruses, only 4 of 25 CTXLP-encoding insertions also had intact Rec protein. The prevalence of CTXLP-encoding insertions in the absence of Rec suggests that alternative splicing is not the mechanism of CTXLP formation. Upstream and downstream of the CTXLP ORF are several stop codons. One mechanism that retroviruses use to compensate for stop codons or a lack of methionine starts is called programmed minus-one ribosomal frameshifting[52]. This involves the formation of an H-type pseudoknot in the RNA transcript at the site of frameshifting. The H-type pseudoknot would likely halt the ribosome from continuing translation, leading to the -1 PRF[52]. The slippery site UUU-AAA-U would re-establish ribosomal tRNA and mRNA base pairing and allow for the continuation of translation after frameshifting[52]. We predict that this mechanism could be used to extend the ORF of the ERVK SU protein by adding on a C-terminal CTXLP domain (**Figures 19 & 20**). Although the predicted structural stability of the ERVK *env* RNA H-type pseudoknot (**Figure 18**) was stronger in some ERVK insertions than others, the presence of this RNA secondary structure was conserved in all sequences examined using RNAfold. To our knowledge, frameshifting under conditions of inflammation has not been previously described. Enhanced frameshifting due to inflammation may be a unique mechanism in retroviruses, particularly upon exposure to TNFα (**Figures 27 and 28**).

[00297] Together, this data suggests that ERVK CTXLP is likely expressed as a cryptic peptide through frameshifted translation of the *env* transcript (**Figures 19 & 20**). Alternative forms of viral *env* proteins (CTXLP proteins may be considered isoforms of *env*) may be translated under specific physiological conditions[53,54]. When ERVK *env* is translated in this proposed alternative ORF, the CTXLP peptide would be expressed within the translated *env* as a cryptic peptide. Cryptic peptides are proposed to have significantly different functions than their precursor protein[55]. Cryptic epitopes within modified HIV proteins have been shown to be immunogenic[56,57].

[00298] Therefore, ERVK CTXLP is likely formed from a -4 PRF occurring slightly upstream of the furin cleavage site in *env*. An IRES sequence likely allows for a shorter ERVK CTXLP protein isoform to be produced, explaining the distinct isoforms of CTXLP identified.

[00299] Prediction of post-translational modifications for CTXLP

[00300] ERVK CTXLP is predicted to have the following post-translational modifications, including, but not limited to phosphorylation, SUMOylation, glycosylation and lipid addition (**Figures 20 & 21, Tables 4-8**) using ELM software[47] and other resources.

Table 4: Predicted phosphorylation sites within ERVK CTXLP protein

Kinase	Predicted phosphorylation sites	ELM	NetPhos3.1	Motif Scan
CDC2	41		●	
	213		●	
	279		●	
	281		●	
	321		●	
	374		●	
	376		●	
	417		●	
	438		●	
	475		●	
	482		●	
CDK	184-191	●		
CDK5	72		●	
	187		●	
	288		●	
	329		●	

CK1	187-193	●		
	190-196	●	●	
	198-204	●		
	235-241	●		
	276-282	●		
	281-287	●		
	293-299	●		
	329-335	●		
	373-379	●		
CK2	32-35	●	●	●
	116-119			●
	260-263			●
	293		●	●
	331-334	●	●	●
	411-414			●
DNAPK	298		●	
	374		●	
	400		●	
EGFR	272		●	
GSK3	54-61	●		
	82-89	●		
	210-217	●		
	281-288	●		
	350-357	●		
	397-404	●		
	409-416	●		
INSR	160		●	
LATS	304-310	●		
MAPK/PDK	69-75	●		
	184-190	●		
	285-291	●		
	326-332	●		
	443-349	●		
NEK2	64-69	●		
	82-87	●		
	86-91	●		
	371-376	●		
	460-465	●		
	495-500	●		
PIKK	54-60	●		
	67-79	●		

	276-282	●		
	397-403	●		
PKA	54-60	●		
	155		●	
	190-196	●		
	250		●	
	340-346	●		
	353		●	
	491		●	
PKC	24-26		●	●
	41-43		●	●
	57		●	
	117		●	
	193		●	
	200-203		●	●
	216-218		●	●
	279		●	
	281		●	
	298		●	
	309		●	
	321-324		●	●
	343		●	
	360		●	
	376		●	
	381		●	
	412		●	
	417		●	
	462-464		●	●
PLK	414-420	●		
p38 MAPK	72		●	
	187		●	
	329		●	

Table 5: Predicted SUMOylation and SUMO interaction sites within ERVK CTXLP protein

Predicted SUMOylation sites	ELM	GPS-SUMO
143-146	●	●
397		●
Predicted SUMO interaction sites		
123-127		●
180-184	●	●
367-374	●	
424-428		●

Table 6: Predicted glycosylation sites within ERVK CTXLP protein

Sugar attachment	Predicted glycosylation sites	ELM	Netglyc4.0	Motif Scan
Fucose	275-282	●		
	408-413	●		
	487-492	●		
Glycosaminoglycan	197-200	●		
	249-252	●		
	275-278	●		
	283-286	●		
	331-334	●		
	352-355	●		
N-glycosylation	99-102	●	●	
	127-130	●	●	
	152-155	●	●	
	273-276	●	●	
	354-357	●	●	
	371-374	●	●	
	460-463	●	●	
	472-475	●	●	
	479-482	●	●	
O-glycosylation	213			●
	217			●
	235			●
	284			●
	288			●
	321			●

Table 7: Predicted lipid attachment sites within ERVK CTXLP protein

Lipid attachment	Predicted lipid addition sites	CSS-Palm	GPS-Lipid
S-Farnesylation	497		●
	503		●
S-Geranylgeranylation	503		●
S-palmitoylation	140	●	
	141	●	
	227		●
	275		●
	382	●	
	408		●
	487	●	●
	488		●

[00301] The N-linked glycosylation of ERVK CTXLP has been verified experimentally using PNGase treatment of CTXLP protein fractions, followed by western blot analysis for shifts in high molecular weight protein banding patterns (Figure 26, n=3). Phosphorylation is postulated to occur to ERVK CTXLP when bound to chromatin, as seen in the 32kDa band shift in Figure 30.

[00302] Moreover, bioinformatic predictions also predicted protein cleavage and interaction sites within CTXLP (**Table 8**) using the PROSPER website (https://prosper.erc.monash.edu.au/home.html). Among the proteins predicted to cleave CTXLP were HIV protease, furin, NEC1, and NEC2. The predicted furin cleavage site is consistent with the location of the known furin cleavage site that typically cleaves the Env polyprotein into discrete SU and TM peptide chains that are then assembled into multimer proteins. However, it is unclear how an overlapping NRS N-linked glycosylation site would impact the ability of furin the cleave the site. As well, the predicted cleavage by HIV protease is interesting as ERVK interactions with HIV proteins have been previously reported[42,58,59]. Of note, cleavage predictions only take into account primary amino acid sequence only, and do not account for how viral protein tertiary structure and cellular factors come into play.

Table 8: Predicted protease cleavage sites within CTXLP, as predicted by PROSPER.

Merops ID	Protease Name	Position	P4-P4' site	N-fragment (kDa)	C-fragment (kDa)	Cleavage score
A02.001	HIV-1 retropepsin	SEQ ID NO: 139 130	TEVL|WEEC	15.57	30.67	1.24
A02.001	HIV-1 retropepsin	SEQ ID NO: 140 348	KGVL|IQKI	41.54	4.69	1.05
A02.001	HIV-1 retropepsin	SEQ ID NO: 141 277	PYML|VVGN	33.14	13.09	0.99
A02.001	HIV-1 retropepsin	SEQ ID NO: 142 71	WLVE|VPTV	8.5	37.74	0.94
C01.036	cathepsin K	SEQ ID NO: 143 268	VPLQ|SCVK	32.12	14.11	1.15
C01.036	cathepsin K	SEQ ID NO: 144 201	FYLW|EWEE	23.97	22.26	1.15
C01.036	cathepsin K	SEQ ID NO: 145 164	HNCS|GQTQ	19.72	26.51	1.14
C01.036	cathepsin K	SEQ ID NO: 146 196	KKLQ|SFYL	23.28	22.96	1.13
C01.036	cathepsin K	SEQ ID NO: 147 143	VILQ|NNEF	17.04	29.2	1.12
C01.036	cathepsin K	SEQ ID NO: 148 78	VSPN|SRFT	9.19	37.04	1.1
C01.036	cathepsin K	SEQ ID NO: 149 341	HILT|EILK	40.67	5.56	1.09
C01.036	cathepsin K	SEQ ID NO: 150 247	QTLE|TRYR	29.6	16.63	1.08
C01.036	cathepsin K	SEQ ID NO: 151 199	QSFY|LWEW	23.68	22.56	1.08
C01.036	cathepsin K	SEQ ID NO: 152 238	HHIR|IWSG	28.46	17.78	1.06
C01.036	cathepsin K	SEQ ID NO: 153 23	VWVP|GPTD	2.67	43.57	0.98
C01.036	cathepsin K	SEQ ID NO: 154 223	SGPE|HPEL	26.62	19.61	0.97
C01.036	cathepsin K	SEQ ID NO: 155 209	KGIS|TPRP	25.05	21.18	0.92
M10.003	matrix metallopeptidase-2	SEQ ID NO: 156 88	MVSG|MSLR	10.47	35.76	1.15
M10.004	matrix metallopeptidase-9	SEQ ID NO: 157 84	FTYH|MVSG	9.98	36.25	1.33
M10.004	matrix metallopeptidase-9	SEQ ID NO: 158 13	NPIE|VYVN	1.51	44.72	1.28
M10.004	matrix metallopeptidase-9	SEQ ID NO: 159 276	PPYM|LVVG	33.03	13.21	1.26
M10.004	matrix metallopeptidase-9	SEQ ID NO: 160 65	MPAV|QNWL	7.73	38.51	1.19
M10.004	matrix metallopeptidase-9	SEQ ID NO: 161 113	RPKG|KTCP	13.6	32.64	1.15
M10.004	matrix metallopeptidase-9	SEQ ID NO: 162 173	CPSA|QVSP	20.69	25.54	1.14
M10.004	matrix metallopeptidase-9	SEQ ID NO: 163 129	NTEV|LWEE	15.45	30.78	1.13
M10.004	matrix metallopeptidase-9	SEQ ID NO: 164 213	TPRP|KIIS	25.5	20.73	1.12
M10.004	matrix metallopeptidase-9	SEQ ID NO: 165 88	MVSG|MSLR	10.47	35.76	1.11
M10.004	matrix metallopeptidase-9	SEQ ID NO: 166 157	APRG|QFYH	18.84	27.39	1.09
M10.004	matrix metallopeptidase-9	SEQ ID NO: 167 6	TPVT|WMDN	0.62	45.61	1.08
M10.004	matrix metallopeptidase-9	SEQ ID NO: 168 90	SGMS|LRPR	10.69	35.54	1.08
M10.004	matrix metallopeptidase-9	SEQ ID NO: 169 138	VANS|VVIL	16.48	29.75	1.08
M10.004	matrix metallopeptidase-9	SEQ ID NO: 170 275	KPPY|MLVV	32.89	13.34	1.08
M10.004	matrix metallopeptidase-9	SEQ ID NO: 171 182	VDSD|LTES	21.59	24.64	1.07
M10.004	matrix metallopeptidase-9	SEQ ID NO: 172 38	EEEG|MMIN	4.48	41.75	1.06
M10.004	matrix metallopeptidase-9	SEQ ID NO: 173 339	SIHI|LTEI	40.46	5.78	1.05

M10.004	matrix metallopeptidase-9	SEQ ID NO: 174	4	MVTP|VTWM	0.42	45.81	1.05
M10.004	matrix metallopeptidase-9	SEQ ID NO: 175	274	VKPP|YMLV	32.73	13.5	1.05
M10.004	matrix metallopeptidase-9	SEQ ID NO: 176	362	SDYG|INCS	43.37	2.86	1.04
M10.004	matrix metallopeptidase-9	SEQ ID NO: 177	281	VVGN|IVIK	33.62	12.61	1.04
M10.004	matrix metallopeptidase-9	SEQ ID NO: 178	351	LIQK|IHFY	41.91	4.32	1.03
M10.004	matrix metallopeptidase-9	SEQ ID NO: 179	266	LTVP|LQSC	31.88	14.35	1.03
M10.004	matrix metallopeptidase-9	SEQ ID NO: 180	313	HRIL|LVRA	37.36	8.88	1.02
M10.004	matrix metallopeptidase-9	SEQ ID NO: 181	347	LKGV|LIQK	41.43	4.81	1.02
M10.004	matrix metallopeptidase-9	SEQ ID NO: 182	39	EEGM|MINI	4.61	41.62	1.01
M10.004	matrix metallopeptidase-9	SEQ ID NO: 183	385	SVSK|LC	46.02	0.21	0.99
M10.005	matrix metallopeptidase-3	SEQ ID NO: 184	173	CPSA|QVSP	20.69	25.54	1.16
M10.005	matrix metallopeptidase-3	SEQ ID NO: 185	137	CVAN|SVVI	16.4	29.84	1.02
M10.005	matrix metallopeptidase-3	SEQ ID NO: 186	288	KPAS|QTIT	34.33	11.9	0.99
M10.005	matrix metallopeptidase-3	SEQ ID NO: 187	276	PPYM|LVVG	33.03	13.21	0.96
S01.001	chymotrypsin A (cattle-type)	SEQ ID NO: 188	142	VVIL|QNNE	16.91	29.33	1.02
S01.001	chymotrypsin A (cattle-type)	SEQ ID NO: 189	83	RFTY|HMVS	9.85	36.39	1.01
S01.001	chymotrypsin A (cattle-type)	SEQ ID NO: 190	228	PELW|RLTV	27.29	18.95	0.97
S01.131	elastase-2	SEQ ID NO: 191	219	ISPV|SGPE	26.14	20.1	1.41
S01.131	elastase-2	SEQ ID NO: 192	352	IQKI|HFYF	42.02	4.21	1.37
S01.131	elastase-2	SEQ ID NO: 193	140	NSVV|ILQN	16.68	29.55	1.21
S01.131	elastase-2	SEQ ID NO: 194	377	RSCI|ALFC	45.19	1.05	1.18
S01.131	elastase-2	SEQ ID NO: 195	52	YPPI|CLGR	6.23	40.01	1.18
S01.131	elastase-2	SEQ ID NO: 196	20	NDSV|WVPG	2.29	43.95	1.17
S01.131	elastase-2	SEQ ID NO: 197	315	ILLV|RARE	37.57	8.66	1.16
S01.131	elastase-2	SEQ ID NO: 198	43	MINI|SIGY	5.08	41.15	1.12
S01.131	elastase-2	SEQ ID NO: 199	278	YMLV|VGNI	33.24	13	1.11
S01.131	elastase-2	SEQ ID NO: 200	86	YHMV|SGMS	10.21	36.02	1.09
S01.131	elastase-2	SEQ ID NO: 201	237	SHHI|RIWS	28.3	17.93	1.09
S01.131	elastase-2	SEQ ID NO: 202	349	GVLI|QKIH	41.65	4.58	1.09
S01.131	elastase-2	SEQ ID NO: 203	65	MPAV|QNWL	7.73	38.51	1.09
S01.131	elastase-2	SEQ ID NO: 204	41	GMMI|NISI	4.86	41.38	1.08
S01.131	elastase-2	SEQ ID NO: 205	5	VTPV|TWMD	0.52	45.71	1.07
S01.131	elastase-2	SEQ ID NO: 206	233	LTVA|SHHI	27.83	18.41	1.07
S01.131	elastase-2	SEQ ID NO: 207	151	GTII|DWAP	18.04	28.19	1.07
S01.131	elastase-2	SEQ ID NO: 208	45	NISI|GYHY	5.28	40.95	1.04
S01.131	elastase-2	SEQ ID NO: 209	129	NTEV|LWEE	15.45	30.78	1.02
S01.131	elastase-2	SEQ ID NO: 210	325	WIPV|STDR	38.88	7.35	1
S01.133	cathepsin G	SEQ ID NO: 211	214	PRPK|IISP	25.63	20.61	1.33

S01.133	cathepsin G	SEQ ID NO: 212	160	GQFY|HNCS	19.28	26.96	1.24
S01.133	cathepsin G	SEQ ID NO: 213	40	EGMM|INIS	4.74	41.49	1.18
S01.133	cathepsin G	SEQ ID NO: 214	277	PYML|VVGN	33.14	13.09	1.17
S01.133	cathepsin G	SEQ ID NO: 215	260	IDLN|SILT	31.27	14.96	1.11
S01.133	cathepsin G	SEQ ID NO: 216	15	IEVY|VNDS	1.77	44.46	1.1
S01.133	cathepsin G	SEQ ID NO: 217	240	IRIW|SGNQ	28.76	17.48	1.09
S01.133	cathepsin G	SEQ ID NO: 218	8	VTWM|DNPI	0.94	45.29	1.06
S01.133	cathepsin G	SEQ ID NO: 219	83	RFTY|HMVS	9.85	36.39	1.05
S01.133	cathepsin G	SEQ ID NO: 220	47	SIGY|HYPP	5.62	40.61	1.04
S01.133	cathepsin G	SEQ ID NO: 221	313	HRIL|LVRA	37.36	8.88	1.03
S01.133	cathepsin G	SEQ ID NO: 222	276	PPYM|LVVG	33.03	13.21	1.01
S01.133	cathepsin G	SEQ ID NO: 223	199	QSFY|LWEW	23.68	22.56	0.99
S01.133	cathepsin G	SEQ ID NO: 224	321	REGM|WIPV	38.38	7.85	0.99
S01.133	cathepsin G	SEQ ID NO: 225	39	EEGM|MINI	4.61	41.62	0.96
S01.133	cathepsin G	SEQ ID NO: 226	361	CSDY|GINC	43.2	3.04	0.94
S01.133	cathepsin G	SEQ ID NO: 227	7	PVTW|MDNP	0.81	45.42	0.93
S01.133	cathepsin G	SEQ ID NO: 228	89	VSGM|SLRP	10.6	35.63	0.92
S01.217	thrombin	SEQ ID NO: 229	214	PRPK|IISP	25.63	20.61	1.03
S01.269	glutamyl peptidase I	SEQ ID NO: 230	319	RARE|GMWI	38.08	8.15	1.04
S01.269	glutamyl peptidase I	SEQ ID NO: 231	185	DLTE|SLDK	21.94	24.3	1
S26.008	thylakoidal processing peptidase	SEQ ID NO: 232	102	QDFS|YQRS	12.12	34.11	0.94
S26.010	signalase (animal) 21 kDa component	SEQ ID NO: 233	317	LVRA|REGM	37.8	8.44	0.95

[00303] Lastly, the predicted protein interactions of cellular proteins and CTXLP were equally intriguing (Table 9). Among the predicted interactions were proteins involved in cell cycle regulation such as MAPK and LATS. As well, interactions were also predicted with proteins involved in innate immunity such as UPS-7/HAUSP, TRAF-2 and TRAF-6, which are upstream of NF-κB in inflammatory signalling.

Table 9: Predicted protein interaction partners of CTXLP using ELM software.

Predicted protein interaction partner	Predicted CTXLP interaction sites
BRCA1	216-220
Calcineurin	90-93
	375-378
Cyclin	216-220
	423-427
Dyenein	66-72
MAPKs	233-242
	361-371
	363-371
	428-437
	462-470
	464-470
MAPKs (ERK1/2 and p38 subfamilies)	425-437
	427-439
Pin1	69-74
	184-189
	285-290
	326-331
	443-448
Protein phosphatase 1 (catalytic subunit)	112-119
	214-221
	465-471
STAT5	63-66
	101-104
	106-109
	126-129
	208-211
	367-370
	471-474
SUMO	179-185
	367-374
TRAF2	32-35
	145-148

TRAF6	310-318
USP7/HAUSP	219-223
	225-229
	229-233
	232-236
	290-294
	301-305
	378-382
	445-449

[00304] **Design of a custom ERVK CTXLP antibody**

[00305] Pierce Custom antibody services has produced a CTXLP-specific polyclonal rabbit antibody, used in all the experiments described below. The predicted epitopes are listed in Figure 22. The rabbit protocol and immunization plan is stated in Figure 23.

[00306] **Design of a custom ERVK CTXLP vector, and complementary ERVK Env and ERVK SU vector**

[00307] GenScript custom plasmid services has produced an ERVK CTXLP-expressing vector within a pcDNA3.1 backbone, used in all the experiments described below. We also synthesized a matching ERVK SU vector as a complementary plasmid devoid of the CTXLP domain. The sequences used to produce the vectors are listed below:

[00308] **ERVK SU + ERVK CTXLP (Based on ERVK113): 57.14 kDa [SEQ ID NO:2]**

[00309] MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPT WAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAANYTYWAYVPFP PLIRAVTWMDNPIEIYVNDSVWVPGPTDDCCPAKPEEEGMMINISIGYRYPPICLGRAPGCL MPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIP KESKNTEVLVWEECVANSAVILQNNEFGTLIDWAPRGQFYHNCSGQTQSCPSAQVSPAVD SDLTESLDKHKHKKLQSFYPWEWGEKGISTARPKIISPVSGPEHPELWRLTVASHHIRIWSG NQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTF NWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKR**QKIHF YFNCSDYGINCSHSYGCCSRSCIALFCSVSKLC***

[00310] First portion is ERVK Env SU, aa residues 1-465 (furin cleavage site) of ERVK 113 (19p12b) with normal frame. Second portion is ERVK CTXLP, aa residues 463-500 of ERVK 113 (19p12b) with +3 reading frame and no start codon bias. Env SU ends at KR before bolded font indicating where Env CTXLP starts (QK). Bolded portion of sequence represents the allele 1 portion of Env CTXLP.

[00311] NRS is an N-linked glycosylation site, RSKR is the furin cleavage site, KRQK is the nuclear localization sequence (NLS). NRS may mask furin site allowing for NLS function.

[00312] **ERVK Env (Based on ERVK113): 79.2 kDa [SEQ ID NO. 234]**

[00313] MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPT WAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAANYTYWAYVPFP PLIRAVTWMDNPIEIYVNDSVWVPGPTDDCCPAKPEEEGMMINISIGYRYPPICLGRAPGCL MPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIPK ESKNTEVLVWEECVANSAVILQNNEFGTLIDWAPRGQFYHNCSGQTQSCPSAQVSPAVDS DLTESLDKHKHKKLQSFYPWEWGEKGISTARPKIISPVSGPEHPELWRLTVASHHIRIWSGN QTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTFN WQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLN**RSKR**FIFTLIAVIMGLIAVTAT AAVAGVALHSSVQSVNFVNDWQNNSTRLWNSQSSIDQK**LANQINDLRQTVIW**MGDRLMSL EHRFQLQCDWNTSDFCITPQIYNESEHHWDMVRCHLQGREDNLTLDISKLKEQIFEASKAHL NLVPGTEAIAGVADGLANLNTVTWVKTIGSTTIINLILILVCLFCLLLVYRCTQQLRRDSDHRER AMMTMVVLSKRKGGNVGKSKRDQIVTVSV*

[00314] →aa residues 1-465 (furin cleavage site- R-X-K/R-R↓) of ERVK 113 (19p12b).

[00315] →aa residues 466-699 (furin cleavage site to end of aa sequence) of ERVK 113 (19p12b). Bolded portion of sequence represent the ISU domain of ERVK TM.

[00316] **ERVK SU (Based on ERVK113): 52.86 kDa [SEQ ID NO: 3]**

[00317] MNPSEMQRKAPPRRRRHRNRAPLTHKMNKMVTSEEQMKLPSTKKAEPPT WAQLKKLTQLATKYLENTKVTQTPESMLLAALMIVSMVVSLPMPAGAAAANYTYWAYVPFP PLIRAVTWMDNPIEIYVNDSVWVPGPTDDCCPAKPEEEGMMINISIGYRYPPICLGRAPGCL MPAVQNWLVEVPTVSPISRFTYHMVSGMSLRPRVNYLQDFSYQRSLKFRPKGKPCPKEIP KESKNTEVLVWEECVANSAVILQNNEFGTLIDWAPRGQFYHNCSGQTQSCPSAQVSPAVD SDLTESLDKHKHKKLQSFYPWEWGEKGISTARPKIISPVSGPEHPELWRLTVASHHIRIWSG NQTLETRDRKPFYTIDLNSSLTVPLQSCVKPPYMLVVGNIVIKPDSQTITCENCRLLTCIDSTF NWQHRILLVRAREGVWIPVSMDRPWEASPSVHILTEVLKGVLNRSKR*

[00318] →aa residues 1-465 (furin cleavage site- R-X-K/R-R↓) of ERVK 113 (19p12b).

[00319] CTXLP biological characterization

[00320] **The ERVK CTXLP domain is found in several distinct proteinaceous forms**

[00321] *Western blot analysis of ERVK-expressing NCCIT cells*

[00322] Whole cell extract of ERVK-expressing NCCIT cells and an immunoprecipitated (IP) fraction enriched for CTXLP were analyzed by Western blot (**Figure 25**). In the whole cell extract, major bands were identified at 90 and 110 kDa, suggesting this may be the predominant form of CTXLP found in the cell. Unmodified CTXLP would be expected to correspond to a 51 kDa band; thus, is it possible that heavier bands are due to PTM such as glycosylation, phosphorylation or sumoylation. It should be noted that higher molecular weight bands are also observed in other cell types (see below). It is likely that these heavier bands reflect PTM including phosphorylation (≈ 2 kDa) and glycosylation (≥ 2 kDa) (**Figures 21 and 26**) or sumoylation. The possibility that the higher molecular weights reflect post-translational modification of the 32 and 51 kDa CTXLP isoforms through glycosylation, would be in accordance with the results of the bioinformatic analysis which indicate that ERVK Env and CTXLP proteins are heavily glycosylated (**Figure 21, Table 6, Figure 26**)[15,60,61]. Another possibility for observing larger than expected protein bands is detergent-resistant multimerization, as retroviral envelope proteins form trimers[15]. Light bands were also observed at 51 kDa and 32 kDa, the latter form possibly due to an alternative start site within or cleavage of CTXLP protein.

[00323] In the CTXLP-enriched fraction, the most heavily enriched band was detected at 51 kDa. In this fraction, CTXLP-reactive bands were also found at 110 and 142 kDa, which are also possible results of protein PTM or multimerization. Lighter bands were found at 26 and 29 kDa, again these bands suggest CTXLP cleavage or alternative methionine start site products.

[00324] **ERVK CTXLP is inducible through the action of pro-inflammatory signalling**

[00325] *Astroctye expression of ERVK CTXLP in the presence of pro-inflammatory agents*

[00326] NCCIT, used as control cells, were cultured along side- SVGA cells. In addition to higher molecular weight bands (not shown) NCCIT cells demonstrated a 51 kDa band and 32 kDa band. Unlike NCCIT cells, in the astrocytic cell line SVGA ERVK CTXLP protein is spontaneously expressed at low levels, with a minor 32 kDa protein being apparent (Figure 25). However, CTXLP expression in SVGA cells was upregulated upon treatment with pro-inflammatory cytokines tumor neurosis factor (TNFα) and LIGHT (lymphotoxin-like inducible member of the TNF superfamily protein) (**Figure 27**). In contrast to CTXLP expression, Env protein expression was not induced by pro-inflammatory stimulus.

[00327] As with NCCIT cells, larger 90/110 kDa CTXLP reactive bands were also observed upon TNFα or LIGHT treated SVGA andReNcell neurons (**Figures 28, 30, 56 and 58**). As detailed above, these larger bands may represent post-translational modification (PTM) of CTXLP isoforms.

[00328] **The localization of ERVK CTXLP expression is cell type and inflammation dependent**

[00329] *Ubiquitous expression of ERVK CTXLP expression in NCCIT cells*

[00330] The localization of CTXLP protein expression was examined through Western blot analysis of cellular fractions and confocal imaging. In NCCIT cells, endogenous CTXLP protein appeared ubiquitously expressed and localizes to the cytoplasm, nucleus and chromatin enriched fractions. CTXLP protein was also found in both the soluble and insoluble NCCIT whole cell lysates (**Figure 29 A and B**). The latter further suggests that there is interaction of CTXLP with cell membranes. The ubiquitous nature of CTXLP was supported by confocal imaging (Figure 27 C).

[00331] CTLXP expression was also associated with indicators of autophagy in NCCIT cells (**Figure 29B**). Autophagy occurs when dysfunctional proteins and damaged organelles accumulate within a cell, and has been associated with cell dysfunction in neurodegenerative disease[62]. NCCIT cells exhibited insoluble caspase-3, which is an indicator for autophagy[63]. There was also evidence of LC3B cleavage (data not shown) which occurs during autophagy.

[00332] *Nuclear Localization of ERVK CTXLP expression in SVGA cells*

[00333] In contrast to NCCIT cells, in SVGA cells endogenous CTXLP protein localization occurred predominantly in the chromatin cellular fraction (**Figure 29A**) and was poorly detected by confocal imaging (**Figure 29C**). However, CTXLP protein levels were strongly elevated in the nucleus upon treatment of 0.1 and 1 ng/mL TNFα when compared to cells alone (**Figure 30**). This increase in CTXLP expression upon treatment of astroctyes with low levels of TNFα or LIGHT, suggests that chronic, low level inflammation may augment CTXLP levels physiologically. Moreover, with pro-inflammatory cytokine exposure, CTXLP puncta formed within the nucleus, but staining was excluded from the nucleoli (**Figure 31**). Higher resolution images of CTXLP expressing astrocytes showed that CTXLP puncta also formed in the cytoplasm and on the cell surface, suggesting that potential isoforms of CTXLP may have location-specific functions (**Figure 31 B**).

[00334] Enhanced CTXLP expression was also associated with enhanced cytoplasmic RT expression (**Figure 30**), suggesting regulation of global ERVK gene expression may be

linked. Pro-inflammatory cytokines have been shown to induce ERVK expression[64-67]. During inflammatory events, ERVK Env has been shown to be either neuroprotective[68] and neuropathological[69]. Indeed, chronic exposure to TNFα may facilitate frameshifting in ERVK *env* translation leading to CTXLP being produced as a cryptic protein. TNFα expression and its subsequent downstream signalling cascade products may result in enhanced IRES-dependant translation[70], promoting the formation of the smaller CTXLP isoform. Nonetheless, the observation that ERVK CTXLP did not co-localize with ERVK Env protein which localized (**Figure 31**), suggesting that these proteins may have different localization sequences and/or patterns. Thus, the relationship between CTXLP and Env in ERVK gene regulation remain unclear.

[00335] Overexpression of the CTXLP cysteine-rich domain in SVGA cells also resulted in increased CTXLP 32 kDa and 90/110 kDa protein bands (**Figure 32**), suggesting that this protein domain is sufficient to enhance its own expression.

[00336] **ERVK CTXLP binds chromatin**

[00337] Consistent with our observation that CTXLP is enriched in the chromatin fraction (**Figure 29**), DNABIND prediction (http://dnabind.szialab.org/)[71] predicts that CTXLP binds DNA. Prediction parameters were as follows: false positive rate of 6%, expected sensitivity of 58.7%, expected Matthews correlation coefficient of 0.55, score threshold is set to 1.577 (threshold probability of 0.8288). Prediction results of the CTXLP sequence were a score of 1.771 and a probability of DNA binding of 0.8546.

[00338] An analysis is DNA interactions revealed that CTXLP bound the interferon response elements (ISREs) within the ERVK viral promoter (5' LTR) (**Figure 33**). This binding is enhanced in the presence of inflammatory cytokine stimuli, and is distinct in select cell types (**Figure 33**). This suggests that CTXLP, and specifically the 32kDa form of CTXLP (Figure 24), may bind DNA and regulate gene expression. Indeed, preliminary data suggest that CTXLP can alter both ERVK expression (**Figure 32**) and the transcription of NF-κB transcripts (see below). Further, CTXLP may alter the gene expression patterns of numerous viral and cellular genes containing an ISRE elements in their promoters[72,73]. Thus, CTXLP may regulate ERVK gene expression, as well as other genes containing ISREs.

[00339] In summary, CTXLP protein isoform expression in NCCIT and SVGA cells was elucidated by Western blots which indicated presumed isoform sizes of 32 kDa, 51 kDa, and 90/110 kDa. In NCCIT cells, endogenous CTXLP is ubiquitously expressed being present in the nucleus and also identified in the cytoplasm and cell membrane, based on cell fractionation and

confocal experiments. In contrast, in SVGA cells basal CTXLP levels are limited, but highly inducible by pro-inflammatory stimuli. In addition, CTXP expression in almost exclusively in the chromatin fraction and demonstrates a prominence in the nucleus upon confocal imaging. The notable exception is that after pro-inflammatory activation for 24 hours CTXLP puncta appear in the cytoplasm and on cellular membranes reminiscent of pathogenic protein aggregates. Moreover, the localization pattern in response to pro-inflammatory activators resulting in a prominence in the nucleus (**Figures 29-31**), ability to bind chromatin (**Figures 29 and 33**) and absence from the nucleoli (**Figure 31**) suggests that CTXLP may be involved in viral transcription. A primary candidate as a viral transcription factor is the 32 kDa CTXLP isoform, as small cysteine-rich proteins have previously been identified as transcriptional activators[74,75], as per HIV-1 Tat (15 kDa) and HTLV Tax (40 kDa) role as viral transcription co-activators[76,77]. Additionally, low basal CTXLP staining in non-diseased cells suggests that it might have a role in normal physiology and gene regulation processes.

[00340] CTXLP Expression in Disease States

[00341] CTXLP is expressed *in vivo* in humans

[00342] *RNAseq analysis of CTXLP+ transcripts in disease states*

[00343] To evaluate the significance of CTXLP in disease, we evaluated the expression of CTXLP encoding ERVK loci in publicly available RNA-Seq datasets in the Sequence Read Archive (SRA) (**Table 10**). This analysis is summarized in **Table 11 and Figures 34-37**. These loci were identified by searching the SRA by disease affiliation and then evaluating each potential study based on samples size, tissue and sequencing quality. Preference was given to studies with large sample sizes, autologous controls, *ex vivo* disease-relevant tissue, and high sequencing quality. Paired-end reads were preferred to single-end. We focused on studies with fewer measures selecting for particular RNA sub-populations, which could have depleted ERVK RNA from the input.

Table 10: RNA-Seq datasets in the Sequence Read Archive (SRA) used for the analysis of ERVK expression.

Characteristics of RNA-Seq Studies

Study	Condition	Tissue	Instrument	Strategy	Source	Selection	Layout	Read Length
SRP090259	schizophrenia	dorsolateral prefrontal cortex	AB SOLiD 4 System	RNA-Seq	transcriptomic	cDNA	single	50 bp
SRP074904	bipolar disorder	putamen or caudate nucleus	Illumina HiSeq 2000	RNA-Seq	transcriptomic	cDNA	single	99 bp
SRP110016	multiple sclerosis	optic chiasm	Illumina HiSeq 3000	RNA-Seq	transcriptomic	cDNA	single	50 bp
SRP102685	rheumatoid arthritis	synovium	Illumina HiSeq 2000	RNA-Seq	transcriptomic	cDNA	paired	101/99 bp
SRP068424	HIV^+/HCV^+	CD4+ T-cells	Illumina HiSeq 2500	RNA-Seq	transcriptomic	random PCR	paired	100/100 bp

[00344] The studies examine included breast cancer, prostate cancer, Amyotrophic Lateral Sclerosis (ALS), Multiple Sclerosis (MS), Rheumatoid Arthritis (RA), schizophrenia, bipolar disorder and HIV+/HCV+ infection. ERVK expression by HML group is summarized in **Table 11**. We found that the overall expression of ERVK in these disease states was low, and all HML groups were expressed. ERVK expression was highest in blood and cancerous tissue. In addition, we found loci with significantly different expression between patients and controls, but these were different for each study. Breast cancer, prostate cancer, and Multiple Sclerosis datasets contained expression patterns which could potentially distinguish patients from controls. These patterns were driven by differences in expression of CTXLP- loci and loci with inactivating Env mutations.

Table 11: ERVK expression by HML group in RNAseq datasets

study	hml	Minimum	Median	Mean	SD	Maximum
	HML1	0	0.0125001	0.1179662	0.3213379	11.465101
	HML10	0	0.0389112	0.1801432	0.3599358	5.855237
	HML2	0	0.0214437	1.0386104	28.9656795	2800.179981
	HML3	0	0.0085536	0.0772390	0.1934535	5.249185
	HML4	0	0.0274862	0.0863698	0.1811918	3.655131
	HML5	0	0.0000000	0.0859580	0.2350117	4.212269
	HML6	0	0.0000000	0.1133778	0.3016479	4.628222
SRP090259	HML7	0	0.0128447	0.0602345	0.1412095	3.037872
	HML8	0	0.0193117	0.2915583	11.7006947	1289.465670
	K14C	0	0.0000000	0.0542510	0.1667601	3.168318
	LTR22B1#LTR/ERVK	0	0.0000000	0.1161527	0.3674700	8.431398
	LTR22B2#LTR/ERVK	0	0.0000000	0.1241151	0.3284346	2.537287
	LTR22C2#LTR/ERVK	0	0.0000000	0.0826217	0.2326312	2.358517
	LTR22E#LTR/ERVK	0	0.0000000	0.1423927	0.3607523	3.823776
	LTR3B_#LTR/ERVK	0	0.0000000	0.0855171	0.2227743	3.975364
	HML1	0	0.0000000	0.1061961	0.3018500	3.824492
	HML10	0	0.0000000	0.1789209	0.4687300	7.527670
	HML2	0	0.0072385	0.1691623	1.8123759	197.784008
	HML3	0	0.0000000	0.0917908	0.3355194	12.499732
	HML4	0	0.0000000	0.1252037	0.2851307	5.880441
	HML5	0	0.0000000	0.0909087	0.3553731	9.824726
	HML6	0	0.0000000	0.1070234	0.2950223	3.474829
SRP074904	HML7	0	0.0000000	0.0786869	0.2005897	2.337186
	HML8	0	0.0000000	1.0845542	56.4008441	5752.139071
	K14C	0	0.0000000	0.0566177	0.1889185	2.725355
	LTR22B1#LTR/ERVK	0	0.0000000	0.0929150	0.2638750	4.531172
	LTR22B2#LTR/ERVK	0	0.0000000	0.1643443	0.5538800	6.122761
	LTR22C2#LTR/ERVK	0	0.0000000	0.0697017	0.2267312	3.086240
	LTR22E#LTR/ERVK	0	0.0000000	0.1307125	0.3980359	4.137923
	LTR3B_#LTR/ERVK	0	0.0000000	0.1079325	0.4247868	11.348338
SRP110016	HML1	0	0.0000000	0.0939917	0.4084061	6.709148
	HML10	0	0.0246443	0.1977949	0.6106722	8.991545
	HML2	0	0.0748145	0.1923879	0.7013925	25.584209
	HML3	0	0.0207453	0.0850121	0.4052984	14.993567
	HML4	0	0.0676325	0.2055983	0.5833871	17.361348
	HML5	0	0.0000000	0.0627547	0.2304348	4.292697
	HML6	0	0.0000000	0.1176631	0.3652445	4.943406
	HML7	0	0.0165461	0.0571758	0.1534795	2.665017

	HML8	0	0.0237677	0.0779974	0.5586158	31.923344
	K14C	0	0.0000000	0.0319094	0.1170643	1.377055
	LTR22B1#LTR/ERVK	0	0.0000000	0.0744216	0.2581293	2.729679
	LTR22B2#LTR/ERVK	0	0.0000000	0.1462319	0.6989967	4.712645
	LTR22C2#LTR/ERVK	0	0.0000000	0.0341152	0.1244866	1.373369
	LTR22E#LTR/ERVK	0	0.0000000	0.0483235	0.1773740	1.849842
	LTR3B_#LTR/ERVK	0	0.0000000	0.0478708	0.1685803	2.643292
	HML1	0	0.0000000	0.1268767	0.3353742	4.052137
	HML10	0	0.0304281	0.1950971	0.3455298	2.487003
	HML2	0	0.0229832	0.1938948	1.8896213	219.947046
	HML3	0	0.0000000	0.1223329	0.3815171	10.135958
	HML4	0	0.0237879	0.1731327	0.3562414	5.434816
	HML5	0	0.0000000	0.1089093	0.3043274	7.381214
	HML6	0	0.0000000	0.1585309	0.3921706	4.257929
SRP102685	HML7	0	0.0000000	0.0912298	0.2878545	7.654548
	HML8	0	0.0081154	2.0901303	98.5146905	8515.001781
	K14C	0	0.0000000	0.0706536	0.2391837	3.356916
	LTR22B1#LTR/ERVK	0	0.0000000	0.1234718	0.3168124	5.835195
	LTR22B2#LTR/ERVK	0	0.0000000	0.1937394	0.7134805	5.204694
	LTR22C2#LTR/ERVK	0	0.0000000	0.0902775	0.2336134	2.321173
	LTR22E#LTR/ERVK	0	0.0000000	0.1764475	0.4852693	4.627528
	LTR3B_#LTR/ERVK	0	0.0000000	0.1258778	0.3518004	4.700366
	HML1	0	0.0000000	0.1546103	0.9128325	32.738476
	HML10	0	0.0000000	0.1910084	0.5156409	8.690614
	HML2	0	0.0000000	0.2408015	2.7415319	195.191293
	HML3	0	0.0000000	0.1016756	0.7212861	32.342361
	HML4	0	0.0000000	0.3064975	1.5651906	56.983694
	HML5	0	0.0000000	0.1737037	0.9668251	22.605106
	HML6	0	0.0000000	0.1513285	0.7387770	15.850478
SRP068424	HML7	0	0.0000000	0.0647419	0.2488850	5.277685
	HML8	0	0.0000000	0.2023157	6.0970185	705.290157
	K14C	0	0.0000000	0.0794524	0.4541036	8.772424
	LTR22B1#LTR/ERVK	0	0.0000000	0.1325097	0.6447649	13.163549
	LTR22B2#LTR/ERVK	0	0.0000000	0.0693838	0.2707259	2.985159
	LTR22C2#LTR/ERVK	0	0.0000000	0.0440100	0.1820465	2.502976
	LTR22E#LTR/ERVK	0	0.0000000	0.0989893	0.3236325	4.761848
	LTR3B_#LTR/ERVK	0	0.0000000	0.1085489	0.5031198	8.617390
SRP064478	HML1	0	0.0000000	0.0601878	0.1912765	3.648491
	HML10	0	0.0046214	0.1085939	0.2846323	3.347339

	HML2	0	0.0114172	0.1068918	1.0920575	60.595378
	HML3	0	0.0000000	0.0663053	0.2855648	11.555712
	HML4	0	0.0122422	0.0818699	0.2004683	2.824911
	HML5	0	0.0000000	0.0544799	0.1695744	3.303590
	HML6	0	0.0000000	0.0687535	0.1899200	1.844958
	HML7	0	0.0000000	0.0453627	0.1214706	1.274142
	HML8	0	0.0000000	1.6252062	86.0880409	9238.633200
	K14C	0	0.0000000	0.0356487	0.1248000	2.091446
	LTR22B1#LTR/ERVK	0	0.0000000	0.0603075	0.1706618	3.324758
	LTR22B2#LTR/ERVK	0	0.0000000	0.0890397	0.3238018	2.795000
	LTR22C2#LTR/ERVK	0	0.0000000	0.0532044	0.1480141	1.725137
	LTR22E#LTR/ERVK	0	0.0000000	0.0890193	0.2572100	2.633204
	LTR3B_#LTR/ERVK	0	0.0000000	0.0548118	0.1628587	2.074693
	HML1	0	0.0000000	0.1099929	0.5374086	19.771974
	HML10	0	0.0000000	0.3120755	4.0788949	209.074420
	HML2	0	0.0000000	0.3083341	7.0141530	716.132793
	HML3	0	0.0000000	0.0819258	0.5721171	44.512233
	HML4	0	0.0055612	0.1631802	0.5669411	15.160832
	HML5	0	0.0000000	0.0740519	0.2991551	8.570903
	HML6	0	0.0000000	0.1508548	0.7572169	35.954051
SRP058722	HML7	0	0.0000000	0.0721502	0.2653196	7.406958
	HML8	0	0.0000000	0.1347603	5.6620404	1058.007060
	K14C	0	0.0000000	0.0942404	1.3013323	77.419996
	LTR22B1#LTR/ERVK	0	0.0000000	0.1381469	1.3368361	54.714245
	LTR22B2#LTR/ERVK	0	0.0000000	0.1202672	0.6000457	6.857656
	LTR22C2#LTR/ERVK	0	0.0000000	0.0524224	0.3472964	9.886468
	LTR22E#LTR/ERVK	0	0.0000000	0.1179183	0.5986011	19.318676
	LTR3B_#LTR/ERVK	0	0.0000000	0.0871452	0.4722096	17.621088
ERP000550	HML1	0	0.0000000	0.0913613	0.6721108	25.754908
	HML10	0	0.0000000	0.5265347	3.7211529	104.449407
	HML2	0	0.0000000	0.3829747	9.4737435	494.573463
	HML3	0	0.0000000	0.0758083	0.6113529	24.861316
	HML4	0	0.0000000	0.1226314	0.5223547	39.400311
	HML5	0	0.0000000	0.0545070	0.3573734	15.142432
	HML6	0	0.0000000	0.1665188	1.1979633	46.109402
	HML7	0	0.0000000	0.0653212	0.5217105	11.243053
	HML8	0	0.0000000	0.0792274	2.3560620	344.064632
	K14C	0	0.0000000	0.0208334	0.0993372	1.730934
	LTR22B1#LTR/ERVK	0	0.0000000	0.0496653	0.2114473	3.282185

LTR22B2#LTR/ERVK	0	0.0000000	0.1243072	0.6755299	6.645951
LTR22C2#LTR/ERVK	0	0.0000000	0.0542908	0.3050922	6.011783
LTR22E#LTR/ERVK	0	0.0000000	0.0386306	0.1391434	2.082330
LTR3B_#LTR/ERVK	0	0.0000000	0.0649027	0.4006698	8.956652

[00345] The lack of differential total RNA expression in controls versus the ALS cohort (**Figure 37**), which is intriguing given data from protein immunostaining showing obvious differences between clinical groups (see below, **Figures 38-42**). PCA analysis reveals that expression of select ERVK CTXLP+ loci cluster in controls versus the ALS cohort (**Figure 37**), suggesting that specific CTXLP loci may drive the expression of CTXLP protein in ALS.

[00346] *ERVK CTXLP expression is enhanced in CNS tissues from patients with Amyotrophic Lateral Sclerosis (ALS)*

[00347] ALS pathology involves degeneration of upper (brain) and lower (spinal cord) motor neurons, leading to muscle weakness and paralysis (reviewed in [78-80]). Brain and spinal cord inflammation is a hallmark of ALS (reviewed in [81,82]). The majority of ALS cases are sporadic, and the cause of this disease remains unknown. Here, we focus on the connection between neuropathology associated with ALS and ERVK CTXLP, such as proteinopathy[83,84], aberrant calcium signalling[85], demyelination[86], and oligodendrocyte dysfunction[87].

[00348] To show that CTXLP protein is not only expressed in *in vitro* cell cultures, but also in *ex vivo* (autopsy) human tissues, spinal cord and brain tissues from neuro-normal controls and patients with ALS were assayed for CTXLP by western blot (**Figure 38**) and confocal microscopy (**Figures 39-42**). Western blot analysis of motor cortex specimens from neuronormal controls and patients with ALS reveals significantly enhanced CTXLP expression in ALS (**Figure 38A**, $p<0.05$). CTXLP was concomitantly expressed with inflammation and tissue injury marker CX3CL1 (**Figure 38B**)[88]. Analysis of cervical spinal cord tissues also demonstrates elevated CTXLP and CX3CL1 expression in ALS as compared to controls, alongside a modest decrease in levels of voltage-gated calcium channel CaV2.2 in ALS (**Figure 38, C-F**). Together, these results point to tissue injury and inflammation in CTXLP+ tissues from patients with ALS.

[00349] In addition, confocal microscopy of cervical spinal cord (**Figure 39A**) and motor cortex specimens (**Figure 39B**) from neuro-normal controls and patients with ALS reveals substantially enhanced CTXLP expression in ALS. In the motor cortex, CTXLP+ cells were neurons (based on MAP2 neuronal marker). This is consistent with previous observations of ERVK proteins present in the motor cortex of patients with ALS[67,89,90]. Notably, basal CTXLP

expression was mostly nuclear in neuronormal tissues, whereas CTXLP exhibited a pattern of cytoplasmic aggregation in motor cortex tissues from patients with ALS (**Figure 39B**). Enhanced MAP2 staining in the axon hillock of CTXLP+ pyramidal neurons may be an indicator of virus activity, as seen during rabies infection[91,92]. This pattern of CTXLP expression coincided with a notable decrease in CaV2.2 expression in ALS as compared to controls. Based on staining pattern, this decrease may represent a loss of CaV2.2 expressing pyramidal neurons, as well as smaller CaV2.2+ cells[93]. Remarkably, CTXLP patterning in the cervical spinal cord exhibited a ring pattern surrounding MAP2+ neurons (MAP2 marks neuronal axons in grey, **Figure 39A**, far right panel).

[00350] Our evidence further indicates that CTXLP can alter oligodendrocyte behavior. In the CNS, highly specialized cells called oligodendrocytes protect neuronal axons by wrapping them in an extensive plasma membrane compacted to produce the myelin sheath[94]. Oligodendrocyte precursor cells (OPCs) are a pool of immature oligodendrocytes, which express characteristic markers such TCF4, Olig1 and Olig2 [95,96]. Upon differentiation into mature oligodendrocytes, they begin to express myelin proteins such as PLP, MOG and MAG[95]. Oligodendrocytes must myelinate early post-differentiation and myelination occurs within a short timeframe (12-18 hours), where their extended processes ensheath 50-60 axonal segments simultaneously[97]. Some CNS regions (spinal cord, brainstem and visual cortex) exhibit early myelination during human development, whereas other regions undergo myelination into adulthood (prefrontal cortex and association fibers). Pools of OPCs can remain in tissues and are capable of migration and later differentiation into mature oligodendrocytes, often in response to brain injury[98]. However, in many disease states, an attempt at remyelination is most often unsuccessful[98]. A prevailing theory surrounding defects in remyelination is that despite increased numbers of OPCs in injured tissue, these precursor cells become stalled in an immature state and fail to properly differentiate into mature oligodendrocytes[96,98]. Alterations in OPC markers, such as enhanced TCF4 and Olig1 occurs in tissue lesions from patients with MS[99,100].

[00351] Our observations show that CTXLP expression occurs in either lateral and/or anterior cortical spinal tracts in ALS (**Figure 40**). Strong CTXLP+ staining coincides with demyelinating lesions, as shown by solochrome cyanine staining of adjacent tissues (**Figure 40**). Increased TCF4 (oligodendrocyte precursor marker) is also evident in association with CTXLP expression in tissue from patients with ALS (**Figure 40**).

[00352] **Figure 41** depicts increased numbers of TCF4+ and Olig1+ cells expressing CTXLP in cervical spinal cord tissue from patients with ALS, as compared to controls. This indicates that CTXLP expression in the spinal cord of patients with ALS does indeed occur in oligodendrocytes (**Figures 39-41**), specifically in cells expressing OPC markers.

[00353] Neurite outgrowth inhibitor (Nogo-A) is a key regulator of oligodendrocyte precursor cell (OPC) differentiation; when OPCs express Nogo-A they are unable to progress towards a mature oligodendrocyte phenotype, which is capable of myelination[101,102]. Thus, enhanced expression of Nogo-A in OPCs in the context of inflammation and disease states prevents axonal regeneration by restricting OPC maturation[103-105]. As an example, demyelinated MS lesions show an increased abundance of Nogo-A+ OPCs, yet the inability of OPCs to mature is proposed as the mechanism driving a non-permissive environment leading to remyelination failure[103,106,107]. In mature oligodendrocytes, Nogo-A expression prevents axonal sprouting and is expressed in these cells until the initiation of active myelination.

[00354] Nogo-A is implicated in a variety of neurological conditions, such as spinal cord injury, peripheral neuropathies, stroke, temporal lobe epilepsy, Alzheimer's disease, ALS, MS and schizophrenia[101,108-110]. Nogo-A has been identified as a prognostic marker and therapeutic target in ALS due to its substantial expression in muscle tissue from patients with motor neuron disease[111,112]. Mechanistically, Nogo-A expression destabilizes neuromuscular junctions[113-116]. Indeed, clinical trials using human anti-Nogo-A antibodies have been performed (ATI 355 from Novartis Pharma and Ozanezumab and GSK1223249 from GlaxoSmithKline)[101,117,118]. These therapies were designed to target Nogo-A expression in the periphery (intravenous infusions), but may fail to block Nogo-A expression in the CNS (**Figure 42**), thus explaining the negative results in Phase II clinical ALS trials with Ozanezumab[119,120]. Of note, genetic polymorphism reticulon 4 receptor (RTN4R) gene encoding the Nogo-A receptor (NgR1), is associated with sporadic ALS[121].

[00355] **Figure 42** demonstrates that CTXLP expression in the spinal cord of patients with ALS is associated with elevated Nogo-A expression, specifically in OPCs and other cell types. This specifically occurs in areas of myelin depletion (see **Figure 41**). It has been demonstrated in human spinal cord, that select myelin protein rings (PLP, MOG, but not MAG) are detectable by immunohistochemistry even 3 years after injury in degenerating fibre tracts exhibiting the absence of intact axons[122]. Nogo-A expression also persists in degenerating spinal tissue and may create a non-permissive environment for axon regeneration[122]. Furthermore, it has been shown that Nogo-A favours a pro-inflammatory context[123], one that

would promote ERVK expression via modulation of NF-κB and pro-inflammatory cytokine secretion[67].

[00356] *ERVK CTXLP expression is enhanced in cancer cells*

[00357] To further evaluate the potential pathogenic activity of CTXLP, we examined CTXLP levels in cancer to follow-up on our observation that NCCIT human embryonic carcinoma line spontaneously expressed CTXLP. The localization pattern that included the cytoplasm also suggested that this represented a stage in course of aberrant CTXLP expression. Thus, we assayed prototypic teratocarcinoma (NCCIT) and breast cancer cells (T47D) for CTXLP expression as compared to the karyotypically normal, non-cancerous, cell lines astrocytic SVGA cells (**Figure 43**). Cancer cells clearly show higher levels of ERVK CTXLP as compared to non-cancerous cells. A cancer screen also reveals several cancers exhibiting enhanced CTXLP levels, including T cell lymphoma, Acute T-cell leukemia, epithelioid carcinoma, Burkitt's lymphoma, neuroepithelioma, prostate, breast, ovary, testis and skin cancers (**Figure 44**).

[00358] In summary, ERVK CTXLP localized to the motor cortex and spinal cord sections from autopsy samples of patients with ALS, but not neuronormal controls. Concomitantly, CTXLP expression was substantially enhanced in diseased ALS tissues aligning with oligodendrocytes, Nogo-A expression and demyelinated lesions. In addition, cancer cell lines and tissue expressed greater levels of CTXLP relative to normal controls. Together, these findings provide significant evidence for the activity of CTXLP in ALS and certain cancers.

[00359] Pathological consequences of CTXLP expression

[00360] ERVK CTXLP enhances NF-κB protein expression, whereas ERVK Env does not

[00361] *Real Time PCR analysis of transfected 293T cells*

[00362] To investigate how cells may react to the expression of CTXLP and SU, RT-PCR analysis was used to measure the expression of the pro-inflammatory NF-κB p65 subunit and the anti-viral response protein IRF7. This analysis showed that both CTXLP and SU triggered a marked increase in the mRNA expression of NF-κB p65. Conversely, neither protein was able to trigger an upregulation of IRF7 (**Figure 45**). The upregulation of NF-κB p65 may be beneficial to the ERVK provirus as it is able to 1) act as a direct transcriptional activator of the ERVK LTR, and 2) trigger inflammatory conditions that are conducive to ERVK activation[64,67].

[00363] *Confocal microscopy analysis of transfected 293T cells*

[00364] SU and CTXLP transfected 293T cells were also stained to determine whether the presence of either of these proteins was sufficient to trigger the expression of NF-κB p65. It was observed that CTXLP was able to trigger NF-κB p65 protein expression, whereas SU was unable (**Figure 46**). This was in stark contrast to the RT-PCR results in **Figure 45**, where both ERVK SU and CTXLP induced NF-κB p65 transcription. As expected, there was a marked increase in SU expression in CTXLP-expressing cells, as the epitope for the SU antibody binds to either SU or CTXLP as both proteins contain the SU amino acid sequence.

[00365] To further confirm whether the effect of NF-κB induction by CTXLP is a general phenomenon occurring the multiple cell types, astrocytic SVGA cells were also transfected as described above and evaluated for NF-κB protein expression. Interestingly, both NF-κB p65 and p50 proteins were induced by CTXLP, but not ERVK SU overexpression (n=4). This finding is notable, considering we have shown that ERVK transcription is mediated by IRF1, p50 and p65 transcription factors, and impacts ERVK expression in ALS[67]. It is also intriguing considering TRAF proteins were predicted to be interacting partners of CTXLP, and may alter NF-κB signalling[124,125].

[00366] **ERVK CTXLP depletes CaV2.2 CCAT protein expression**

[00367] A surprizing feature of several voltage-gated calcium channels (VGCCs) is the ability of their C-terminal fragments to translocate to the nucleus and impact gene expression. Termed calcium channel-associated transcription regulator (CCAT) by Gomez-Ospina et al. in 2006[126], these novel gene products encoded within the VGCC sequences. In most cases, an antibody targeting a C-terminal CACNA1 epitope will identify an approximately 75 kDa CCAT fragment with a cellular distribution within the nucleus (or nuclear fractions), unlike the intact channel protein localized to the cytoplasm and membrane[126,127]. VGCC CCAT proteins can be regulated by cell signalling events. For example, cellular signals that promote CaV1.2 CCAT nuclear localization include treatment of neurons with 2.5mM EGTA (a chelator which reduces free extracellular calcium), whereas 65mM KCl treatment (mimicking tonic activity of VGCC) decreased nuclear CCAT levels[126]. Several signals which drive high intracellular calcium levels in neurons, including 100μM glutamate, depolarization and NMDA signalling, lead to decreased nuclear CCAT levels[126].

[00368] We have previously demonstrated an inverse correlation between CTXLP and voltage-gated calcium channel CaV2.2 expression in ALS brain and spinal cord tissues (**Figures 38 and 39**). To further extend this observation, we performed *in vitro* experiments of

CTXLP and ERVK SU exposure by overlaying immunoprecipitation products on human astrocytes (**Figure 47**). Treatment of SVGA cells with CTXLP, but not ERVK SU, resulted in the depletion of Cav2.2 CCAT (75 kDa), as measured by western blot analysis and confocal microscopy. A decreased in the full size CaV2.2 channel (220 kDa) was also observed (n=2, data not shown). Thus, CTXLP in the CNS may lead to an overall decrease in CaV2.2 channel and CaV2.2 CCAT expression, thus explaining the observed decreased expression of CaV2.2 in ALS (**Figures 38 and 39**). The regulatory role of CaV2.2 on cellular transcription is currently unknown.

[00369] ERVK CTXLP is toxic via mechanisms that differ from ERVK Env-mediated toxicity

[00370] *Treatment of SVGA astrocyte cells with ERVK Env proteins isolated via immunoprecipitation*

To determine the neurotoxicity of CTXLP, SVGA astrocyte cells were treated with CTXLP proteins isolated from NCCIT cells via immunoprecipitation (IP). This simulates conditions wherein CTXLP would enter the cell from the outside and possibly exert its effects by binding to cell surface receptors (such as calcium channels). There was considerable variation in the neurotoxicity assays, which may be due in part to the fact that cells were dosed by volume of CTXLP. There was no reliable way to measure the concentration of the protein in the IP product, as protein concentration was well below sensitivity of our in-house BCA assay (20 µg/ml). However, a much higher number of cells treated with CTXLP expressed caspase-3 (apoptosis marker used in the toxicity assays) than control cells demonstrating that CTXLP was toxic to astrocytes, even at unmeasurably low concentrations (**Figure 48**).

[00371] A separate toxicity assay was performed by treating astrocytes with SU and CTXLP (respectively) in the presence and absence of calcium. Theoretically, if CTXLP does in fact contain an ω-conotoxin domain, by flooding cells with calcium and thus saturating calcium channels, it's ability to exert toxic effects on cells via calcium channel binding should be blocked. This is what was seen in CTXLP, but not SU, toxicity assays. In the presence of calcium, the levels of caspase expression in CTXLP-treated cells were similar to controls conditions. They were also less than those of cells treated with CTXLP in the absence of calcium. Further, cells treated with SU in the presence and absence of calcium expressed similar levels of caspase-3 in comparison to CTXLP alone (Figures 49 and 50).

[00372] Cells in this same toxicity assay were also analyzed at later time points. It was observed that CTXLP and SU-treated cells appeared to be able to continue to replicate

despite high levels of caspase expression. After 8 days, cells in both conditions increased in cell density despite high levels of caspase-3 expression and without the addition of media. These observations suggest that these cells may have been transformed oncogenically[129-131]. Conversely, control cells were not viable after 8 days in culture (**Figure 51**). It is interesting to note that some viruses, such as the influenza virus, require caspase expression to replicate efficiently[132]. Moreover, HIV Tat, the viral transactivator, induces caspase activation as part of its neurotoxic mechanism[38].

[00373] It is notable that the trend of enhanced caspase-3 positivity is seen in both treated (**Figures 48-51**) and transfected cells (**Figure 52**), suggesting that exposure to CTXLP and/or cellular production of CTXLP *in vivo* may be toxic to cells.

[00374] **ERVK CTXLP expression drives morphological changes in cells**

[00375] ERVK CTXLP-treated SVGAs and 293T cells transfected with ERVK CTXLP-encoding plasmids were imaged to observe the cellular morphology. Though many of the cells looked normal in appearance, it was observed that an increased number of the cells produced long filipodia (**Figure 53**). As well, the formation of syncytia (multinucleated cells) was also observed amongst cells exposed to CTXLP. These features are characteristic of retrovirus-infected cells and are known to promote virus transfer between cells[133].

[00376] This data indicates that ERVK CTXLP expression has the capacity to enhance NF-κB p65/p50 and CaV2.2 proteins that play a critical role in ALS pathogenesis. In addition, CTXLP administration or transfection induced significant levels of capase-3. The induction of caspase-3 activation and apoptosis by CTXLP was inhibited by excess extracellular calcium pointing to a calcium channel mediated activation of toxicity. Remarkably, despite the initial die off of cells, cells remaining in the cultures appeared to demonstrate appreciable cellular proliferation relative to control suggesting the induction of a carcinogenic process.

[00377] **ERVK CTXLP can be targeted by small molecule therapeutics**

[00378] Taken together this data strongly indicates that targeting CTXLP would have significant therapeutic value in ALS. To this end, we have began investigating A small molecule inhibitors to capable of counteracting the pathological effects associated with CTXLP expression is of therapeutic value. A drugs screen in ERVK CTXLP-expressing NCCIT cells was performed to evaluate potential efficacity against CTXLP (**Figure 54**). Michael acceptor electrophile (MAE) compounds are known to inhibit HIV Tat-dependent transcription by interfering with thiols in its cysteine-rich domain[134,135]. As CTXLP and HIV Tat share commonality in their cysteine-rich domains (**Figure 14**), we evaluated a series of

MAE compounds including curcumin, rosmarinic acid, gambogic acid and celastrol. Two MAE drugs, celastrol (Cel) and gambogic acid (GA), were identified as suppressing CTXLP expression in NCCIT cells in the low micromolar range (**Figure 54**).

[00379] Derived as an active compound from the Thunder God vine (*Tripterygium wilfordii* Hook F), celastrol (https://pubchem.ncbi.nlm.nih.gov/compound/celastrol) is a plant-derived triterpene with antioxidant, anti-viral and anti-inflammatory activity[134,136,137]. Celastrol is currently used as a therapeutic agent for rheumatoid arthritis (RA) and lupus in China[136,138]. Celastrol has also been shown to impact pathological outcomes and symptoms in animal models of RA[139], as well as inflammatory markers in activated fibroblast-like synoviocytes from patients with rheumatoid arthritis[140]. Celastrol has been shown to limit beta-amyloid pathology and neuronal degeneration in Alzheimer's disease models[141]. Its anti-cancer properties are also under investigation[142].

[00380] Gambogic acid (https://pubchem.ncbi.nlm.nih.gov/compound/16072310) is an active compound from the Gamboge tree (*Garcinia hanburyi*), with antioxidant, anti-viral and anti-inflammatory properties)[134,143-145]. It has been shown to be neuroprotective[146], and inhibit spinal cord injury and inflammation in a rat model[147]. Both celastrol and gambogic acid can prevent mutant huntingtin protein aggregation and its neuronal toxicity[148]. The anti-cancer properties of gambogic acid are also under investigation[144,145].

[00381] Improvements on drug efficacy, toxicity and tissue-targeting are possible by using MAE-derivatives, related compounds (**Table 12**), soluble analogues and nanosystem delivery to the brain[149,150]. Here we provide proof-of-concept that CTXLP is druggable using small molecules (**Figures 54-58**); further drug development may improve upon the anti-CTXLP effects of celastrol and gambogic acid.

Table 12: Compounds with similar structures and/or bioactivities with celastrol and gambogic acid found using PubChem open chemistry database.

Celastrol (CID 122724)	PubChem CID	Gambogic Acid (CID 16072310)	PubChem CID
Celestrol methyl ester	10004662	Gambogic acid	86767267
Salaspermic acid	44593364	Gambogic amide	6710783
Demethylzeylasteral	10322911	Gambogellic Acid	10651612
Wilforol A	10096097	Isomorellin	12313004
Isoiguesterin	11373102	Gambogin	15298998
Tingenone	441687	Isogambogenic acid	15299002
ZXENWDWQTWYUGY-DISOOBLMSA-N	122852	Gambogic acid	15559465
JFACETXYABVHFD-YDUKQFKJSA-N	90488873	Desoxymorellin	16078248
JFACETXYABVHFD-ZAZHENERSA-N	86280086	Isomorellin	16078249
Tripterygone	197388	Morellic acid	16078251
JFACETXYABVHFD-NLFRDLPRSA-N	51455907	Forbesione	16078254
Maitenin	101520	Cochinchinone C	70697833
Sandorinic acid C	11048975	Cochinchinone C	73019072
Bryonolic acid	472768	Neogambogic acid	6438568
Bryonolic acid	6712218	Neogambogic acid	92132426
Tingenone	355376	Gambogic acid	99639195
Isoiguesterin	328559	Isogambogenic acid	101389903
Pristimerine	264268	Gambogin	101690778
Polpunonic acid	169521	Gaudichaudione A	101949804
22-Hydroxytingenone	73147	Gambogenic Acid	102004807
Iguesterin	46881919	Gambogin	102004809
Celastrol methyl ester	159516	Neogambogic acid	102004925
3-Oxoglycyrrhetinic acid	111253	Forbesione	102303099
Tripterin	4274774	Morellic acid	102533562
QVCXDBCRBHXXAD-WFVGHVPHSA-N	24861322	Gambogic acid	126963722

KQJSQWZMSAGSHN-PACOHSDFSA-N	16757868	Gambogic acid	134129562
20alpha-Hydroxytingenone	10717799	Gambogenic Acid	9895478
JFACETXYABVHFD-LQSRWFAZSA-N	16757909	Gambogic acid	9852185
20beta-Hydroxytingenone	44559597	Gambogic acid	11599836
ZTCAJLZRROIDHU-WAFCJUBUSA-N	44572792	Gambogic amide	25252739
FXLVCCDIUNLGKU-BRUCSKOJSA-N	5701992	Gambogellic Acid	52945437
Remangilone D	44558996	Morellic acid	54580250
ZKJXUKUIPQCUMI-QSZQYENJSA-N	10624078	Gambogic acid	3451
AIXBDINLSQYZGS-JWNVRJNBSA-N	45482038	Acetyl isogambogic acid	6857789
GSIDIGLXCWJQPN-YMEQZMBDSA-N	25424418	Gambogin	11753679
DMHJWMODPYSFCD-DUSJHGMNSA-N	71498413	Gambogic amide	56951189
PDMJIWXURDBAOZ-CIFOREJBSA-N	45482047	Gambogic acid	91668478
21-Oxopristimerine	50907761	Gambogic acid	5281632
Dispermoquinone	53320376	Gambogic acid	5380091
QHYPSOWPDMYTTQ-UNNRZNSMSA-N	71498452	Gambogic acid	5353639
QIRUFAFQGKOTKA-YKUCPAPWSA-N	389017	Isomorellic acid	9915833
OQLDDXDMTOPTDO-QRARIYCASA-N	10695614	Gambogenic Acid	10794070
QGWDYPREORDRIT-DGRUGRQQSA-N	16745529	Gambogic acid	20054919
QGWDYPREORDRIT-DSIOGZMYSA-N	229868	Isogambogenic acid	70639870
ZNFSSQAJGMMWBY-WXPPGMDDSA-N	25197280	Gambogic acid	70639872
WHDKOWNIOGJXHK-PTRAYGLTSA-N	45482045	10-Hydroxygambogic Acid	71450485
OQLDDXDMTOPTDO-PEKWGEHZSA-N	44289021	BLDWFKHVHHINGR-UHFFFAOYSA-N	125071
GAPWCQHXCIXKLV-RKHSXEAASA-N	6708798	Deoxymorellin	635828
QGWDYPREORDRIT-UHFFFAOYSA-N	3129312	Morellic acid	5319893
RRRZQVJZDVPAJN-ZRCCSVPJSA-N	25197172	Isomorellin	5364585
Polpunonic acid	169521	Isomorellic acid	5366120
KZCZQJNPWZPAEJ-ZRCCSVPJSA-N	25195610	DRRWWKSGTSQOON-DXMWQDMHSA-N	52947871

Dihydrocelastrol	10411574	7-Methoxyepigambogic acid	45270567
FATJTRUVRFSESL-HZYNXAPGSA-N	44566365	UONSIJYWYKPEDK-NXWWLHKBSA-N	45272194
Isoiguesterinol	10477355	BIGAHFWHALQTRA-TYAVBBKTSA-N	45272195
Wilfolic acid C	44559659	GJFGYTWPUIBTJN-UOCUBHIGSA-N	45272197
HVRSOVWJUJGHSI-JHGSJXKWSA-N	44558965	LJUARHDVFLLQMF-KELXBUOKSA-N	45272281
11-Oxoursonic acid	22210052	BMPKGNCYXQRUMA-BLACGAOQSA-N	52941797
6-Oxopristimerol	11754914	JEWKRGWMKMUBBF-JFDIIJRYSA-N	52943022
MAAVNXPPBHQXNL-INMPMWFSSA-N	6710688	ORRQVYBMKAECLD-KMIBLQPDSA-N	52946647
GZKMFYLBPHPWEI-DGRUQNLJSA-N	363631	QQQNFIABWPXSFD-PUYFONRISA-N	52946668
LUSKWXDBHNQVEL-JXZITWAVSA-N	53323497	JTTOKKDLXZMVBA-IHECDDROSA-N	52947870
QGWDYPREORDRIT-NFGWGJPRSA-N	73758286	XSDFUWMDPUDYHN-CLKBEGISSA-N	52948934
CDOKUYLTAYCBST-GKMFGOQDSA-N	9869964	Gaudichaudione H	57345562
CPFUJAKVJBYWJA-BTRWLOMLSA-N	10411393	6-Hydroxycluvenone	57390441
YXEPIOPYEDGEEP-IITIDZKMSA-N	71498373	KCALBYYRHOKKBO-BCFMDSCZSA-N	57395879
KGQOHECLTONQOT-HKFGFSCZSA-N	10645721	CSIZKMXLDPOBKM-PQIHDOCZSA-N	57395880
QEACUYQTRMGOSD-MZRSIZMESA-N	14465806	XZPYBDPGHSLINX-OHTINKFRSA-N	71720741
CVAILKMOFONEDU-KRJMWWHISA-N	15765122	KCYKVBCVFBSOKZ-RWPPGCTJSA-N	118737824
2-Picenecarboxylic acid	23757062	PVRDWAUIIJESEC-WFSGSTODSA-N	118737833
QLTFHGMEDZMMTF-VUQPYPCZSA-N	44298352	OIZCJCKDTJMDFV-CVTCUNNCSA-N	118737835
OKOGABAEYJRJOB-JSJVQHDDSA-N	46184390	KSEYPOHNDKJEPF-ILHGWRPKSA-N	45267055
UZEJIMRGSSLBIV-SAQIBKBSSA-N	25197281	Dimethyl gambogate	6857785
Fupenzic acid	12045007	Garcinolic acid	6857794
WZAUFGYINZYCKH-AQCDROMSSA-N	6708673	Methyl gambogate	12113746
Amazoquinone	44559090	Decahydrogambogic acid	5149276
Dihydrocelastryl diacetate	9828620	REDMIYQFNIRTDF-WIKVJIRTSA-N	16758035
20-Epi-isoiguesterinol	21575471	VZXLWEWYBUGLJA-SAABBMRESA-N	44449775
TTWPKNPRMGVGJO-QZLVDJLTSA-N	49797930	FJRORJDZZLUAPP-BUJCIKCXSA-N	25208438

23-Nor-Blepharodol	53320826	Dimethyl-Ga	44449753
GSIDIGLXCWJQPN-BRLXHVQISA-N	5336986	7-Methoxygambogic acid	45268014
SAOOBRUHTPONGX-UANCAJPASA-N	6708713	7-Methoxygambogellic acid	25208761
PLXMKOYILCBYGS-UXDHXBHYSA-N	6710689	VZIUOBFNNRUPAK-QBELJZEMSA-N	71717656
QGWDYPREORDRIT-VLYMFKARSA-N	44435792	JDGCURVTMMXMDY-TYTBCFIUSA-N	71718886
DQHHRVQZUPBARM-LRRZNWEGSA-N	90233240	FTTRVECHPLAERQ-UZTNAXEXSA-N	71717668
YQUGEUGWFBESNQ-WXPPGMDDSA-N	118408942	UCLJDAOAFDHFEA-XYBONBLRSA-N	71717669
QOGSXJHNNDQXSS-YNJIRTJXSA-N	88303297	PUDIAMZKKXFSOI-UZTNAXEXSA-N	71717674
SKMCTUIWOMRTKB-RLLZTQLFSA-N	68198583	CDIIOQLRYIQLOZ-JSJFTWSHSA-N	71717675
GAVQRDLYJPTRLX-NGXGXHCGSA-N	68028960	KKPBMYHMXYLZMU-KDBWZQHXSA-N	71718259
QGWDYPREORDRIT-XYMLVXMESA-N	60103590	PYCHDQHGAJKLEB-ALPBHPBSSA-N	71718272
XLZGVFDYBDKLSJ-HKUCNJHLSA-N	59492093	XTQBIQOZQZIJJW-HQCJEUBQSA-N	71718866
INHUILMVLIXIOP-LLYXLEJJSA-N	59428287	CTSIBWDVPNXDDY-WSOPLZEESA-N	71718874
OZKBTRIBGQYSPM-VRAWMDIMSA-N	59350382	GSULQLPVUMJWJN-YWJDODGHSA-N	71718876
PNZQDIUZNCHSAL-ZRCCSVPJSA-N	58338790	OSDWBDHHKBVHAI-CNADFBRCSA-N	71717053
ILAWPUDRPIJSHC-CDKFMWKUSA-N	58338787	BZRJTOIKOJHZGQ-ALPBHPBSSA-N	71716428
QGWDYPREORDRIT-OUMYWODESA-N	57321836	QSJAAIXXDPZJDI-QVZJNGHZSA-N	71717051
FGUPEMLTKNSPDS-WYUYVVTISA-N	56847557	KLEWBUIXEFWIOX-QBELJZEMSA-N	71717031
TZYINTOVRMOTDT-WXPPGMDDSA-N	56847496	NPPZUPSAZVYWJX-RGAUDQMMSA-N	71716434
JKZDPNSGTLVKFS-JSJVQHDDSA-N	25197171	CRZXNKVWVIOGLG-KCJGJVMNSA-N	71717027
PPBNDNNFAYLUKF-SUNMMQDUSA-N	25197168	MGBNLVNNMKFWHO-MNDRQJQGSA-N	71717026
XAWKZQBUEYUMJQ-WXPPGMDDSA-N	118404337	PJRSCTLUXRMVLG-LPIQBNQASA-N	71717025
LMRBFTMNXRSIDU-ANCVDJAISA-N	56846808	YRQDRRMNQYXEQF-TYTBCFIUSA-N	71716450
GKBQZTNFLYSBDI-HKUCNJHLSA-N	90922624	LKTCFJPAKUCNIB-QBELJZEMSA-N	71716443
KTHCLZLOFJSMHN-NGXGXHCGSA-N	68029427	YKSIZBBLYFQODP-YWJDODGHSA-N	71716435
XFFJOOGKLFMNFN-NOZRFFRFSA-N	68029424	FJYQZSOKEMDMHD-RGAUDQMMSA-N	71720740
MEVHISVZZRNRNQ-HJJLTIBASA-N	67406451	VOYHKSWHADONRT-UHFFFAOYSA-N	117591472

QGWDYPREORDRIT-GWEJSANNSA-N	60103581	VJXSSTZKBZWPPU-UHFFFAOYSA-N	117591470
CWNGKYCBOAGDTO-MIZXVQKXSA-N	59546528	FWOBTOPXTHIFAR-UHFFFAOYSA-N	117591530
HMWZFAJRZJAGAL-CDKFMWKUSA-N	58338783	PEVFZCIVSYJXJJ-UHFFFAOYSA-N	117591589
RZBXUSAFLXVWGM-ANCVDJAISA-N	56847608	YPFPFCJFKUOQMA-UHFFFAOYSA-N	117592388
WDJQPABVFANMAQ-XIRGYHLMSA-N	118408890	DQPKYEHRFQDOLK-RWPPGCTJSA-N	118737825
QGWDYPREORDRIT-BPTQZAATSA-N	54227450	ZYUWYMMXGSCUQT-UHFFFAOYSA-N	117591047
YGLDJGRHKCVIQN-ZRCCSVPJSA-N	46184627	YANZBSJNCIQKSX-ZFHAUAHYSA-N	118737828
DHGOQCNNFIELMK-ABJUJWBKSA-N	42630196	KJBDNAFUXXSLAD-UHFFFAOYSA-N	117590910
HKRKUMPPXYIXCF-JSJVQHDDSA-N	25197278	AICGDFMIRHRSPB-MMEAOPOPSA-N	118737829
ZEXSHDSVCZQLCO-WFVGHVPHSA-N	25197169	CXFIQFGADOTDPF-CMLKYFHXSA-N	76311369
QGWDYPREORDRIT-ZUZOHOKMSA-N	21118757	XUYFKRZDNDQABJ-XDSJHDTISA-N	71718887
11-Oxooleanonic acid	11576389	BAHZVPGPIWTLLZ-RGAUDQMMSA-N	71720723
KQJSQWZMSAGSHN-AHWKKIARSA-N	58636951	UEMVXKVCUBRCBG-KCJGJVMNSA-N	71720719
KQJSQWZMSAGSHN-WDVYODBHSA-N	118655639	XAUUFNQHVHEWRB-RGAUDQMMSA-N	71720718
YDUSFXWVWNIQDZ-IGTSBUIGSA-N	118404346	YGVBIDJABJLVOL-DGTITLQCSA-N	71720708
HCAYAMHXSDCPTR-MRZDQBBQSA-N	118404343	COEJNDYMOAULJK-JCJNIYCKSA-N	71720089
KQJSQWZMSAGSHN-MRHUJCAJSA-N	90233238	JESNVXLQVMHXLZ-JSJFTWSHSA-N	71720088
FESQODXDXOHLEO-JNEFGXKCSA-N	71249962	XTBWCIYKONYIDG-AQOIPSNVSA-N	71720068
SMYCYEXFRJMQLW-ANQVMFJUSA-N	71167315	WSCCRZWUYZCJAZ-KAJKVYITSA-N	71719495
FMTPULGTIHBJRT-LGVWSNLESA-N	69574940	LWIGRTRTVVPXOZ-XDSJHDTISA-N	71719493
UAJBCGCAPNHLHM-QUYLDEAFSA-N	68028959	ZMMPPBWNSJCZGR-QVZJNGHZSA-N	71719492
LLKPYONQSFCTMG-IGTSBUIGSA-N	123598084	XCRBRZWMQVMPIY-KENSWSBLSA-N	23629033
UQSBWMQFUNJXTI-AZUGTCGHSA-N	58338788	BIMUUWRNJBALEP-AWNOVZCOSA-N	45268895
REKHQMDGAPXWJP-ANCVDJAISA-N	56847556	DBWZFQAUMYEWML-WWZJDETNSA-N	45267910
GBQYERWLNHTHAH-ZRCCSVPJSA-N	56847495	34-Hydroxy-gambogic acid	45267909
VMVBZDHQRFGSLA-FAWNWTIBSA-N	56846810	DXSGQRROBDYCSJ-YVMJPLCQSA-N	45267154
AQKDBFWJOPNOKZ-MTWWMYJUSA-N	53656716	XRBAWHATHDIBFU-XOSCNRPVSA-N	45266948

MXMNIRXPSWDEED-GTKRWHGSSA-N	25197170	XCRBRZWMQVMPIY-OQOGLVOPSA-N	44583737
RPGDRWMPFKEMPI-JSJVQHDDSA-N	25177624	JBRMVLBIZUTECR-FXRUCJBFSA-N	44403668
QGWDYPREORDRIT-OHHDNCQJSA-N	16401165	VZQQLPACAVHZQT-OLZUXEKSSA-N	44403667
GYUVZGGERRSPQY-UHKCKZGUSA-N	66583327	Gambogenific acid	25208911
SYLIRTRYLBYOBO-JJWQIEBTSA-N	91408393	UQHRXFAVXKZFRU-RGAUDQMMSA-N	71716426
VDDPQVQCXWFZJM-LDLRJHFFSA-N	91367949	RIFZOYVQOFHOCS-WIKVJIRTSA-N	16758013
LSSXGFNMUHCIAI-WXPPGMDDSA-N	91190512	REDMIYQFNIRTDF-LARDOQITSA-N	5469880
SYLIRTRYLBYOBO-AHWKKIARSA-N	91065125	VLADFNOTLYCWMW-UHFFFAOYSA-N	52916109
WNCVCMKJUFWTFA-OWDORJPTSA-N	71249923	XZPYBDPGHSLINX-UHFFFAOYSA-N	52916002
QGWDYPREORDRIT-AVJZJKPBSA-N	70532695	Tetrahydrogambogic acid	9917275
DNLGFGXSBOIDNV-ZRCCSVPJSA-N	67421869	Dihydrogambogic acid	6857793
URDWIVHPIKIQNK-JJWQIEBTSA-N	66583411	HHOLRSVIZVDOIV-OXFPAKKDSA-N	52949100
WJLHDYQXCJXZGV-WXPPGMDDSA-N	91463431	XUYFKRZDNDQABJ-DLPQTZSGSA-N	57400963
QGWDYPREORDRIT-WBTUSMEDSA-N	60103570	HHVDVNJHGHGGHI-LDNSNGAXSA-N	57394941
QGWDYPREORDRIT-VPHAENBISA-N	57301674	YTIQONQLSSBXHE-MOWCMFFRSA-N	70691871
QOGSXJHNNDQXSS-ZHITZLKESA-N	57051700	LWIGRTRTVVPXOZ-DLPQTZSGSA-N	57393978
LMHNQDYMADJAAM-CDKFMWKUSA-N	56847555	TZXRZTWZRWZRQS-WSOPLZEESA-N	71716427
GICPBFRHOLICHK-UHKCKZGUSA-N	56847494	LPYYTLGGSMEQMH-ZSJJNDTGSA-N	45269643
WWKHRRYBBUOLCO-CDKFMWKUSA-N	25197277	Cochinchinoxanthone	53355017
QGWDYPREORDRIT-MQLBBMOOSA-N	18637982	BKRLQHWNGLIVCW-NFWYAXIXSA-N	52945436
		33-Chlorogambogellic Acid	52943021
		GBQLXOPZKHBGOY-SCWSFWMSSA-N	46886397
		DRRWWKSGTSQOON-WIRZGQEJSA-N	46886396
		QYRPARUSUFWOPG-HBWLMKOJSA-N	45272280
		GITYGECAVAWXHS-FZGWIHBJSA-N	45272279
		(9,10)-Dihydroxy-gambogic acid	45270476
		7-Methoxyisomorellinol	45269745
		ONKMNKXXFSJVSP-UHFFFAOYSA-N	117591965
		PLPLFPMHLSHHDS-UHFFFAOYSA-N	117593100
		MUQLGQHFYINEFE-UHFFFAOYSA-N	117592428
		YHHMSCSXVOQXAF-UHFFFAOYSA-N	117592536
		IVZPDDZYGYQLHF-UHFFFAOYSA-N	117592735
		RZRZWSHHUHKZRH-UHFFFAOYSA-N	117592761

		XCNLXYWAJHGEPU-UHFFFAOYSA-N	117592772
		HAZSRZGEUAECEB-UHFFFAOYSA-N	117592892
		PWVDTBGVDGFEKA-UHFFFAOYSA-N	117592295
		ILJYFDSGSZVXFL-LQEUQGNQSA-N	118753349
		XXHKTHKJENJGLT-UHFFFAOYSA-N	117592031
		DQLSLPKMHSAVGY-UHFFFAOYSA-N	117592271
		HEZLRSFHGIYFNV-UHFFFAOYSA-N	117592219
		WVBHHKPTCIYZRW-UHFFFAOYSA-N	117592147
		ANXHWYZDFGXCSL-UHFFFAOYSA-N	117592105
		FEUKMNUTHOBIFJ-UHFFFAOYSA-N	117592081
		JEHWFPNSWRPRFY-UHFFFAOYSA-N	117593116
		NBOTXPPFMTZDJX-UHFFFAOYSA-N	117593375
		RIZDOCXODMLFMH-UHFFFAOYSA-N	117593544
		RJZOSQYZDIYLOD-UHFFFAOYSA-N	117593652
		CIMLDIMAIQTAAF-UHFFFAOYSA-N	117593685
		SLOASZPUTSRJOS-UHFFFAOYSA-N	117594239
		KQBIZWOEXACMRJ-UHFFFAOYSA-N	117594907
		HQPAKWNGDAIVMM-UHFFFAOYSA-N	117594911
		XITFXYOJRIAYLM-UHFFFAOYSA-N	117595036
		UUNPNKKRLMOBNZ-UHFFFAOYSA-N	117595578
		XWFNYKWKDWAAMZ-QKBJRNKPSA-N	118707564
		ZISRIFHOONSTEW-XQUYNDDWSA-N	118753102
		MQXZYUNEWMQRJD-MIRSFJNZSA-N	118753301
		WCBINTABDRSBOM-BXQNXPOQSA-N	118753302
		TZPROOSLLJNQHV-YBSJKGMBSA-N	118753348
		Scortechinone A	44559179
		9,10-Dihydrogambogic Acid	71459533
		GEZHEQNLKAOMCA-UOONSFDBSA-N	58209843
		Gambogic acid amide	16725080
		REDMIYQFNIRTDF-UCQKPKSFSA-N	5475311
		Acetyl isoallogambogic acid	6857765
		Gamboginic acid, methyl ester	23806091
		REDMIYQFNIRTDF-OZWPVNNZSA-N	44449776
		Decahydro-Ga	44449798
		Tetrahydro-Ga	44449824
		UYPYPAISERHQAO-PBBIOFTGSA-N	44452392
		FNJGRUCXYDWBQQ-UHFFFAOYSA-N	117591871
		Scortechinone B	44559180
		Scortechinone I	44559181
		Scortechinone-Q	44559270
		Scortechinone R	44559271
		Scortechinone S	44559272
		Bractatin	44583731
		1-O-Methylbractatin	44583733
		Methyl 8, 8a-dihydromorellate	45268013
		YTIQONQLSSBXHE-OGRXGPJBSA-N	70687664
		OLVQCRKEVCKWSS-WNDLJKFFSA-N	71717041
		UPJGOGQGKKPFQF-VJEMJKLZSA-N	71720078
		WNQJCSUJMQRMEE-KWXZSCLYSA-N	76315019

		FBNIACMJHDKGKH-UHFFFAOYSA-N	117591045
		MRFFDQCGXSUJFO-UHFFFAOYSA-N	117591692
		KGPHOVCVHPOWBR-UHFFFAOYSA-N	117591869
		GEZHEQNLKAOMCA-UBYIDDGGSA-N	91332450
		QOZHTUAZXBIGBU-WRXOINPPSA-N	117647595
		GEZHEQNLKAOMCA-KSZVLNGESA-N	91351716
		GEZHEQNLKAOMCA-IGPPFNQUSA-N	91356587
		GEZHEQNLKAOMCA-WQMCTBSRSA-N	91395869
		REDMIYQFNIRTDF-CLWCGEPSSA-N	91421299
		DVTLNRRWCRGSEB-QLMUFRIZSA-N	91507797
		DCUBEADPOQPJCP-WWYBWCOQSA-N	91525565
		30-Hydroxygambogic acid	102004804
		CCEGWRPYDCDELZ-JDHSLWBYSA-N	117640037
		REDMIYQFNIRTDF-UVYBHTOASA-N	117640050
		CXFIQFGADOTDPF-VAVNHFACSA-N	91116286
		GEZHEQNLKAOMCA-QSNZZALHSA-N	91081424
		MNNVIONVHRRQPF-IGPPFNQUSA-N	90998876
		CGTWSSOGZQAVNI-YUTXIDHZSA-N	90956184
		MNNVIONVHRRQPF-QSNZZALHSA-N	90904334
		GEZHEQNLKAOMCA-BMAVOULBSA-N	90837811
		MFUIGIDUBRLELJ-RHDRSXQYSA-N	90802529
		FAEQAXFMLPWRFS-UHFFFAOYSA-N	90793624
		XAPLNRWTXVBXJQ-UHFFFAOYSA-N	89737453
		AAEQTEKIFSEBLF-JCDNVTHQSA-N	123214118
		IQDYCICYXWCEEX-QTFYUPPWSA-N	89593387
		PHWVEYPUZJUGEV-OAWOWVGUSA-N	121241349
		IWZRSTDKYHZSQF-SDRUQSECSA-N	88870689
		AORIIYVDTXBCHV-YCVDEPICSA-N	123197254
		JTEORTUOYDVEOM-FOQNCPQJSA-N	123187933
		GCHJONZZLLRAEM-IGQYWBJASA-N	123186661
		LPPVILAKSVOTHC-VABJNMDGSA-N	123168249
		DUZIVTBZXZNFKW-LGYDYSPQSA-N	123153387
		XDJCBNWCKIBHCH-JLZOOELISA-N	123146135
		IWZRSTDKYHZSQF-FBFXOJLPSA-N	123143901
		UJPXBLUJNQMYIY-VRJUNHGMSA-N	122542892
		FHJLQIOSAMVKBE-UTKNUOMGSA-N	121241350
		MPVLKYHKVPLUBC-WRXOINPPSA-N	117649446
		RUOIRPONLNREJT-IXTCAIOQSA-N	121241347
		RUOIRPONLNREJT-GAZVMYCTSA-N	121241346
		WVBSIKJNTVCEQJ-NBQSLMHUSA-N	121241345
		RCWNBHCZYXWDOV-WPKINVRVSA-N	121241344
		WTHZNKDLVYBPIU-XKZIYDEJSA-N	121241343
		VZXLWEWYBUGLJA-KCZYMQEJSA-N	118218885
		KWSMUTWPBWYJTJ-GBFWCEHUSA-N	118215929
		YTINOMMNLVRQDZ-ODZJVPPQSA-N	118215928
		PNVQUHOJEOSPST-CZHHEZJISA-N	117649448
		AORIIYVDTXBCHV-NTXDIHSUSA-N	88870699
		TZPSXPDSLJISGI-JSTMFIRTSA-N	88870710
		SYPMLUQDIRBAJH-DPGBVESVSA-N	88870709

		TZPSXPDSLJISGI-SMCHVARRSA-N	88870708
		KIIICIKEPTYGHX-HQTPSEOASA-N	88870707
		KIIICIKEPTYGHX-WLVTUAKASA-N	88870706
		SYPMLUQDIRBAJH-GKPRHQBLSA-N	88870705
		MPUTYJMBORRTBJ-BJGVHYDOSA-N	88870704
		GCHJONZZLLRAEM-ZWNPRXLMSA-N	88870703
		AORIIYVDTXBCHV-XENTYZTMSA-N	88870702
		LPPVILAKSVOTHC-MGXSGBCKSA-N	88870701
		XTMOYKJZVWYKPJ-BEYSKSGQSA-N	88870711
		XDJCBNWCKIBHCH-USNSJPIISA-N	88870698
		PVGHFWKFADSYSJ-MUUFOGJZSA-N	88870697
		TZPSXPDSLJISGI-FZVGPGJDSA-N	88870696
		DUZIVTBZXZNFKW-DJXAADCISA-N	88870695
		FMCZWXSKQSSEHP-DJXAADCISA-N	88870694
		DUZIVTBZXZNFKW-HAAYKULCSA-N	88870693
		KIIICIKEPTYGHX-LNYPENFMSA-N	88870692
		KLXFRVJTZKOFKV-SXQTYUKPSA-N	88870691
		TZPSXPDSLJISGI-OUIKJMRCSA-N	88870690
		XIDKYIKGGBTUPH-LFVJCYFKSA-N	126602554
		PVGHFWKFADSYSJ-VSBOKRGHSA-N	88870721
		AFLHWBGXZWEULQ-PKAZHMFMSA-N	89410167
		DUEZXUMGZFEZCZ-OYKKKHCWSA-N	89409368
		KFTSCWCFQNHXEF-OYKKKHCWSA-N	89409366
		MPUTYJMBORRTBJ-ZRJUZVLNSA-N	88870748
		IWZRSTDKYHZSQF-SAJNXLGZSA-N	88870747
		AORIIYVDTXBCHV-PKRKJABKSA-N	88870746
		KLXFRVJTZKOFKV-PQINOVKQSA-N	88870744
		MVRLZFHWUTWTPZ-OAWOWVGUSA-N	88870743
		IWZRSTDKYHZSQF-WASDJRSKSA-N	88870742
		XDJCBNWCKIBHCH-FXXVQEQYSA-N	88870740
		VJGFCQXTEBDXCL-XYGWBWBKSA-N	89410235
		FJTHIYKOKGKDFM-YECKHLLKSA-N	88870720
		MPUTYJMBORRTBJ-ZFGQNZLVSA-N	88870719
		GCHJONZZLLRAEM-RHZAVJPWSA-N	88870718
		FQCIAFWZLWMCNQ-HSULCKAZSA-N	88870717
		XTMOYKJZVWYKPJ-REPUEAQBSA-N	88870716
		SYPMLUQDIRBAJH-VGRXJTFRSA-N	88870715
		NHNBZYVAEYGXJD-PYOCCJRJSA-N	88870714
		KKKVOYLLLKLJGN-MWJHYMAZSA-N	88870713
		FMCZWXSKQSSEHP-XENTYZTMSA-N	88870712
		FJTHIYKOKGKDFM-WHJBVOODSA-N	123809343
		GFZFBIVRUFQXDE-RAKWAVLCSA-N	123867307
		JTEORTUOYDVEOM-AKJUZXHISA-N	123858966
		KLXFRVJTZKOFKV-RWJQYVGMSA-N	123858384
		XTMOYKJZVWYKPJ-IIIUNIONSA-N	123851264
		COVMVPHACFXMAX-NJEUQTODSA-N	123849167
		PVGHFWKFADSYSJ-QFSWFWNDSA-N	123845773
		TZPSXPDSLJISGI-OWMZLRFQSA-N	123844233
		DUZIVTBZXZNFKW-YCVDEPICSA-N	123821973

		AORIIYVDTXBCHV-UOVBMFSZSA-N	123819800
		IWZRSTDKYHZSQF-OWDHWTJPSA-N	123811651
		KWSMUTWPBWYJTJ-GTUNQJGZSA-N	123867860
		XTMOYKJZVWYKPJ-ZPHHXPIHSA-N	123801882
		IWZRSTDKYHZSQF-STFZBRADSA-N	123799471
		FMCZWXSKQSSEHP-LGYDYSPQSA-N	123783895
		MPUTYJMBORRTBJ-UIVVFIOZSA-N	123768632
		FMCZWXSKQSSEHP-NRBGCZKASA-N	123764889
		TZPSXPDSLJISGI-HHTAKYNCSA-N	123747944
		FJTHIYKOKGKDFM-VABJNMDGSA-N	123730782
		COQAPWLZSHQTKA-WQXODUOJSA-N	123725925
		SYPMLUQDIRBAJH-QHKZOBPHSA-N	123713443
		GCHJONZZLLRAEM-MNTUFJQYSA-N	123689815
		GEZHEQNLKAOMCA-MAKUZWOISA-N	123970099
		MPUTYJMBORRTBJ-GQPSOBIZSA-N	124083343
		MVRLZFHWUTWTPZ-PLUQQRNKSA-N	124083342
		TZNZFVPLCOBOHE-SRHIZFSVSA-N	124083341
		KRGVLKLMSZLNJD-CZBSRGPZSA-N	124083340
		UNPJLJMTRSOEBG-YTJXBEJASA-N	124083339
		MPUTYJMBORRTBJ-ORYAWNIFSA-N	124011815
		NHNBZYVAEYGXJD-WQXODUOJSA-N	124009406
		FMCZWXSKQSSEHP-UOVBMFSZSA-N	124004813
		XDJCBNWCKIBHCH-PCCITZADSA-N	123995047
		IWZRSTDKYHZSQF-IVPRPUKMSA-N	123970404
		PVGHFWKFADSYSJ-AHWIIWHVSA-N	123672510
		REDMIYQFNIRTDF-SNSZSSRMSA-N	123962837
		NFVXKLYWFCNBCO-BIDYHREASA-N	123946655
		FQCIAFWZLWMCNQ-MZQQFRDZSA-N	123931583
		FMCZWXSKQSSEHP-YCVDEPICSA-N	123926824
		SYPMLUQDIRBAJH-JKFLJFCISA-N	123921560
		SYPMLUQDIRBAJH-MXTXYYSDSA-N	123903462
		KIIICIKEPTYGHX-KAXRJKLWSA-N	123902138
		KLXFRVJTZKOFKV-XLPBWHEJSA-N	123898922
		KKKVOYLLLKLJGN-QHQAMMJOSA-N	123885028
		GEZHEQNLKAOMCA-RAKWAVLCSA-N	123307604
		TZPSXPDSLJISGI-GTHFLSHASA-N	123415116
		KIIICIKEPTYGHX-FEHYYDPSSA-N	123392802
		KIIICIKEPTYGHX-QGNRMNGYSA-N	123385862
		FBJVPBMWPVODRO-PCCITZADSA-N	123363376
		GEZHEQNLKAOMCA-ADCYJAEYSA-N	123352107
		DUZIVTBZXZNFKW-UOVBMFSZSA-N	123351208
		SYPMLUQDIRBAJH-KUWKLSGISA-N	123340327
		AORIIYVDTXBCHV-NRBGCZKASA-N	123338137
		TZPSXPDSLJISGI-OYYJVTFHSA-N	123324331
		COVMVPHACFXMAX-WJQTUEGHSA-N	123322435
		AORIIYVDTXBCHV-NWODWGALSA-N	123437820
		NHNBZYVAEYGXJD-RNLYVTMESA-N	123307294
		XDJCBNWCKIBHCH-MEXQKCLWSA-N	123306297
		QOZHTUAZXBIGBU-UHFFFAOYSA-N	123299953

		SYPMLUQDIRBAJH-CFCCRWGCSA-N	123290623
		UWZMGTSPGQXAAP-WQXODUOJSA-N	123272237
		NFVXKLYWFCNBCO-VDMQVCGESA-N	123263934
		BYSLEZZCJZXNQG-RWJQYVGMSA-N	123248797
		DUEZXUMGZFEZCZ-UHFFFAOYSA-N	123248044
		HWTUJOKAWVHJCJ-OPUOJSSUSA-N	123242824
		RCWNBHCZYXWDOV-KBGSTRFQSA-N	123550305
		XDJCBNWCKIBHCH-RAKWAVLCSA-N	123661979
		SYPMLUQDIRBAJH-XZGVVOIPSA-N	123658150
		UWZMGTSPGQXAAP-RNLYVTMESA-N	123650727
		KFTSCWCFQNHXEF-UHFFFAOYSA-N	123649704
		KIIICIKEPTYGHX-ILHXXLRDSA-N	123628892
		GEZHEQNLKAOMCA-VRYKAPMJSA-N	123626409
		AORIIYVDTXBCHV-LGYDYSPQSA-N	123615972
		DUZIVTBZXZNFKW-NRBGCZKASA-N	123581184
		GEZHEQNLKAOMCA-CCZYIYSKSA-N	123570917
		XTMOYKJZVWYKPJ-NVDLJSELSA-N	123554777
		BYSLEZZCJZXNQG-TYTNRNEYSA-N	123234071
		PVGHFWKFADSYSJ-KFMJCXDSSA-N	123549300
		PVGHFWKFADSYSJ-JTFGTGAKSA-N	123542499
		AAEQTEKIFSEBLF-DZGHJPMGSA-N	123522015
		FQCIAFWZLWMCNQ-MLFAQUDHSA-N	123514172
		XTMOYKJZVWYKPJ-GMLGDLCKSA-N	123504656
		MPUTYJMBORRTBJ-WHGJWNQISA-N	123494893
		RCWNBHCZYXWDOV-NWPLKAIBSA-N	123475850
		YTINOMMNLVRQDZ-MCGXLDAJSA-N	123469661
		GEZHEQNLKAOMCA-BIDYHREASA-N	123464815
		IWZRSTDKYHZSQF-GATNSQFVSA-N	58545200
		RCWNBHCZYXWDOV-YZYLOZNXSA-N	58545211
		SYPMLUQDIRBAJH-HYTANOMUSA-N	58545210
		XTMOYKJZVWYKPJ-ZOLZCLEESA-N	58545209
		GEZHEQNLKAOMCA-BSGXHHMHSA-N	58545208
		AORIIYVDTXBCHV-NTJOPWEGSA-N	58545207
		UWZMGTSPGQXAAP-CAQBGEGQSA-N	58545206
		VZQQLPACAVHZQT-IYYXIFPBSA-N	58545204
		DMEVOSRJMNDQOB-XKZIYDEJSA-N	58545203
		DUZIVTBZXZNFKW-SMSNCEFTSA-N	58545202
		COVMVPHACFXMAX-PWZQPALBSA-N	58545201
		KLXFRVJTZKOFKV-XFTJLXKISA-N	58545213
		AAEQTEKIFSEBLF-CEGNIYDJSA-N	58545199
		KIIICIKEPTYGHX-FBQUOBMKSA-N	58545198
		IWZRSTDKYHZSQF-HJBKNYNFSA-N	58545197
		FJTHIYKOKGKDFM-IDACPZNESA-N	58545196
		PVGHFWKFADSYSJ-SZYQWLCWSA-N	58545195
		KIIICIKEPTYGHX-ASZVOQTMSA-N	58545193
		AORIIYVDTXBCHV-SMSNCEFTSA-N	58545192
		SYPMLUQDIRBAJH-WVJIZLELSA-N	58545190
		GEZHEQNLKAOMCA-ITMOWBSKSA-N	58545188
		TZPSXPDSLJISGI-IOHWOTBESA-N	58545187

		LPPVILAKSVOTHC-IDACPZNESA-N	58545228
		CXFIQFGADOTDPF-DLLRBFTDSA-N	59248849
		RAWMYNQUVMHIBX-IADYIPOJSA-N	59060966
		MFUIGIDUBRLELJ-SJZKZEADSA-N	58802103
		MKRYDGCYDKJGSE-OMRLFUBUSA-N	58802102
		MVRLZFHWUTWTPZ-UZAZBKDBSA-N	58554586
		KIIICIKEPTYGHX-MGTQMBPOSA-N	58554585
		MPUTYJMBORRTBJ-QMNYGIFOSA-N	58554584
		KKKVOYLLLKLJGN-JTZZIZQTSA-N	58545231
		PHWVEYPUZJUGEV-UZAZBKDBSA-N	58545230
		SYPMLUQDIRBAJH-GUGAKOSKSA-N	58545229
		FQCIAFWZLWMCNQ-PJEVLYNISA-N	58545186
		HWTUJOKAWVHJCJ-OFNKHKRGSA-N	58545227
		MPUTYJMBORRTBJ-UCUGODPPSA-N	58545224
		NHNBZYVAEYGXJD-SORCPTHGSA-N	58545221
		XTMOYKJZVWYKPJ-JBZGCEAXSA-N	58545220
		BYSLEZZCJZXNQG-PSRDIXOWSA-N	58545219
		XDJCBNWCKIBHCH-ASCSEZEHSA-N	58545218
		TZPSXPDSLJISGI-HEERNYNQSA-N	58545217
		COQAPWLZSHQTKA-CAQBGEGQSA-N	58545215
		DUZIVTBZXZNFKW-NTJOPWEGSA-N	58545214
		VDSCKSOYNLTQSY-KKQCBWBVSA-N	6710618
		MNNVIONVHRRQPF-HMMDVQROSA-N	16750435
		MNNVIONVHRRQPF-JDFKUOOISA-N	16750413
		GEZHEQNLKAOMCA-ZEYIWNDBSA-N	11556381
		KCALBYYRHOKKBO-QDTIIGTASA-N	11468283
		DWYQBZCXZCGVHI-YSMPRRRNSA-N	11422617
		KVXKEBWQNIIKMB-MTJSOVHGSA-N	11308419
		AANGJSKPBIKCLU-ITYLOYPMSA-N	9896558
		MFUIGIDUBRLELJ-UCQKPKSFSA-N	9895689
		KYPSMUUXSFJTAR-HEEAUFFFSA-N	9851944
		LFSCNWNADRUBLS-UHFFFAOYSA-N	6710687
		GEZHEQNLKAOMCA-DTWORVFFSA-N	16750462
		COQAPWLZSHQTKA-FRMWRBSQSA-N	6419330
		VDSCKSOYNLTQSY-KKQCBWBVSA-N	6710618
		Isomorellic acid	6419329
		VZQQLPACAVHZQT-RNLYVTMESA-N	6325059
		CXFIQFGADOTDPF-LPYMAVHISA-N	6284659
		VZXLWEWYBUGLJA-UHFFFAOYSA-N	5205218
		CXFIQFGADOTDPF-UHFFFAOYSA-N	5149277
		COVMVPHACFXMAX-UHFFFAOYSA-N	550587
		COQAPWLZSHQTKA-RNLYVTMESA-N	442607
		GEZHEQNLKAOMCA-PCCITZADSA-N	442595
		REDMIYQFNIRTDF-UHFFFAOYSA-N	421874
		PZOHDYPLDDMKLL-BRXPKNJNSA-N	56595878
		FMCZWXSKQSSEHP-NTJOPWEGSA-N	58545185
		FMCZWXSKQSSEHP-SMSNCEFTSA-N	58545184
		GCHJONZZLLRAEM-VSNLTSTASA-N	58545183
		XDJCBNWCKIBHCH-KKCBTXSASA-N	58545182

		PVGHFWKFADSYSJ-QEDXCBQSSA-N	58545181
		NFVXKLYWFCNBCO-UOONSFDBSA-N	58209844
		CGTWSSOGZQAVNI-NXHYFTOVSA-N	57941599
		GEZHEQNLKAOMCA-AOJNUVNQSA-N	57845639
		REDMIYQFNIRTDF-YOXDLCKMSA-N	57586028
		QHQBTUGYLFRMGB-UHFFFAOYSA-N	57332076
		GEZHEQNLKAOMCA-WOMUXGJCSA-N	59248851
		AAEQTEKIFSEBLF-CPNRCEQSSA-N	56595835
		DRCNCMDYOLGEQM-BRXPKNJNSA-N	54764387
		PVRRTDHRPRHFPD-BRXPKNJNSA-N	54764299
		JTEORTUOYDVEOM-LMZOBULNSA-N	25134602
		UYPYPAISERHQAO-XKZIYDEJSA-N	23391922
		YXDVFXVYFZUHNH-OYKKKHCWSA-N	23391915
		WCBINTABDRSBOM-BKUYFWCQSA-N	23391867
		CAZOJRLTSYFYHR-HMAPJEAMSA-N	23391861
		IQYGGNGWJAGSBX-QTSGYQIKSA-N	21603452
		BZFVSJCAEZAUPC-UHFFFAOYSA-N	76658597
		GAIPRNHDISFSOS-UHFFFAOYSA-N	78056261
		JMMQRHVWNJXTCK-UHFFFAOYSA-N	78056259
		ZOKLYUGLLJUVLW-UHFFFAOYSA-N	78056187
		SUOOENMABGOQCH-UHFFFAOYSA-N	78056180
		KCYKVBCVFBSOKZ-UHFFFAOYSA-N	77153183
		JAWDBXPEURIBJV-UHFFFAOYSA-N	77152099
		GUASLOULTWPTKR-UHFFFAOYSA-N	76658664
		SUFYIURUDYGNLI-UHFFFAOYSA-N	76658661
		WCVGFLPSZFYRCL-UHFFFAOYSA-N	76658621
		KRUKCSSBDGDWBK-UHFFFAOYSA-N	76658617
		FSSXEBRQIMPUGN-UHFFFAOYSA-N	78056262
		YHKGUOGCUACMHC-UHFFFAOYSA-N	76658582
		BVWTXTULIJKQBW-UHFFFAOYSA-N	76658577
		UYPYPAISERHQAO-UHFFFAOYSA-N	74047027
		YXDVFXVYFZUHNH-UHFFFAOYSA-N	74047023
		WCBINTABDRSBOM-UHFFFAOYSA-N	74046984
		CAZOJRLTSYFYHR-UHFFFAOYSA-N	74046983
		RCWNBHCZYXWDOV-UHFFFAOYSA-N	73008268
		KCALBYYRHOKKBO-UHFFFAOYSA-N	72973410
		DWYQBZCXZCGVHI-UHFFFAOYSA-N	72955606
		KVXKEBWQNIIKMB-UHFFFAOYSA-N	72795177
		AORIIYVDTXBCHV-HAAYKULCSA-N	88870677
		IWZRSTDKYHZSQF-SEXUYDOESA-N	88870687
		NHNBZYVAEYGXJD-BUZUJXMISA-N	88870686
		KIIICIKEPTYGHX-GWVHQLHWSA-N	88870685
		SYPMLUQDIRBAJH-VWXYWCCPSA-N	88870684
		XTMOYKJZVWYKPJ-XPIJMDORSA-N	88870683
		SYPMLUQDIRBAJH-QNAMZJFPSA-N	88870682
		XTMOYKJZVWYKPJ-KZHARUQXSA-N	88870681
		HWTUJOKAWVHJCJ-KJLFQBLLSA-N	88870680
		PHWVEYPUZJUGEV-PLUQQRNKSA-N	88870679
		AORIIYVDTXBCHV-DJXAADCISA-N	88870678

		RAWMYNQUVMHIBX-UHFFFAOYSA-N	72503918
		XDJCBNWCKIBHCH-NDZGPYJESA-N	88870676
		PVGHFWKFADSYSJ-SQHFVPGGSA-N	88870675
		PVGHFWKFADSYSJ-CZBSRGPZSA-N	88870674
		DUZIVTBZXZNFKW-XENTYZTMSA-N	88870673
		FMCZWXSKQSSEHP-NTXDIHSUSA-N	88870672
		FMCZWXSKQSSEHP-HAAYKULCSA-N	88870671
		DUZIVTBZXZNFKW-NTXDIHSUSA-N	88870670
		FJTHIYKOKGKDFM-MGXSGBCKSA-N	88870669
		AANGJSKPBIKCLU-UHFFFAOYSA-N	85062448
		GEZHEQNLKAOMCA-FSLBZXJLSA-N	66603456
		GEZHEQNLKAOMCA-DDBFOHIYSA-N	70639874
		Deoxymorellin	70639873
		BYSLEZZCJZXNQG-FEOFYTQISA-N	70639871
		GEZHEQNLKAOMCA-ZESOCHHDSA-N	70639869
		Isomorellinol	70639868
		AAEQTEKIFSEBLF-HABQCUFLSA-N	70639867
		CXFIQFGADOTDPF-DXJQLTJMSA-N	70235747
		GEZHEQNLKAOMCA-WFPDQCIUSA-N	70235746
		KCYKVBCVFBSOKZ-KPKJPENVSA-N	66612138
		JAWDBXPEURIBJV-WUXMJOGZSA-N	66603468
		UWZMGTSPGQXAAP-XVQCHYCYSA-N	70639876
		LWIGRTRTVVPXOZ-FNUJIKEJSA-N	66561235
		GEZHEQNLKAOMCA-UTDYKPHNSA-N	66509103
		COQAPWLZSHQTKA-XVQCHYCYSA-N	59895966
		LWIGRTRTVVPXOZ-BVLCDUKVSA-N	59895965
		QDXKAHJQAXFABR-JQJLJWNFSA-N	59895964
		GEZHEQNLKAOMCA-HWOJJXKDSA-N	59895963
		DCUBEADPOQPJCP-YGJDYZIVSA-N	59607534
		HNEQMSDUEKZLSM-HGSFHZCQSA-N	59248855
		LJFULGOYVXFMAL-SSQGKGTLSA-N	59248854
		QGMGFULXKBKTCL-KIXMQPMBSA-N	70641345
		MFUIGIDUBRLELJ-UHFFFAOYSA-N	72428343
		MKRYDGCYDKJGSE-UHFFFAOYSA-N	72428342
		NTOPURIAGNZMKL-IMRQLAEWSA-N	71262488
		FSSXEBRQIMPUGN-UCQKPKSFSA-N	71261269
		GAIPRNHDISFSOS-HMAPJEAMSA-N	71261268
		JMMQRHVWNJXTCK-UCQKPKSFSA-N	71261266
		ZOKLYUGLLJUVLW-ITYLOYPMSA-N	71261175
		SUOOENMABGOQCH-ITYLOYPMSA-N	71261168
		MFUIGIDUBRLELJ-LOSVMQNTSA-N	71215924
		WCILAHSZCPIJPY-NTDTWHTNSA-N	70968535
		FQCIAFWZLWMCNQ-JTWMGGOBSA-N	88870688
		BLDWFKHVHHINGR-BBHBLYOZSA-N	70641343
		NPTGLFQRDIIEBF-IZWSUVCCSA-N	70641230
		NPTGLFQRDIIEBF-BXIWITQHSA-N	70641229
		JEGOOEVCLQIINX-IWBBQFJISA-N	70641225
		IXIIUJBIELTYTO-WECGGMARSA-N	70641219
		FGVLMIINZOOWDO-BRWUPGPDSA-N	70641217

		AAEQTEKIFSEBLF-WXZUYDNESA-N	70639883
		Desoxygambogenin	70639878
		COVMVPHACFXMAX-GJFVWJTOSA-N	70639877

[00382] *Gambogic acid remedies the levels of CTXLP-associated pathological markers CaV2.2 CCAT and Nogo-A*

[00383] **Figure 56** highlights than not only CTXLP is reduced in the presence of gambogic acid, but that the drug treatment also restores CaV2.2 CCAT expression (known to be suppressed by CTXLP, **Figure 47**). Furthermore, we demonstrate show that gambogic acid is a potent inhibitor of Nogo-A (which inhibits remyelination) expression in cancerous NCCIT cells (**Figure 56**). As the decrease in Nogo-A expression directly correlates with a drop in CTXLP expression, this therapeutic strategy may also reduce pathogenic Nogo-A expression in ALS (**Figure 42**). Nogo-A inhibitors have been previously identified, such as green tea polyphenols and other natural product extracts[151]. Proteolytic turnover of Nogo-A is a physiological mechanism to reduce Nogo-A levels[152]. Gambogic acid may reduce Nogo-A expression by having an effect of CTXLP-driven pathology or a more direct effect on specific protein turnover[153,154].

[00384] Development of cell and animal models to investigate CTXLP pathogenesis

[00385] *Identification of CTXLP-encoding ERVK loci in primate genomes and their human homologues*

[00386] Our close relatives also encode ERVK, but some ERVK loci are unique to humans. Examination of the ERVK content of three non-human primate genomes, *Pan troglodytes* (Common chimpanzee), *Gorilla gorilla gorilla* (Western lowland gorilla), and *Cercocebus atys* (Sooty Mangabey) shows that CTXLP is not limited to humans.

[00387] The most recent genomic assembly for each primate species was searched for CTXLP in the same manner as the human genome (**Table 13**). panTro5 and gorGor5 were retrieved from UCSC, and Caty_1.0 was retrieved from NCBI. Chimpanzee ERVK were identified using UCSC table panTro5.nestedRepeats, but no such table exists for the Gorilla or Sooty Mangabey. Gorilla and Mangabey ERVK were identified directly from RepeatMasker output. To reduce the number of small ERV fragments to be BLASTed and to increase the accuracy of orthology predictions by including flanking genomic regions, the loci annotated in RepeatMasker were extended by 1000 bp to either side and then any less than

10 bp apart were merged.

Table 13: ERVK and CTXLP loci in primates.

Orthologous ERVK fragments identified by BLAST

	H. sapiens	*P. troglodytes*	*G. gorilla*	*C. atys*
H. sapiens	**7358**	6060	5853	2585
P. troglodytes	6060	**7389**	5812	2581
G gorilla	5853	5812	**6504**	2562
C atys	2585	2581	2565	**7411**

Number of PF08087* and PF13804* Loci by Species

Species	CTXLP	Env
H. sapiens	28	383
P. troglodytes	33	402
G gorilla	39	318
C atys	31	379

[00388] The expected relationship between the four species is displayed by the number of orthologs recognized. Humans are most closely related to Chimpanzees, then to Gorillas, and most distantly to the Sooty Mangabey. We can see in the second table that the number of CTXLP and Env positive ERVK loci varies minimally between species.

[00389] Only two human CTXLP+ loci are present in all four species. There is also 1 mangabey locus present in all four. No loci were CTXLP+ and present in all four species. **Figure 59** depicts MUSCLE alignments of tBLASTx results from loci in human, chimpanzee, gorilla, and mangabey where an orthologue in at least one species encodes a Toxin_18+ ORF. The first four alignments are split in two, where the top half contains sequences belonging to orthologous sets aligned horizontally, and the bottom half contains the remaining sequences from each species. The fifth alignment is tBLASTx results for the cd-hit cluster representative sequence of the largest cluster of Toxin_18+ ORFs from each species. Only tBLASTx results for loci which encoded a representative sequence of a cluster containing more than one member were included. This alignment was generated by MUSCLE, then curated so that the Cys residues of chimpanzee_108932 aligned better with the other sequences. **Figure 60** suggests that there is more conservation between orthologues in different species than between paralogues from the same species. This pattern is much more apparent for CTXLP than for Env reading frame, where CTXLP appear to be under diversifying selection as indicated by increased dissimilarity between the CTXLP reading frame as compared to the Env reading frame of given sequences. differences are

smaller if present at all. Taken together, this suggests that non-human primates are viable models for CTXLP research.

[00390] *Murine model of ERVK CTXLP*

[00391] Avindra Nath's group has successfully developed a murine model which supports the neurotoxic potential of the ERVK envelope gene towards motor neurons[69]. ERVK env gene transgenic mice exhibit progressive motor dysfunction and hallmark pathology associated with ALS such decreased motor cortex volume and injury to pyramidal neurons and anterior horn cells in the spinal cord. This murine model is a solid platform for ERVK research, yet it remains unclear whether pathology and clinical outcomes were driven by canonical retroviral envelope proteins or CTXLP. This is because the insert used to generate the transgenic mice has the capacity to encode CTXLP (**Figures 57 and 58**). However, these mice do represent a putative model of CTXLP-driven neuropathology and neurodegeneration.

[00392] *Drosophila models of ERVK Env gene products, including CTXLP*

[00393] *Drosophila* (fruit flies) are a widely-used model organism, often used to study the cellular effects of pathogenic human viruses[155]. Moreover, TDP-43 null and TDP-43 mutant flies develop measurable motor deficits[156-158], making this model an exceptional tool to evaluate the impact of ERVK on ALS-like neuropathology and clinical outcomes. In collaboration with Dr. Alberto Civetta (University of Winnipeg), have designed an animal model system in which the ERVK proteins are transgenically expressed in *Drosophila melanogaster*.

[00394] ERVK Env, SU, TM and CTXLP open reading frames have been cloned (by GenScript, USA) into a pUAST vector (Drosophila Genomic Resource Center, #1000), allowing for Gal4 control of transgene expression patterns (see section above on design of custom CTXLP, SU and Env vectors). Generation of an ERVK protein transgenic flies is outsourced to BestGene Inc. (Chino Hills, CA). Flies will be crossed with neuronal (ELAV[156]), glia (repo[159]) or astrocyte (alrm[160])-restricted Gal4 fly strains (Bloomington Drosophila Stock Center 8760, 7415 & 67032, respectively) to generate flies selectively expressing cell-type specific ERVK proteins. For each experimental group, lifespan analysis and locomotor impairment (# of walks/focal) will be monitored as previously described[156,161,162] ENREF 80. To perform pathological examinations, flies will be cold-sacrificed, heads removed, and tissue either flash frozen for western blot or fixed for immunohistochemical analyses. Biological readouts will be correlated with survival and motor-impairment metrics, as to assess how pathological events track with clinical outcomes. In a second series of experiments, fly models exhibiting clinical

impairment will be used to assess the efficacy of a panel of CTXLP inhibitors, such as celastrol and gambogic acid (**Figures 54-58**). Inhibitors will be dissolved in DMSO and spiked into standard fly food just before solidification, at therapeutic concentrations of drug. Impact of drug administration on neuropathology and clinical outcomes will be evaluated.

[00395] *Human tissue culture models of CTXLP expression*

[00396] CTXLP was detectable in all human cell lines assayed (SVGA, ReNcell CX, NCCIT, T47D, cancer cell line panel), with varying degrees of expression. Based on the data shown above (**Figures 54-58**), ERVK CTXLP-expressing NCCIT teratocarcinoma cells are a viable model for drug screening applications. Additionally, we have developed a transient vector (pcDNA3.1, **Figures 46 and 52**) and a drug (cumate)-inducible lentiviral system (SBI SparQ QM812B-1[163]) allowing for stable overexpression of ERVK CTXLP, SU and Env. By using the feeder-independent pluripotent stem cell line WA09 (WiCell, mTeSR™1/Matrigel™ Platform) we can establish cerebral organoids, as previously described[164]. The above described protocols will form the foundation for generating ERVK CTXLP-expressing cerebral organoids, a human, druggable, three-dimensional brain model. Lentivirus transduction and flow cytometry selection will be used to generate WA09 stem cells containing the previously described cumate-inducible CTXLP vector. To establish a model of ERVK CTXLP-mediated pathology in intact cerebral tissue, these genetically modified WA09 cells will be grown to maximally sized cerebral organoids in the absence of cumate. Both wild-type and CTXLP-inducible cerebral organoids will be treated with varying doses of cumate, to allow for dose-response and time course experiments. Other potential human systems to study CTXLP in the future includes patient-derived human induced pluripotent stem cells systems[165-167].

[00397] In summary, Human tissue and animal models for the study of CTXLP in ALS and cancer are needed. We are actively working to further develop our human tissue culture models. In addition, together with Dr. Alberto Civetta, we are in the process of developing a model in *Drosophilia* at the University of Winnipeg. Importantly, we will continue to pursue mammalian models with our collaborators which offer an opportunity to explore multiple features of pathogenesis as we continue to elucidate the processes involved in CTXLP pathogenesis.

[00398] **Discussion**

[00399] Endogenous retroviruses (ERVs) are host genetic elements, representing approximately 8% of human genomic DNA. ERV activation can benefit their host, or in other contexts are proposed to be involved in pathogenesis and disease[6]. Our interest in ERVK and

the CTXLP protein lies in its association to motor neuron conditions such as *Amyotrophic Lateral Sclerosis* (ALS), as well in cancers.

[00400] ERVK is known to be up-regulated in the neurons of many individuals with ALS[67,89,90,168]. The motor impairment in ALS is linked to calcium channel dysfunction, which is considered a viable therapeutic target. Thus, we were particularly intrigued by the CTXLP region of the ERVK genome when we discovered that it encoded a conotoxin-like peptide indicative of a pathogenic mechanism in ERVK associated disease. O-superfamily conotoxins are known to inhibit voltage gated calcium ion channels[169]. These channels are predominantly expressed at the presynaptic terminal of neural synapses[170]. When an impulse reaches this region, they allow calcium ions into the neuron, thereby increasing calcium ion concentration. This increase in concentration leads to fusion of synaptic vesicles containing neurotransmitters with the presynaptic membrane. The neurotransmitters are then released into the synaptic cleft where they bind to receptors on the postsynaptic terminal and stimulate downstream signaling. Inhibition of these channels with conotoxins can lead to tremors and an inhibition of motor function[23,24]. Our findings indicate that ERVK CTXLP may likewise be able to inhibit VGCCs and their calcium channel-associated transcriptional regulator (CCAT) and elicit similar responses resulting in impaired motor function.

[00401] It may be that ERVK CTXLP was previously implicated in ALS pathology. Notably, a 1997 study found that sera from 5 out of 6 ALS patients was able to reduce calcium ion currents when applied to mouse dorsal root ganglia[171]. The sera from a variety of disease control groups did not exhibit any effect on calcium ion currents. The authors concluded that "serum factors" from ALS patients can be passively transferred to affect calcium ion channel activity. It is possible that ERVK CTXLP may be the mediator of this effect. If this was the case, the protein would either have to be produced in non-brain cells or tissues. ERVK reverse transcriptase has been detected in the serum of many ALS patients[172,173]. If this enzyme originates from ERVK, it would demonstrate that ERVK proteins enter the serum during ALS. Thus, it would be possible that CTXLP could be present in the serum as well. In addition, a compromised blood-brain barrier observed in ALS and other neurology inflammatory events would allow CTXLP to cross into the serum[174].

[00402] Conotoxins are able to specifically inhibit ion channels of certain types of neurons[169], which may correlate with the loss of motor function and neurocognitive decline that is observed in ALS[175]. Other neurotoxin models have been proposed as etiological agents of ALS. The most prominent example is the suspected link between beta-N-methylamino-L-

alanine (BMAA), a neurotoxin produced by a group of terrestrial cyanobacterial symbionts in cycad plants, and ALS (or an ALS-like syndrome)[176]. However, large scale spatial clustering of individuals with ALS has been inconsistent with the range of BMAA-producing cyanobacteria and other suspected environmental risk factors (although small, regional concordances have been identified)[176]. That is, no yet-proposed neurotoxin etiology has been able to explain the vast majority of cases of ALS. Therefore, although an environmental neurotoxin model for ALS makes sense at a physiological level, a genetic-based model (with environmental/epigenetic influence) seems more likely at an epidemiological level. A genetically-encoded neurotoxin etiology of ALS, such as ERVK CTXLP, would be consistent with both of these approaches. This model would not rely on the requirement to identify unique genes in individuals with ALS, as the different phenotypes (having ALS or not having it) could be caused by differential expression of the same genetic material. Additionally, two active ERVK loci unique to ALS patients have been identified[90]. It is possible that the CTXLP proteins of these loci are more functional than other ERVK sequences. Another alternative possibility is that insertional polymorphisms or single nucleotide polymorphisms result in differential ERVK CTXLP production or function. Thus, it is possible that sequence variation in ALS patients leads to differentially functional ERVK CTXLP proteins.

[00403] Additionally, there are several seemingly disparate features of conotoxin toxicity and ALS pathophysiology that would have to be resolved in order for such a link to be possible. For instance, calcium ion concentrations in the neurons of many ALS patients are elevated[175,177]. Since omega-conotoxins inhibit calcium influx into neurons[169], elevated calcium levels are the opposite of what would normally be expected in an ERVK CTXLP etiology of ALS[175]. However, it may be possible that these two features are not inconsistent, as many factors (aside from VGCCs) control neuronal calcium ion concentrations[170]. For instance, calcium-binding proteins such as calbindin-D28K and parvalbumin are absent in motor neurons lost early in ALS[170,177]. These proteins were present in significantly higher concentrations in healthy motor neurons, and in those affected later in the course of the disease[170]. Impaired mitochondrial calcium buffering has also been observed in ALS neurons[170,177].

[00404] Apart from pathology associated with VGCC disruption, we have also shown that CTXLP expression is correlated with elevated Nogo-A in the spinal cord of patients with ALS. Nogo-A is a key regulator of oligodendrocyte precursor cell (OPC) differentiation, ultimately negatively impacting remyelination and tissue repair. Demyelinated spinal cord

lesions show an increased abundance of Nogo-A+ OPCs, yet the inability of OPCs to mature is proposed as the mechanism driving a non-permissive environment leading to remyelination failure[103-107]. Additionally, Nogo-A favours a pro-inflammatory context[123], one that would promote ERVK expression via modulation of NF-κB and pro-inflammatory cytokine secretion[67].

[00405] Nogo-A is implicated in a variety of neurological conditions, such as spinal cord injury, peripheral neuropathies, stroke, temporal lobe epilepsy, Alzheimer's disease, ALS, MS and schizophrenia[101,108-110]. Nogo-A has been identified as a prognostic marker and therapeutic target in ALS due to its substantial expression in muscle tissue from patients with motor neuron disease[111,112]. Mechanistically, Nogo-A expression destabilizes neuromuscular junctions[113-116]. Indeed, clinical trials using human anti-Nogo-A antibodies have been performed (ATI 355 from Novartis Pharma and Ozanezumab and GSK1223249 from GlaxoSmithKline)[101,117,118]. These therapies were designed to target Nogo-A expression in the periphery (intravenous infusions), but may fail to block Nogo-A expression in the CNS, thus explaining the negative results in Phase II clinical ALS trials with Ozanezumab[119,120]. Yet, anti-Nogo-A and remyelination-based therapies may be of value in the treatment of CTXLP+ disease states.

[00406] As ERVK CTXLP is present in the tissues of ALS patients, it may be used as a biomarker for the disease. This is significant given that ALS is often difficult to diagnose in its initial stages[175]. Furthermore, if it is found to be an etiological agent of disease, ERVK CTXLP levels could be useful in assessing disease progression or prognosis. Perhaps most importantly, therapeutics could be designed to target it in order to reduce motor function deficits and increase longevity. For instance, a humanized monoclonal antibody could be designed against ERVK CTXLP for intravenous immunoglobulin (IVIG) therapy. Alternatively, small molecule inhibitors, such as MAEs celastrol and gambogic acid, could be used to target ERVK CTXLP DNA binding, gene transactivation effects, enhancement of NF-κB expression and modulation of pathogenic biomarkers. If ERVK CTXLP is found to play pathological roles in other diseases, for example spinal cord injury, multiple sclerosis, schizophrenia or cancers to name a few, this avenue of research could have implications on the diagnosis and treatment of these diseases as well.

[00407] ERVK expression is up-regulated in schizophrenia and bipolar disorders[178] (unpublished data). This may be worth investigating further if ERVK CTXLP production is confirmed, given the fact that omega-conotoxins can cause emotional distress and prolonged delirium with psychotic features[24,179].

[00408] Additionally, HIV and HTLV infections are known to lead to up-regulation of ERVK expression[180,181]. Both of these infections are associated with poorly understood, reversible ALS-like syndromes in a small number of patients[182-184]. HIV-associated ALS can be treated effectively with highly active antiretroviral therapy (HAART)[182,183]. It is possible that ERVK CTXLP is a pathological contributor to exogenous retrovirus infections and these ALS-like diseases.

[00409] Many cancers are associated with ERVK expression[185]. Evidence that increased ERVK CTXLP expression occurs in cancers cells implicates this viral protein in oncogenesis and possibly metastasis. Specifically, the induction of NF-κB is likely a key feature of ERVK CTXLP activity which may facilitate cancer development and progression[186,187].

[00410] Together, the results of this analysis provide a basis for further research into the ERVK genome and the relationship between ERVK and inflammatory disease. Given the possible correlations between ERVK CTXLP and disease pathology, this line of research deserves further study.

[00411] Figures 63-65 summarize possible implications of our discoveries.

[00412] **References**

[00413] 1. Hohn, O., Hanke, K. & Bannert, N. HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. Frontiers in oncology 3, 246 (2013).

[00414] 2. Lesbats, P., Engelman, A.N. & Cherepanov, P. Retroviral DNA Integration. Chemical reviews (2016).

[00415] 3. Christensen, T. HERVs in neuropathogenesis. J Neuroimmune Pharmacol 5, 326-335 (2010).

[00416] 4. Weiss, R.A. The discovery of endogenous retroviruses. Retrovirology 3, 67 (2006).

[00417] 5. Lokossou, A.G., Toudic, C. & Barbeau, B. Implication of human endogenous retrovirus envelope proteins in placental functions. Viruses 6, 4609-4627 (2014).

[00418] 6. Manghera, M., Ferguson, J. & Douville, R. Endogenous retrovirus-K and nervous system diseases. Curr Neurol Neurosci Rep 14, 488 (2014).

[00419] 7. Macfarlane, C. & Simmonds, P. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. J Mol Evol 59, 642-656 (2004).

[00420] 8. Shin, W., et al. Human-specific HERV-K insertion causes genomic variations in the human genome. PLoS One 8, e60605 (2013).

[00421] 9. Marchi, E., Kanapin, A., Magiorkinis, G. & Belshaw, R. Unfixed endogenous retroviral insertions in the human population. J Virol 88, 9529-9537 (2014).

[00422] 10. Seifarth, W., et al. Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. J Virol 79, 341-352 (2005).

[00423] 11. Griffiths, D.J. Endogenous retroviruses in the human genome sequence. Genome Biol 2, REVIEWS1017 (2001).

[00424] 12. Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E. & Sverdlov, E. At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. J Virol 80, 10752-10762 (2006).

[00425] 13. Ruggieri, A., et al. Human endogenous retrovirus HERV-K(HML-2) encodes a stable signal peptide with biological properties distinct from Rec. Retrovirology 6, 17 (2009).

[00426] 14. Denne, M., et al. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. J Virol 81, 5607-5616 (2007).

[00427] 15. Contreras-Galindo, R., et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. J Virol 86, 262-276 (2012).

[00428] 16. Dewannieux, M., Blaise, S. & Heidmann, T. Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. J Virol 79, 15573-15577 (2005).

[00429] 17. Julien, J.P., et al. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science 342, 1477-1483 (2013).

[00430] 18. Daly, N.L. & Craik, D.J. Bioactive cystine knot proteins. Curr Opin Chem Biol 15, 362-368 (2011).

[00431] 19. Iyer, S. & Acharya, K.R. Tying the knot: the cystine signature and molecular-recognition processes of the vascular endothelial growth factor family of angiogenic cytokines. FEBS J 278, 4304-4322 (2011).

[00432] 20. McNulty, J.C., et al. Structures of the agouti signaling protein. J Mol Biol 346, 1059-1070 (2005).

[00433] 21. Zhu, S., Darbon, H., Dyason, K., Verdonck, F. & Tytgat, J. Evolutionary origin of inhibitor cystine knot peptides. FASEB J 17, 1765-1767 (2003).

[00434] 22. Becker, S. & Terlau, H. Toxins from cone snails: properties, applications and biotechnological production. Appl Microbiol Biotechnol 79, 1-9 (2008).

[00435] 23. Anderson, P.D. Bioterrorism: toxins as weapons. Journal of pharmacy practice 25, 121-129 (2012).

[00436] 24. Obafemi, A. & Roth, B. Prolonged delirium with psychotic features from omega conotoxin toxicity. Pain Med 14, 447-448 (2013).

[00437] 25. Norton, R.S. & Olivera, B.M. Conotoxins down under. Toxicon 48, 780-798 (2006).

[00438] 26. Su, S.C., et al. Regulation of N-type voltage-gated calcium channels and presynaptic function by cyclin-dependent kinase 5. Neuron 75, 675-687 (2012).

[00439] 27. Adams, D.J. & Berecki, G. Mechanisms of conotoxin inhibition of N-type (Ca(v)2.2) calcium channels. Biochim Biophys Acta 1828, 1619-1628 (2013).

[00440] 28. Eldridge, R., Li, Y. & Miller, L.K. Characterization of a baculovirus gene encoding a small conotoxinlike polypeptide. J Virol 66, 6563-6571 (1992).

[00441] 29. Gracy, J., et al. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res 36, D314-319 (2008).

[00442] 30. Kearse, M., et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649 (2012).

[00443] 31. Pettersen, E.F., et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605-1612 (2004).

[00444] 32. Daly, N.L., Rosengren, K.J., Henriques, S.T. & Craik, D.J. NMR and protein structure in drug design: application to cyclotides and conotoxins. Eur Biophys J 40, 359-370 (2011).

[00445] 33. Jiang, L., et al. 111In-labeled cystine-knot peptides based on the Agouti-related protein for targeting tumor angiogenesis. J Biomed Biotechnol 2012, 368075 (2012).

[00446] 34. Li, D., et al. Function and solution structure of hainantoxin-I, a novel insect sodium channel inhibitor from the Chinese bird spider Selenocosmia hainana. FEBS Lett 555, 616-622 (2003).

[00447] 35. Sato, K., et al. Binding of Ala-scanning analogs of omega-conotoxin MVIIC to N- and P/Q-type calcium channels. FEBS Lett 469, 147-150 (2000).

[00448] 36. Jackson, P.J., et al. Structural and molecular evolutionary analysis of Agouti and Agouti-related proteins. Chem Biol 13, 1297-1305 (2006).

[00449] 37. Bagashev, A. & Sawaya, B.E. Roles and functions of HIV-1 Tat protein in the CNS: an overview. Virol J 10, 358 (2013).

[00450] 38. Kruman, II, Nath, A. & Mattson, M.P. HIV-1 protein Tat induces apoptosis of hippocampal neurons by a mechanism involving caspase activation, calcium overload, and oxidative stress. Exp Neurol 154, 276-288 (1998).

[00451] 39. Popper, S.J., et al. Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. J Infect Dis 180, 1116-1121 (1999).

[00452] 40. Dhamija, N., Choudhary, D., Ladha, J.S., Pillai, B. & Mitra, D. Tat predominantly associates with host promoter elements in HIV-1-infected T-cells - regulatory basis of transcriptional repression of c-Rel. FEBS J 282, 595-610 (2015).

[00453] 41. Kim, J., Yoon, J.H. & Kim, Y.S. HIV-1 Tat interacts with and regulates the localization and processing of amyloid precursor protein. PLoS One 8, e77972 (2013).

[00454] 42. Gonzalez-Hernandez, M.J., et al. Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 Tat protein. J Virol 88, 8924-8935 (2014).

[00455] 43. Subramanian, R.P., Wildschutte, J.H., Russo, C. & Coffin, J.M. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology 8, 90 (2011).

[00456] 44. Gifford, R. & Tristem, M. The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26, 291-315 (2003).

[00457] 45. Barbulescu, M., et al. Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Curr Biol 9, 861-868 (1999).

[00458] 46. Schmitt, K., Reichrath, J., Roesch, A., Meese, E. & Mayer, J. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. Genome biology and evolution 5, 307-328 (2013).

[00459] 47. Dinkel, H., et al. ELM 2016--data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res 44, D294-300 (2016).

[00460] 48. Vagner, S., et al. Alternative translation initiation of the Moloney murine leukemia virus mRNA controlled by internal ribosome entry involving the p57/PTB splicing factor. J Biol Chem 270, 20376-20383 (1995).

[00461] 49. Buck, C.B., et al. The human immunodeficiency virus type 1 gag gene encodes an internal ribosome entry site. J Virol 75, 181-191 (2001).

[00462] 50. Lopez-Lastra, M., Rivas, A. & Barria, M.I. Protein synthesis in eukaryotes: the growing biological relevance of cap-independent translation initiation. Biol Res 38, 121-146 (2005).

[00463] 51. Boulikas, T. Putative nuclear localization signals (NLS) in protein transcription factors. J Cell Biochem 55, 32-58 (1994).

[00464] 52. Chang, K.C. Revealing -1 programmed ribosomal frameshifting mechanisms by single-molecule techniques and computational methods. Comput Math Methods Med 2012, 569870 (2012).

[00465] 53. Touriol, C., et al. Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biol Cell 95, 169-178 (2003).

[00466] 54. Buee, L., Bussiere, T., Buee-Scherrer, V., Delacourte, A. & Hof, P.R. Tau protein isoforms, phosphorylation and role in neurodegenerative disorders. Brain Res Brain Res Rev 33, 95-130 (2000).

[00467] 55. Ueki, N., Someya, K., Matsuo, Y., Wakamatsu, K. & Mukai, H. Cryptides: functional cryptic peptides hidden in protein structures. Biopolymers 88, 190-198 (2007).

[00468] 56. Ho, O. & Green, W.R. Cytolytic CD8+ T cells directed against a cryptic epitope derived from a retroviral alternative reading frame confer disease protection. J Immunol 176, 2470-2475 (2006).

[00469] 57. Garrison, K.E., et al. Transcriptional errors in human immunodeficiency virus type 1 generate targets for T-cell responses. Clin Vaccine Immunol 16, 1369-1371 (2009).

[00470] 58. Brinzevich, D., et al. HIV-1 interacts with human endogenous retrovirus K (HML-2) envelopes derived from human primary lymphocytes. J Virol 88, 6213-6223 (2014).

[00471] 59. Terry, S.N., et al. Expression of HERV-K108 envelope interferes with HIV-1 production. Virology 509, 52-59 (2017).

[00472] 60. Hanke, K., et al. Reconstitution of the ancestral glycoprotein of human endogenous retrovirus k and modulation of its functional activity by truncation of the cytoplasmic domain. J Virol 83, 12790-12800 (2009).

[00473] 61. Lemaitre, C., Harper, F., Pierron, G., Heidmann, T. & Dewannieux, M. The HERV-K human endogenous retrovirus envelope protein antagonizes Tetherin antiviral activity. J Virol 88, 13626-13637 (2014).

[00474] 62. Kesidou, E., Lagoudaki, R., Touloumi, O., Poulatsidou, K.N. & Simeonidou, C. Autophagy and neurodegenerative disorders. Neural Regen Res 8, 2275-2283 (2013).

[00475] 63. Jang, G.Y., et al. Transglutaminase 2 suppresses apoptosis by modulating caspase 3 and NF-kappaB activity in hypoxic tumor cells. Oncogene 29, 356-367 (2010).

[00476] 64. Manghera, M. & Douville, R.N. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? Retrovirology 10, 16 (2013).

[00477] 65. Manghera, M., Ferguson, J. & Douville, R. ERVK polyprotein processing and reverse transcriptase expression in human cell line models of neurological disease. Viruses 7, 320-332 (2015).

[00478] 66. Manghera, M., Ferguson-Parry, J. & Douville, R.N. TDP-43 regulates endogenous retrovirus-K viral protein accumulation. Neurobiol Dis 94, 226-236 (2016).

[00479] 67. Manghera, M., Ferguson-Parry, J., Lin, R. & Douville, R.N. NF-kappaB and IRF1 Induce Endogenous Retrovirus K Expression via Interferon-Stimulated Response Elements in Its 5' Long Terminal Repeat. J Virol 90, 9338-9349 (2016).

[00480] 68. Bhat, R.K., et al. Human Endogenous Retrovirus-K(II) Envelope Induction Protects Neurons during HIV/AIDS. PLoS One 9, e97984 (2014).

[00481] 69. Li, W., et al. Human endogenous retrovirus-K contributes to motor neuron disease. Science translational medicine 7, 307ra153 (2015).

[00482] 70. Zhou, J., Callapina, M., Goodall, G.J. & Brune, B. Functional integrity of nuclear factor kappaB, phosphatidylinositol 3'-kinase, and mitogen-activated protein kinase signaling allows tumor necrosis factor alpha-evoked Bcl-2 expression to provoke internal ribosome entry site-dependent translation of hypoxia-inducible factor 1alpha. Cancer Res 64, 9041-9048 (2004).

[00483] 71. Szilagyi, A. & Skolnick, J. Efficient prediction of nucleic acid binding function from low-resolution protein structures. J Mol Biol 358, 922-933 (2006).

[00484] 72. Marsili, G., et al. On the role of interferon regulatory factors in HIV-1 replication. Ann N Y Acad Sci 1010, 29-42 (2003).

[00485] 73. Williams, B.R. Transcriptional regulation of interferon-stimulated genes. Eur J Biochem 200, 1-11 (1991).

[00486] 74. Garcia, J.A., Harrich, D., Pearson, L., Mitsuyasu, R. & Gaynor, R.B. Functional domains required for tat-induced transcriptional activation of the HIV-1 long terminal repeat. EMBO J 7, 3143-3147 (1988).

[00487] 75. Keller, G., Gross, C., Kelleher, M. & Winge, D.R. Functional independence of the two cysteine-rich activation domains in the yeast Mac1 transcription factor. J Biol Chem 275, 29193-29199 (2000).

[00488] 76. Vocero-Akbani, A., Lissy, N.A. & Dowdy, S.F. Transduction of full-length Tat fusion proteins directly into mammalian cells: analysis of T cell receptor activation-induced cell death. Methods Enzymol 322, 508-521 (2000).

[00489] 77. Sheehy, N., et al. Functional analysis of human T lymphotropic virus type 2 Tax proteins. Retrovirology 3, 20 (2006).

[00490] 78. Hardiman, O., et al. Amyotrophic lateral sclerosis. Nat Rev Dis Primers 3, 17071 (2017).

[00491] 79. Eisen, A., et al. Cortical influences drive amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry 88, 917-924 (2017).

[00492] 80. Cappello, V. & Francolini, M. Neuromuscular Junction Dismantling in Amyotrophic Lateral Sclerosis. Int J Mol Sci 18(2017).

[00493] 81. Moisse, K. & Strong, M.J. Innate immunity in amyotrophic lateral sclerosis. Biochim Biophys Acta 1762, 1083-1093 (2006).

[00494] 82. McCombe, P.A. & Henderson, R.D. The Role of immune and inflammatory mechanisms in ALS. Curr Mol Med 11, 246-254 (2011).

[00495] 83. Scotter, E.L., Chen, H.J. & Shaw, C.E. TDP-43 Proteinopathy and ALS: Insights into Disease Mechanisms and Therapeutic Targets. Neurotherapeutics 12, 352-363 (2015).

[00496] 84. Kwong, L.K., Neumann, M., Sampathu, D.M., Lee, V.M. & Trojanowski, J.Q. TDP-43 proteinopathy: the neuropathology underlying major forms of sporadic and familial frontotemporal lobar degeneration and motor neuron disease. Acta Neuropathol 114, 63-70 (2007).

[00497] 85. Leal, S.S. & Gomes, C.M. Calcium dysregulation links ALS defective proteins and motor neuron selective vulnerability. Frontiers in cellular neuroscience 9, 225 (2015).

[00498] 86. Zhou, T., et al. Implications of white matter damage in amyotrophic lateral sclerosis (Review). Mol Med Rep 16, 4379-4392 (2017).

[00499] 87. Tognatta, R. & Miller, R.H. Contribution of the oligodendrocyte lineage to CNS repair and neurodegenerative pathologies. Neuropharmacology 110, 539-547 (2016).

[00500] 88. Poniatowski, L.A., et al. Analysis of the Role of CX3CL1 (Fractalkine) and Its Receptor CX3CR1 in Traumatic Brain and Spinal Cord Injury: Insight into Recent Advances in Actions of Neurochemokine Agents. Mol Neurobiol 54, 2167-2188 (2017).

[00501] 89. Manghera, M., Ferguson-Parry, J. & Douville, R.N. TDP-43 regulates endogenous retrovirus-K viral protein accumulation. Neurobiol Dis 94, 226-236 (2016).

[00502] 90. Douville, R., Liu, J., Rothstein, J. & Nath, A. Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. Ann Neurol 69, 141-151 (2011).

[00503] 91. Hurtado, A.P., Rengifo, A.C. & Torres-Fernández, O. Immunohistochemical Overexpression of MAP-2 in the Cerebral Cortex of Rabies-Infected Mice. International Journal of Morphology 33, 465-470 (2015).

[00504] 92. Monroy-Gomez, J., Santamaria, G. & Torres-Fernandez, O. Overexpression of MAP2 and NF-H Associated with Dendritic Pathology in the Spinal Cord of Mice Infected with Rabies Virus. Viruses 10(2018).

[00505] 93. Timmermann, D.B., Westenbroek, R.E., Schousboe, A. & Catterall, W.A. Distribution of high-voltage-activated calcium channels in cultured gamma-aminobutyric acidergic neurons from mouse cerebral cortex. J Neurosci Res 67, 48-61 (2002).

[00506] 94. McTigue, D.M. & Tripathi, R.B. The life, death, and replacement of oligodendrocytes in the adult CNS. J Neurochem 107, 1-19 (2008).

[00507] 95. Othman, A., et al. Olig1 is expressed in human oligodendrocytes during maturation and regeneration. Glia 59, 914-926 (2011).

[00508] 96. Chong, S.Y. & Chan, J.R. Tapping into the glial reservoir: cells committed to remaining uncommitted. J Cell Biol 188, 305-312 (2010).

[00509] 97. Watkins, T.A., Emery, B., Mulinyawe, S. & Barres, B.A. Distinct stages of myelination regulated by gamma-secretase and astrocytes in a rapidly myelinating CNS coculture system. Neuron 60, 555-569 (2008).

[00510] 98. de Faria, O., Jr., Gonsalvez, D., Nicholson, M. & Xiao, J. Activity-dependent central nervous system myelination throughout life. J Neurochem (2018).

[00511] 99. Fancy, S.P., et al. Dysregulation of the Wnt pathway inhibits timely myelination and remyelination in the mammalian CNS. Genes Dev 23, 1571-1585 (2009).

[00512] 100. Arnett, H.A., et al. bHLH transcription factor Olig1 is required to repair demyelinated lesions in the CNS. Science 306, 2111-2115 (2004).

[00513] 101. Schmandke, A., Schmandke, A. & Schwab, M.E. Nogo-A: Multiple Roles in CNS Development, Maintenance, and Disease. Neuroscientist 20, 372-386 (2014).

[00514] 102. Pernet, V., Joly, S., Christ, F., Dimou, L. & Schwab, M.E. Nogo-A and myelin-associated glycoprotein differently regulate oligodendrocyte maturation and myelin formation. J Neurosci 28, 7435-7444 (2008).

[00515] 103. Miron, V.E., Kuhlmann, T. & Antel, J.P. Cells of the oligodendroglial lineage, myelination, and remyelination. Biochim Biophys Acta 1812, 184-193 (2011).

[00516] 104. Talbott, J.F., et al. Endogenous Nkx2.2+/Olig2+ oligodendrocyte precursor cells fail to remyelinate the demyelinated adult rat spinal cord in the absence of astrocytes. Exp Neurol 192, 11-24 (2005).

[00517] 105. Cafferty, W.B. & Strittmatter, S.M. The Nogo-Nogo receptor pathway limits a spectrum of adult CNS axonal growth. J Neurosci 26, 12242-12250 (2006).

[00518] 106. Kuhlmann, T., et al. Differentiation block of oligodendroglial progenitor cells as a cause for remyelination failure in chronic multiple sclerosis. Brain 131, 1749-1758 (2008).

[00519] 107. Theotokis, P., et al. Time course and spatial profile of Nogo-A expression in experimental autoimmune encephalomyelitis in C57BL/6 mice. J Neuropathol Exp Neurol 71, 907-920 (2012).

[00520] 108. Wojcik, S., Engel, W.K. & Askanas, V. Increased expression of Noga-A in ALS muscle biopsies is not unique for this disease. Acta myologica : myopathies and cardiomyopathies : official journal of the Mediterranean Society of Myology / edited by the Gaetano Conte Academy for the study of striated muscle diseases 25, 116-118 (2006).

[00521] 109. Teng, F.Y. & Tang, B.L. Nogo signaling and non-physical injury-induced nervous system pathology. J Neurosci Res 79, 273-278 (2005).

[00522] 110. McDonald, C.L., Bandtlow, C. & Reindl, M. Targeting the Nogo receptor complex in diseases of the central nervous system. Curr Med Chem 18, 234-244 (2011).

[00523] 111. Dupuis, L., et al. Nogo provides a molecular marker for diagnosis of amyotrophic lateral sclerosis. Neurobiol Dis 10, 358-365 (2002).

[00524] 112. Pradat, P.F., et al. Muscle Nogo-A expression is a prognostic marker in lower motor neuron syndromes. Ann Neurol 62, 15-20 (2007).

[00525] 113. Sui, Y.P., Zhang, X.X., Lu, J.L. & Sui, F. New Insights into the Roles of Nogo-A in CNS Biology and Diseases. Neurochem Res 40, 1767-1785 (2015).

[00526] 114. Bruneteau, G., et al. Endplate denervation correlates with Nogo-A muscle expression in amyotrophic lateral sclerosis patients. Annals of clinical and translational neurology 2, 362-372 (2015).

[00527] 115. Bros-Facer, V., et al. Treatment with an antibody directed against Nogo-A delays disease progression in the SOD1G93A mouse model of Amyotrophic lateral sclerosis. Hum Mol Genet 23, 4187-4200 (2014).

[00528] 116. Jokic, N., et al. The neurite outgrowth inhibitor Nogo-A promotes denervation in an amyotrophic lateral sclerosis model. EMBO Rep 7, 1162-1167 (2006).

[00529] 117. Ineichen, B.V., et al. Nogo-A Antibodies for Progressive Multiple Sclerosis. CNS Drugs 31, 187-198 (2017).

[00530] 118. Meininger, V., et al. Safety, pharmacokinetic, and functional effects of the nogo-a monoclonal antibody in amyotrophic lateral sclerosis: a randomized, first-in-human clinical trial. PLoS One 9, e97803 (2014).

[00531] 119. Meininger, V., et al. Safety and efficacy of ozanezumab in patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Neurol 16, 208-216 (2017).

[00532] 120. Wills, A.M. Blockade of the neurite outgrowth inhibitor Nogo-A in amyotrophic lateral sclerosis. Lancet Neurol 16, 175-176 (2017).

[00533] 121. Amy, M., et al. A common functional allele of the Nogo receptor gene, reticulon 4 receptor (RTN4R), is associated with sporadic amyotrophic lateral sclerosis in a French population. Amyotrophic lateral sclerosis & frontotemporal degeneration 16, 490-496 (2015).

[00534] 122. Buss, A., et al. Sequential loss of myelin proteins during Wallerian degeneration in the human spinal cord. Brain 128, 356-364 (2005).

[00535] 123. Fang, Y., et al. The Nogo/Nogo Receptor (NgR) Signal Is Involved in Neuroinflammation through the Regulation of Microglial Inflammatory Activation. J Biol Chem 290, 28901-28914 (2015).

[00536] 124. Ghosh, S. & Dass, J.F. Study of pathway cross-talk interactions with NF-kappaB leading to its activation via ubiquitination or phosphorylation: A brief review. Gene 584, 97-109 (2016).

[00537] 125. Yang, X.D. & Sun, S.C. Targeting signaling factors for degradation, an emerging mechanism for TRAF functions. Immunol Rev 266, 56-71 (2015).

[00538] 126. Gomez-Ospina, N., Tsuruta, F., Barreto-Chang, O., Hu, L. & Dolmetsch, R. The C terminus of the L-type voltage-gated calcium channel Ca(V)1.2 encodes a transcription factor. Cell 127, 591-606 (2006).

[00539] 127. Gomez-Ospina, N., et al. A promoter in the coding region of the calcium channel gene CACNA1C generates the transcription factor CCAT. PLoS One 8, e60526 (2013).

[00540] 128. Nielsen, K.J., Schroeder, T. & Lewis, R. Structure-activity relationships of omega-conotoxins at N-type voltage-sensitive calcium channels. J Mol Recognit 13, 55-70 (2000).

[00541] 129. Takata, T., et al. Clinical significance of caspase-3 expression in pathologic-stage I, nonsmall-cell lung cancer. Int J Cancer 96 Suppl, 54-60 (2001).

[00542] 130. Pu, X., et al. Caspase-3 and caspase-8 expression in breast cancer: caspase-3 is associated with survival. Apoptosis 22, 357-368 (2017).

[00543] 131. Chen, H., et al. Prognostic value of Caspase-3 expression in cancers of digestive tract: a meta-analysis and systematic review. Int J Clin Exp Med 8, 10225-10234 (2015).

[00544] 132. Wurzer, W.J., et al. Caspase 3 activation is essential for efficient influenza virus propagation. EMBO J 22, 2717-2728 (2003).

[00545] 133. Do, T., et al. Three-dimensional imaging of HIV-1 virological synapses reveals membrane architectures involved in virus transmission. J Virol 88, 10327-10339 (2014).

[00546] 134. Narayan, V., et al. Celastrol inhibits Tat-mediated human immunodeficiency virus (HIV) transcription and replication. J Mol Biol 410, 972-983 (2011).

[00547] 135. Kalantari, P., Narayan, V., Henderson, A.J. & Prabhu, K.S. 15-Deoxy-Delta12,14-prostaglandin J2 inhibits HIV-1 transactivating protein, Tat, through covalent modification. FASEB J 23, 2366-2373 (2009).

[00548] 136. Cascao, R., Fonseca, J.E. & Moita, L.F. Celastrol: A Spectrum of Treatment Opportunities in Chronic Diseases. Front Med (Lausanne) 4, 69 (2017).

[00549] 137. Salminen, A., Lehtonen, M., Paimela, T. & Kaarniranta, K. Celastrol: Molecular targets of Thunder God Vine. Biochem Biophys Res Commun 394, 439-442 (2010).

[00550] 138. Tao, X., Sun, Y. & Zhang, N. [Treatment of rheumatoid arthritis with low doses of multi-glycosides of Tripterygium wilfordii]. Zhong Xi Yi Jie He Za Zhi 10, 289-291, 261-282 (1990).

[00551] 139. Venkatesha, S.H., Astry, B., Nanjundaiah, S.M., Yu, H. & Moudgil, K.D. Suppression of autoimmune arthritis by Celastrus-derived Celastrol through modulation of pro-inflammatory chemokines. Bioorg Med Chem 20, 5229-5234 (2012).

[00552] 140. Fang, Z., et al. High-Throughput Study of the Effects of Celastrol on Activated Fibroblast-Like Synoviocytes from Patients with Rheumatoid Arthritis. Genes 8(2017).

[00553] 141. Paris, D., et al. Reduction of beta-amyloid pathology by celastrol in a transgenic mouse model of Alzheimer's disease. J Neuroinflammation 7, 17 (2010).

[00554] 142. Kashyap, D., et al. Molecular targets of celastrol in cancer: Recent trends and advancements. Critical reviews in oncology/hematology 128, 70-81 (2018).

[00555] 143. Reutrakul, V., et al. Cytotoxic and anti-HIV-1 caged xanthones from the resin and fruits of Garcinia hanburyi. Planta Med 73, 33-40 (2007).

[00556] 144. Banik, K., et al. Therapeutic potential of gambogic acid, a caged xanthone, to target cancer. Cancer Lett 416, 75-86 (2018).

[00557] 145. Kashyap, D., Mondal, R., Tuli, H.S., Kumar, G. & Sharma, A.K. Molecular targets of gambogic acid in cancer: recent trends and advancements. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine 37, 12915-12925 (2016).

[00558] 146. Jang, S.W., et al. Gambogic amide, a selective agonist for TrkA receptor that possesses robust neurotrophic activity, prevents neuronal cell death. Proc Natl Acad Sci U S A 104, 16329-16334 (2007).

[00559] 147. Fu, Q., Li, C. & Yu, L. Gambogic acid inhibits spinal cord injury and inflammation through suppressing the p38 and Akt signaling pathways. Mol Med Rep 17, 2026-2032 (2018).

[00560] 148. Wang, J., Gines, S., MacDonald, M.E. & Gusella, J.F. Reversal of a full-length mutant huntingtin neuronal cell phenotype by chemical inhibitors of polyglutamine-mediated aggregation. BMC Neurosci 6, 1 (2005).

[00561] 149. Chen, S.R., et al. A Mechanistic Overview of Triptolide and Celastrol, Natural Products from Tripterygium wilfordii Hook F. Frontiers in pharmacology 9, 104 (2018).

[00562] 150. Saini, P., Ganugula, R., Arora, M. & Kumar, M.N. The Next Generation Non-competitive Active Polyester Nanosystems for Transferrin Receptor-mediated Peroral Transport Utilizing Gambogic Acid as a Ligand. Scientific reports 6, 29501 (2016).

[00563] 151. Fan, T.K., Gundimeda, U., Mack, W.J. & Gopalakrishna, R. Counteraction of Nogo-A and axonal growth inhibitors by green tea polyphenols and other natural products. Neural Regen Res 11, 545-546 (2016).

[00564] 152. Sepe, M., et al. Proteolytic control of neurite outgrowth inhibitor NOGO-A by the cAMP/PKA pathway. Proc Natl Acad Sci U S A 111, 15729-15734 (2014).

[00565] 153. Yu, X.J., Zhao, Q., Wang, X.B., Zhang, J.X. & Wang, X.B. Gambogenic acid induces proteasomal degradation of CIP2A and sensitizes hepatocellular carcinoma to anticancer agents. Oncol Rep 36, 3611-3618 (2016).

[00566] 154. Wang, J., et al. Gambogic acid-induced degradation of mutant p53 is mediated by proteasome and related to CHIP. J Cell Biochem 112, 509-519 (2011).

[00567] 155. Hughes, T.T., et al. Drosophila as a genetic model for studying pathogenic human viruses. Virology 423, 1-5 (2012).

[00568] 156. Krug, L., et al. Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. PLoS Genet 13, e1006635 (2017).

[00569] 157. Chang, J.C., Hazelett, D.J., Stewart, J.A. & Morton, D.B. Motor neuron expression of the voltage-gated calcium channel cacophony restores locomotion defects in a Drosophila, TDP-43 loss of function model of ALS. Brain Res 1584, 39-51 (2014).

[00570] 158. Estes, P.S., et al. Motor neurons and glia exhibit specific individualized responses to TDP-43 expression in a Drosophila model of amyotrophic lateral sclerosis. Dis Model Mech 6, 721-733 (2013).

[00571] 159. Ghosh, A., et al. Targeted ablation of oligodendrocytes triggers axonal damage. PLoS One 6, e22735 (2011).

[00572] 160. Huang, Y., Ng, F.S. & Jackson, F.R. Comparison of larval and adult Drosophila astrocytes reveals stage-specific gene expression profiles. G3 (Bethesda) 5, 551-558 (2015).

[00573] 161. Civetta, A. & Clark, A.G. Correlated effects of sperm competition and postmating female mortality. Proc Natl Acad Sci U S A 97, 13162-13165 (2000).

[00574] 162. Civetta, A., Montooth, K.L. & Mendelson, M. Quantitative trait loci and interaction effects responsible for variation in female postmating mortality in Drosophila simulans and D. sechellia introgression lines. Heredity (Edinb) 94, 94-100 (2005).

[00575] 163. Mullick, A., et al. The cumate gene-switch: a system for regulated expression in mammalian cells. BMC biotechnology 6, 43 (2006).

[00576] 164. Lancaster, M.A., et al. Cerebral organoids model human brain development and microcephaly. Nature 501, 373-379 (2013).

[00577] 165. Rosati, J., et al. Establishment of stable iPS-derived human neural stem cell lines suitable for cell therapies. Cell death & disease 9, 937 (2018).

[00578] 166. Dimos, J.T., et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 321, 1218-1221 (2008).

[00579] 167. Egawa, N., et al. Drug screening for ALS using patient-specific induced pluripotent stem cells. Science translational medicine 4, 145ra104 (2012).

[00580] 168. Douville, R.N. & Nath, A. Human Endogenous Retrovirus-K and TDP-43 Expression Bridges ALS and HIV Neuropathology. Frontiers in microbiology 8, 1986 (2017).

[00581] 169. Olivera, B.M. & Teichert, R.W. Diversity of the neurotoxic Conus peptides: a model for concerted pharmacological discovery. Mol Interv 7, 251-260 (2007).

[00582] 170. Marambaud, P., Dreses-Werringloer, U. & Vingtdeux, V. Calcium signaling in neurodegeneration. Molecular neurodegeneration 4, 20 (2009).

[00583] 171. Yan, H.D., Lim, W., Lee, K.W. & Kim, J. Sera from amyotrophic lateral sclerosis patients reduce high-voltage activated Ca2+ currents in mice dorsal root ganglion neurons. Neurosci Lett 235, 69-72 (1997).

[00584] 172. MacGowan, D.J., Scelsa, S.N. & Waldron, M. An ALS-like syndrome with new HIV infection and complete response to antiretroviral therapy. Neurology 57, 1094-1097 (2001).

[00585] 173. McCormick, A.L., Brown, R.H., Jr., Cudkowicz, M.E., Al-Chalabi, A. & Garson, J.A. Quantification of reverse transcriptase in ALS and elimination of a novel retroviral candidate. Neurology 70, 278-283 (2008).

[00586] 174. Garbuzova-Davis, S. & Sanberg, P.R. Blood-CNS Barrier Impairment in ALS patients versus an animal model. Frontiers in cellular neuroscience 8, 21 (2014).

[00587] 175. Mitchell, J.D. & Borasio, G.D. Amyotrophic lateral sclerosis. Lancet 369, 2031-2041 (2007).

[00588] 176. Caller, T.A., et al. Spatial clustering of amyotrophic lateral sclerosis and the potential role of BMAA. Amyotroph Lateral Scler 13, 25-32 (2012).

[00589] 177. Appel, S.H., Beers, D., Siklos, L., Engelhardt, J.I. & Mosier, D.R. Calcium: the Darth Vader of ALS. Amyotroph Lateral Scler Other Motor Neuron Disord 2 Suppl 1, S47-54 (2001).

[00590] 178. Frank, O., et al. Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. J Virol 79, 10890-10901 (2005).

[00591] 179. Thompson, J.C., Dunbar, E. & Laye, R.R. Treatment challenges and complications with ziconotide monotherapy in established pump patients. Pain physician 9, 147-152 (2006).

[00592] 180. Toufaily, C., Landry, S., Leib-Mosch, C., Rassart, E. & Barbeau, B. Activation of LTRs from different human endogenous retrovirus (HERV) families by the HTLV-1 tax protein and T-cell activators. Viruses 3, 2146-2159 (2011).

[00593] 181. Vincendeau, M., et al. Modulation of human endogenous retrovirus (HERV) transcription during persistent and de novo HIV-1 infection. Retrovirology 12, 27 (2015).

[00594] 182. Verma, A. & Berger, J.R. ALS syndrome in patients with HIV-1 infection. J Neurol Sci 240, 59-64 (2006).

[00595] 183. Alfahad, T. & Nath, A. Retroviruses and amyotrophic lateral sclerosis. Antiviral Res 99, 180-187 (2013).

[00596] 184. Matsuzaki, T., et al. HTLV-I-associated myelopathy (HAM)/tropical spastic paraparesis (TSP) with amyotrophic lateral sclerosis-like manifestations. J Neurovirol 6, 544-548 (2000).

[00597] 185. Anwar, S.L., Wulaningsih, W. & Lehmann, U. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. Int J Mol Sci 18(2017).

[00598] 186. Vlahopoulos, S.A., et al. Dynamic aberrant NF-kappaB spurs tumorigenesis: a new model encompassing the microenvironment. Cytokine Growth Factor Rev 26, 389-403 (2015).

[00599] 187. Bradford, J.W. & Baldwin, A.S. IKK/nuclear factor-kappaB and oncogenesis: roles in tumor-initiating cells and in the tumor microenvironment. Adv Cancer Res 121, 125-145 (2014).

[00600] References

[00601]

[00602] 1. Eldridge, R., Li, Y. & Miller, L.K. Characterization of a baculovirus gene encoding a small conotoxinlike polypeptide. J Virol 66, 6563-6571 (1992).

[00603] 2. Gracy, J., et al. KNOTTIN: the knottin or inhibitor cystine knot scaffold in 2007. Nucleic Acids Res 36, D314-319 (2008).

[00604] 3. Kearse, M., et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647-1649 (2012).

[00605] 4. Pettersen, E.F., et al. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605-1612 (2004).

[00606] 5. Dinkel, H., et al. ELM 2016--data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res 44, D294-300 (2016).

[00607] 6. Brohawn, D.G., O'Brien, L.C. & Bennett, J.P., Jr. RNAseq Analyses Identify Tumor Necrosis Factor-Mediated Inflammation as a Major Abnormality in ALS Spinal Cord. PLoS One 11, e0160520 (2016).

[00608] 7. Nielsen, K.J., Schroeder, T. & Lewis, R. Structure-activity relationships of omega-conotoxins at N-type voltage-sensitive calcium channels. J Mol Recognit 13, 55-70 (2000).

[00609] 8. Li, W., et al. Human endogenous retrovirus-K contributes to motor neuron disease. Science translational medicine 7, 307ra153 (2015).

[00610] The embodiments described herein are intended to be examples only. Alterations, modifications and variations can be effected to the particular embodiments by those of skill in the art. The scope of the claims should not be limited by the particular embodiments set forth herein, but should be construed in a manner consistent with the specification as a whole.

[00611] All publications, patents and patent applications mentioned in this Specification are indicative of the level of skill those skilled in the art to which this invention pertains and are herein incorporated by reference to the same extent as if each individual publication patent, or patent application was specifically and individually indicated to be incorporated by reference.

[00612] The invention being thus described, it will be obvious that the same may be varied in many ways. Such variations are not to be regarded as a departure from the spirit

and scope of the invention, and all such modification as would be obvious to one skilled in the art are intended to be included within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated polypeptide that comprises or consists of: an amino acid sequence having at least about 90% identity with the amino acid sequence set forth in SEQ ID NO:1 (CSDYGINCSHSYGCCSRSCIALFC).

2. The isolated polypeptide of claim 1, wherein the isolated polypeptide comprises or consists of an amino acid sequence having at least about 95% identity with the amino acid sequence set forth in SEQ ID NO:1.

3. The isolated polypeptide of claim 1 or 2, wherein the isolated polypeptide comprises or consists of the amino acid sequence set forth in SEQ ID NO:1.

4. An isolated nucleic acid molecule comprising a nucleotide sequence encoding a peptide comprising or consisting of an amino acid sequence having at least about 90% identity with the amino acid sequence set forth in SEQ ID NO:1.

5. The isolated nucleic acid molecule of claim 4, wherein the isolated nucleic acid molecule comprises or consists of a nucleotide sequence having at least about 90% identity with the nucleotide acid sequence encoding the polypeptide of SEQ ID NO: 1.

6. The isolated nucleic acid molecule of claim 4, wherein the isolated nucleic acid molecule comprises a nucleotide sequence having at least about 95% identity with the nucleotide acid sequence encoding the polypeptide of SEQ ID NO: 1.

7. A vector comprising the nucleic acid molecule according to any one of claims 4 to 6.

8. A mammalian cell comprising the nucleic acid molecule of any one of claims 4 to 6.

9. The mammalian cell of claim 8, wherein said mammalian cell is a human cell or non-human primate cell.

10. A host cell comprising the nucleic acid molecule of any one of claims 4 to 6.

11. The host cell of claim 10, wherein said host cell is a mammalian cell, an insect cell (such as *Drosophila melanogaster*), a bacteria cell, or a fungal cell.

12. A method for producing the peptide according to any one of claim 1 to 3, comprising: culturing a mammalian cell according to claim 8 or 9, or a host cell of claim 10 or 11, in a culture medium; and isolating the peptide from the mammalian cell of claims 8 or 9, or host cell of claim 10 or 11, or culture medium thereof.

13. An antibody that specifically recognizes the peptide of any one claims 1 to 3.

14. The antibody of claim 11, wherein said antibody is a monoclonal antibody or a polyclonal antibody.

15. A method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

16. A method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology.

17. The method of claim 15 or 16, wherein said condition or disorder is an infectious disease.

18. The method of claim 17, wherein said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

19. The method of claim 15 or 16, wherein said condition or disorder is a neurological disease.

20. The method of claim 19, wherein said neurological disease is amyotrophic lateral sclerosis (ALS), bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

21. The method of claim 15 or 16, wherein said condition or disorder is a cancer.

22. The method of claim 21, wherein said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

23. The method of claim 15 or 16, wherein said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease.

24. The method of any one of claim 15 to 23, wherein said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

25. The method of any one of claims 15 to 23, wherein said active agent comprises a Michael acceptor electrophile (MAE).

26. The method of any one of claims 15 to 23, wherein said active agent comprises gambogic acid.

27. The method of any one of claims 15 to 23, wherein said active agent comprises celastrol.

28. The method of any one of claims 15 to 23, wherein said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

29. The method of any one of claims 15 to 23, where said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

30. The method of any one of claims 15 to 23, further comprising administering a human anti-Nogo-A antibody.

31. The method of any one of claims 15 to 23, wherein said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

32. Use of a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology for treating or preventing conditions or disorders associated with CTXLP in a subject.

33. Use of a therapeutically effective amount of active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology in the manufacture of a medicament for treating or preventing conditions or disorders associated with CTXLP in a subject.

34. Use of a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology for treating or preventing conditions or disorders associated with ERVK in a subject.

35. Use of a therapeutically effective amount of an active agent optionally in a physiological carrier, or a pharmaceutically acceptable salt thereof, wherein the active agent blocks or inhibits CTXLP activity and/or CTXLP associated pathology in the manufacture of a medicament for treating or preventing conditions or disorders associated with ERVK in a subject.

36. The use of any one of claims 32 to 35, wherein said condition or disorder is an infectious disease.

37. The use of claim 36, wherein said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

38. The use of any one of claims 32 to 35, wherein said condition or disorder is a neurological disease.

39. The use of claim 38, wherein said neurological disease is amyotrophic lateral sclerosis (ALS), bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

40. The use of any one of claims 32 to 35, wherein said condition or disorder is a cancer.

41. The use of claim 40, wherein said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

42. The use of any one of claims 32 to 35, wherein said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease.

43. The use of any one of claims 32 to 42, wherein said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

44. The use of any one of claims 32 to 43, wherein said active agent comprises a Michael acceptor electrophile (MAE).

45. The use of any one of claims 32 to 44, wherein said active agent comprises gambogic acid.

46. The use of any one of claims 32 to 44, wherein said active agent comprises celastrol.

47. The use of any one of claims 32 to 43, wherein said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

48. The use of any one of claims 32 to 43, where said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

49. The use of any one of claims 32 to 48, further comprising the use of a human anti-Nogo-A antibody.

50. The use of any one of claims 32 to 49, wherein said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

51. A method for transcriptional activation, comprising contacting a DNA molecule comprising a gene with a peptide of any one of claims 1 to 3.

52. A diagnostic reagent for use in the detection of CTXLP protein in a subject, comprising an antibody of claims 11 or 12.

53. A diagnostic reagent for use in the detection of CTXLP mRNA in a subject, comprising an isolated nucleic acid according to any one of claims 4 to 6.

54. A diagnostic reagent for use in the detection CTXLP activity in a subject, comprising a peptide of any one of claims 1 to 3.

55. A method for treating or preventing conditions or disorders associated with CTXLP in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier or a pharmaceutically acceptable salt thereof when the amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

56. A method for treating or preventing conditions or disorders associated with ERVK in a subject, comprising: measuring an amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA; and administering to a subject in need thereof a therapeutically effective amount of an active agent optionally in a physiological carrier or a pharmaceutically acceptable salt thereof when the amount of CTXLP polypeptide, or CTXLP activity, or CTXLP mRNA, is high, optionally compared to a control, wherein the active agent blocks or inhibits the CTXLP activity and/or CTXLP associated pathology.

57. The method of claim 55 or 56, wherein said condition or disorder is an infectious disease.

58. The method of claim 57, wherein said infection disease is HSV infection, HIV infection, EBV infection, HTLV infection, toxoplasma gondii infection, HSV infection, or prion disease.

59. The method of claim 55 or 56, wherein said condition or disorder is a neurological disease.

60. The method of claim 59, wherein said neurological disease is amyotrophic lateral sclerosis, bipolar disorder, Kennedy's disease, multiple sclerosis, or schizophrenia.

61. The method of claim 55 or 56, wherein said condition or disorder is a cancer.

62. The method of claim 61, wherein said cancer is breast cancer, chronic myelogenous leukemia, colon cancer, gastric cancer, germ cell tumours, germinogenic tongue tumours, gonadoblastomas, hepatocellular carcinoma, adenocarcinoma, epithloid carcinoma, Acute T-cell leukemia, leukemia, lymphoma, T-cell lymphoma, Burkitt's lymphoma, neuroepithelioma, melanoma, myelodysplastic syndrome, nasopharyngeal carcinoma, ovarian cancer, pancreatic cancer, prostate cancer, testicular cancer, lung cancer, stomach cancer, skin cancer, trophoblastic tumours, tumorigenesis (e.g., via AR interaction), thyroid adenoma, or ERVK in cancerous tissues.

63. The method of claim 55 or 56, wherein said associated pathology is a change in CNS function of said subject, a developmental disorder, a stroke, Alzheimer's disease, spinal cord injury, cerebral ischemia, Huntington's disease, Parkinson's disease, a peripheral neuropathy, or epilepsy ocular disease.

64. The method of any one of claim 55 to 63, wherein said active agent a small molecule, an antibody, a nucleic acid, an aptamer, or a peptide.

65. The method of any one of claims 55 to 63, wherein said active agent comprises a Michael acceptor electrophile (MAE).

66. The method of any one of claims 55 to 63, wherein said active agent comprises gambogic acid.

67. The method of any one of claims 55 to 63, wherein said active agent comprises celastrol.

68. The method of any one of claims 55 to 63, wherein said active agent is a small molecule inhibitor of HIV Tat, for example a Michael acceptor electrophile (MAE) such as curcumin, rosmarinic acid, gambogic acid, celastrol (15-deoxy-Δ(12,14)-prostaglandin J(2) (15d-PGJ(2)), cyclopentenone prostaglandins (CyPG), such as 15-deoxy-Delta(12,14)-PGJ(2) (15d-PGJ(2)), N-acetylcysteine amide (NACA), or D-penicillamine (also called Cuprimine); a sulfhydryl compound with chelating properties such as N-(2-Mercapto-propionyl)-glycin (MPG), 2,3-Dimercapto-propanol (DMP), 2,3-Dimercapto-propane-sulfonic acid (DMPS), Nitric oxide (NO), or sulphated polysaccharides; or a Thioredoxin reductase 1 (TRR1) inhibitor, such as B5 (curcumin analog).

69. The method of any one of claims 55 to 63, where said active agent is a small molecule or antibody reversing CTXLP blockade on oligodendrocyte precursor cell maturation and oligodendrocyte myelination, such as clemastine fumarate.

70. The method of any one of claims 55 to 69, further comprising administering a human anti-Nogo-A antibody.

71. The method of any one of claims 55 to 70, wherein said active agent is a small molecule enhancer of CaV2.2 and its calcium channel associated transcription regulator (CaV2.2 CCAT) expression or activity, such as EGTA, or glutamate.

72. The method of any one of claims 55 to 71, wherein the amount of CTXLP polypeptide is determined using an antibody of claim 13 or 14.

73. A kit comprising:(a) a container comprising a pharmaceutical composition containing the peptide of any one of claims 1 to 3, and/or a nucleic acid according to any one of claim 4 to 6, and/or a vector of claim 7, a mammalian cell of claims 8 or 9, a host cell of claims 10 or

11, and/or an antibody of claim 13 or 14, in solution or in lyophilized form;(b) optionally, a second container containing a diluent or reconstituting solution for the lyophilized formulation; and(c) optionally, instructions for use.

74. The kit according to claim 73, further comprising one or more of (iii) a buffer, (iv) a diluent, (v) a filter, (vi) a needle, or (v) a syringe.

75. A method identifying CTXLP inhibitors, comprising: contacting a mammalian cell or a host cell, such as an insect cell (such as *Drosophila melanogaster*), a bacteria cell, or a fungal cell, with a test compound or test composition, and measuring an amount of CTXLP protein, CTXLP-mRNA, CTXLP-regulated gene, or CTXLP-associate biomarker.

76. A method identifying CTXLP inhibitors, comprising: contacting a organoid, with a test compound or test composition, and measuring an amount of CTXLP protein, CTXLP-mRNA, CTXLP-regulated gene, or CTXLP-associate biomarker

FIG 1

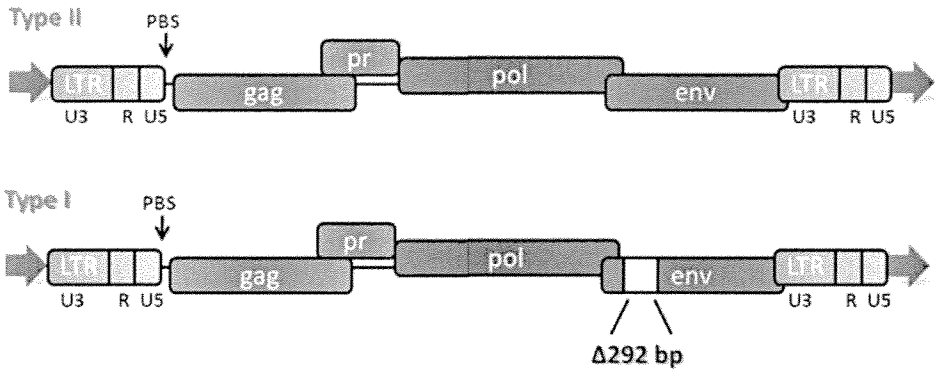


Figure 1: Two types of ERVK genomes. The ERVK genome consists of four main Retroviridae genes, which are from 5' to 3': *gag*, *pro*, *pol*, and *env*. These viral genes are flanked by long terminal repeats (LTRs) containing U3, R and U5 regions. Two types of ERVK genomes can be distinguished based on a 292 bp deletion in the *env* gene.

FIG 2

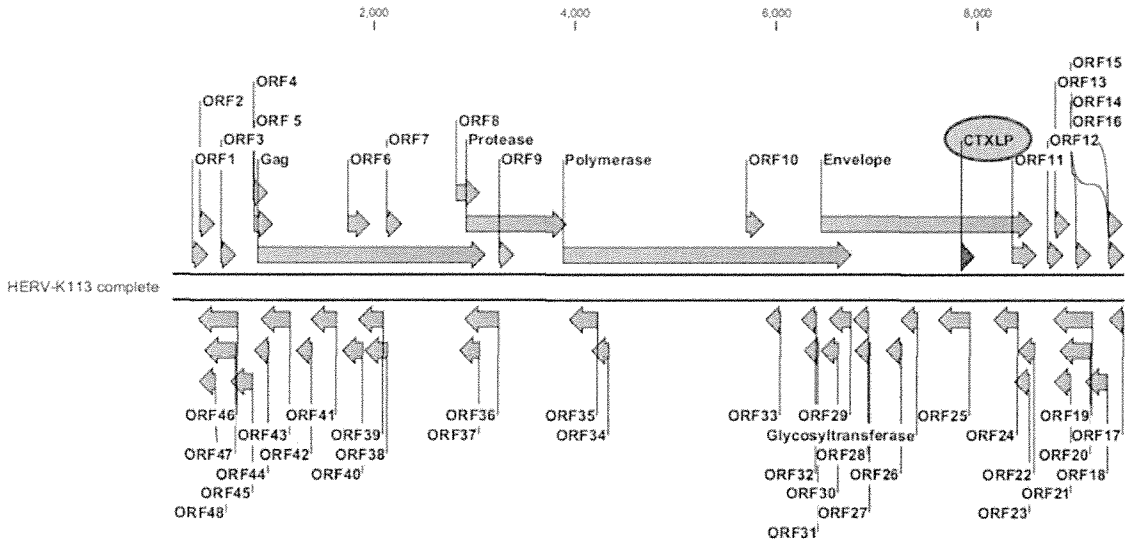


Figure 2: **Open reading frames on both strands of the endogenous retrovirus K-113 genome**. ORFs bigger than 50 nucleotides (yellow) on both the sense and antisense strands were predicted using CLCbio software. The ERVK conotoxin-like protein (CTXLP) ORF is indicated in red. Any amino acid-encoding codon was accepted as an ORF start, although each ended with a stop codon. Note the overlapping ORFs within known ERVK genes, such as *gag*, *protease*, *polymerase* and *envelope*.

FIG 3

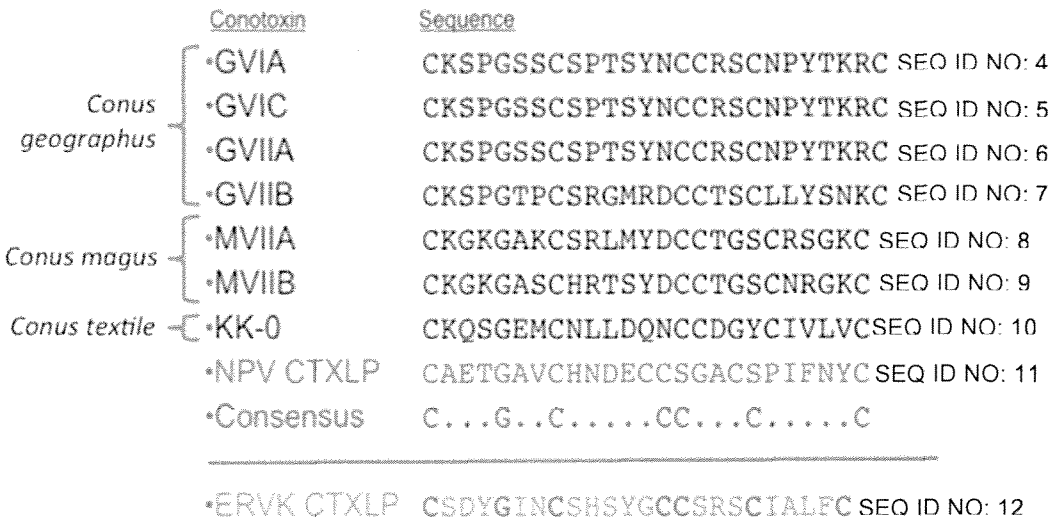


Figure 3: Alignment of cone snail and viral omega conotoxin domain sequences. Sequences from 3 *Conus* species (black), 1 conotoxin-like protein domain sequence from *Autographa Californica* Nuclear Polyhedrosis Virus (blue), the consensus sequence generated from the aforementioned sequences (red) and the sequence of the putative Endogenous Retrovirus K-113 conotoxin-like protein domain (green). Modified from[1]. Note the characteristic C-C-CC-C-C knottin folding motif[2].

FIG 4

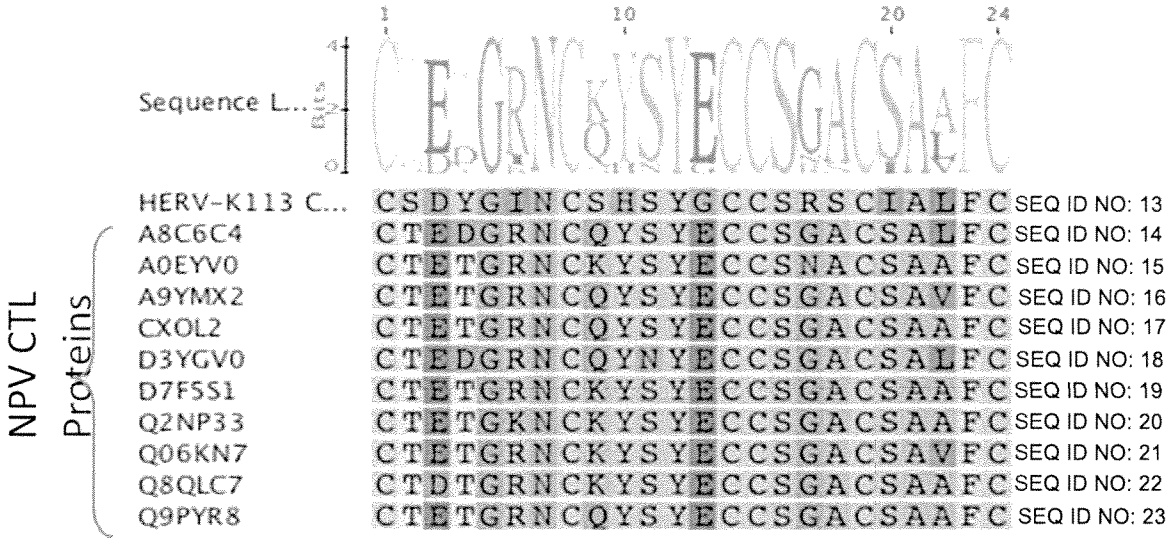


Figure 4: Alignment and sequence logo of the putative endogenous retrovirus K-113 conotoxin-like protein domain and 10 Nuclear Polyhedrosis Virus conotoxin-like protein domain sequences. Sequences were aligned and sequence logo was assessed using Geneious v5 software[3]. Note the conserved C-G-NC-Y-CCS-C-A-FC sequence logo in these viral conotoxin-like proteins.

FIG 5

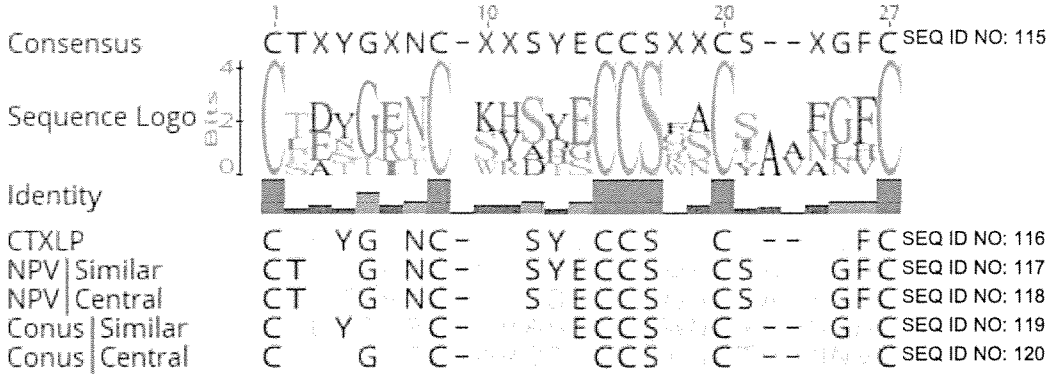


Figure 5: ERVK CTXLP cysteine-rich motif has strong similarity to both nuclear polyhedrosis virus (NPV, 46.2%) and *Conus* (45.8%) conotoxin proteins. Nuclear Polyhedrosis Virus (NPV) and Cone snail conotoxin sequences which are typical ("Central") or which most closely resemble the ERVK conotoxin-like protein encoded in the human genome ("Similar") were identified in Uniprot and hg38 with reference to the protein database Pfam and the repetitive element database Repbase. These were aligned to the Pfam model of the Conotoxin family and the alignment was trimmed to the knottin domain. For details about which search algorithms and measures of centrality were used, see the methods section below. *Methods:* ERVK sequences were identified with ORFfinder in DNA corresponding to entries in UCSC genome browser track rmskJoinedBaseline, derived from Repbase, whose "name" field matches "%#LTR/ERVK". ORFs matching PF08087 were taken to be CTXLP encoding and these were clustered with cd-hit with default settings, identifying the major variant designated "CTXLP". Cone snail sequences were obtained from Uniprot using the search terms "conus conotoxin". Nuclear Polyhedrosis Virus (NPV) sequences were obtained from Uniprot using the search terms "conotoxin taxonomy:'dsDNA viruses, no RNA stage [35237]'". The results of each search were clustered using cd-hit with default settings. The cluster representative sequences for each group were searched using hmmsearch for hits against the models in Pfam clan CL0083. The sequences corresponding to PF02950 (Conotoxin) and PF08087 (toxin_18) with the highest average pairwise percent identity to all other matches for that HMM were designated "Central". BLAST databases were constructed for sequences matching each HMM, and these were queried with CTXLP using blastp. The result with the lowest e-value from each database was designated "Similar". The two "Central" and two "Similar" sequences, along with CTXLP, were aligned to PF02950 using hmmalign, and this alignment was trimmed to the Knottin domain.

FIG 6

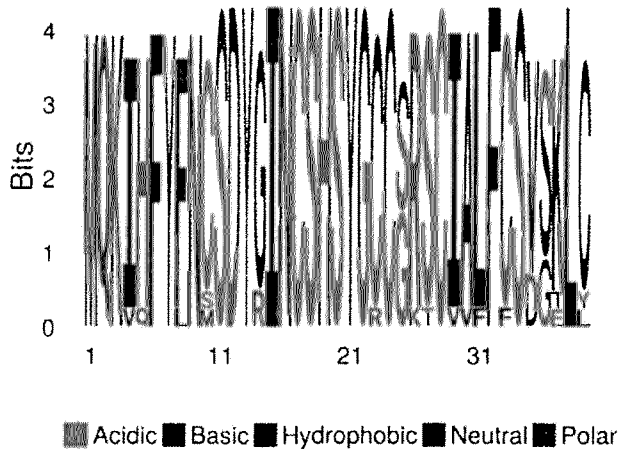


Figure 6: Logo of the knottin domain from cluster representative CTXLP sequences. ERVK CTXLP has 7 cysteines. The amino acid composition includes 1/39 acidic, 5/39 basic, 11/39 hydrophobic, 19/39 polar and 3/39 neutral residues. The bitscores represent the meaningfulness of particular amino-acid residues given their conservation and the composition of the sequences overall. This was produced from a0053/data/ctxlp_variants.cluster.aligned.fasta using the R packages "msa" to load the alignment and "ggseqlogo" to produce the figure.

FIG 7

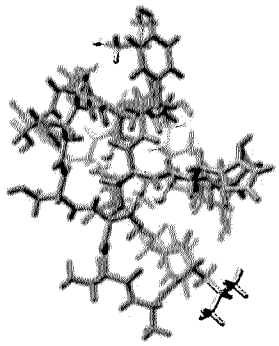


Figure 7: Modeled 3-dimensional structure of the putative Endogenous Retrovirus-K113 conotoxin-like protein domain. Protein tertiary structure was predicted using Knotter1D3D software (gray sticks = carbon, green sticks = hydrogen, red sticks = oxygen, blue sticks = nitrogen and yellow spheres = sulfur). Note the interactions of the yellow cysteine residues, as they form disulfide bonds.

FIG 8

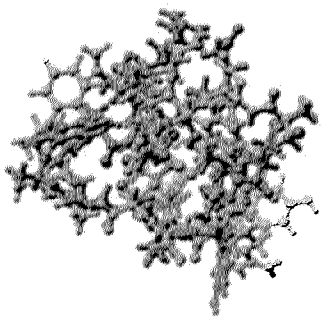


Figure 8: Aligned overlap of the predicted structures of viral conotoxin-like proteins from ERVK-113 and *Ecotropis obliqua* NPV. Knotter1D3D was used to predict the structures of putative ERVK-113 CTXLP domain (blue) and *Ecotropis obliqua* NPV CTXLP domain (red). Structure alignment is based on sequence alignment and was prepared using UCSF Chimera software[4].

FIG 9

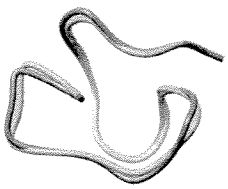


Figure 9: Aligned overlap of the predicted structures of viral conotoxin-like protein backbones from ERVK-113 and *Ecotropis obliqua* NPV. Knotter1D3D was used to predict the structures of putative ERVK-113 CTXLP domain (blue) and *Ecotropis obliqua* NPV CTXLP domain (red). Structure alignment is based on sequence alignment and was prepared using UCSF Chimera software[4].

FIG 10

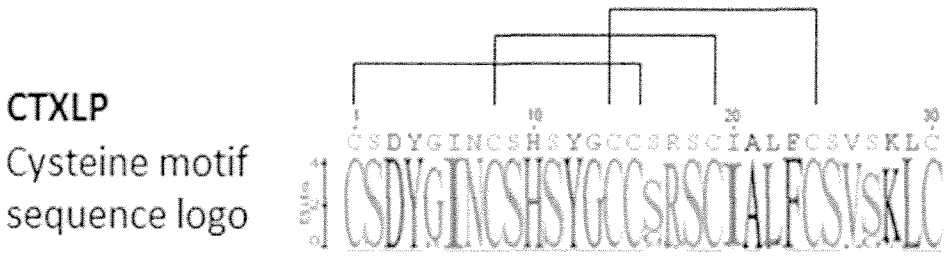


Figure 10: **Predicted inhibitor cysteine knot fold of ERVK CTXLP cysteine-rich peptide.** Disulfide bonds connect cysteine 1 to cysteine 4, cysteine 2 to cysteine 5, and cysteine 3 to cysteine 6, resulting in an inhibitor cysteine knot fold.

FIG 11

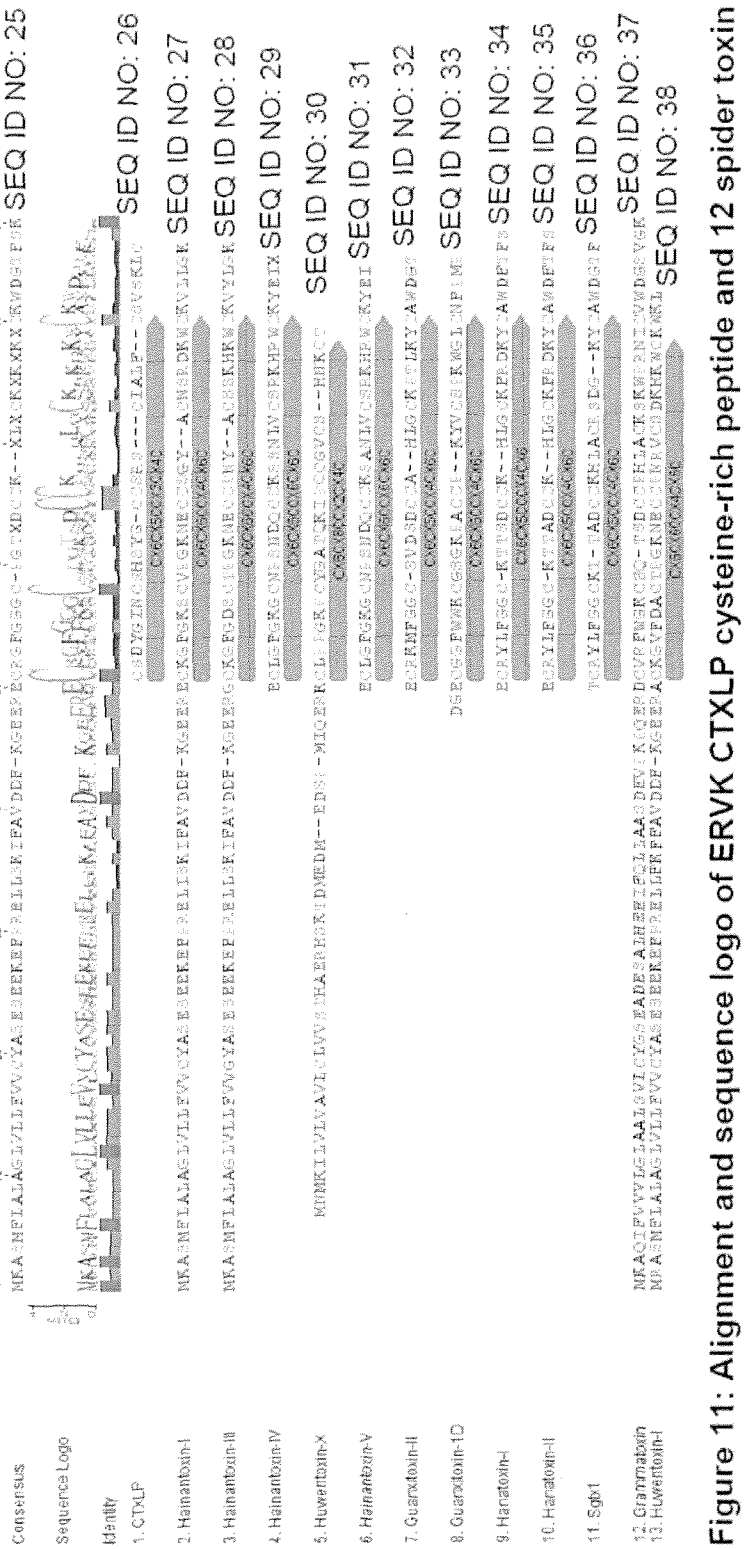


Figure 11: Alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and 12 spider toxin ICK peptides. Sequences were aligned and sequence logo generated using Geneious Software. A conserved C-C-CC-C-C motif is observed in all sequences. CTXLP, all Hainantoxin and one Guanxitoxin contained a conserved G between the first and second cysteine, as indicated by the star. Grey callouts indicate the cysteine motif spacing of each toxin.

FIG 12

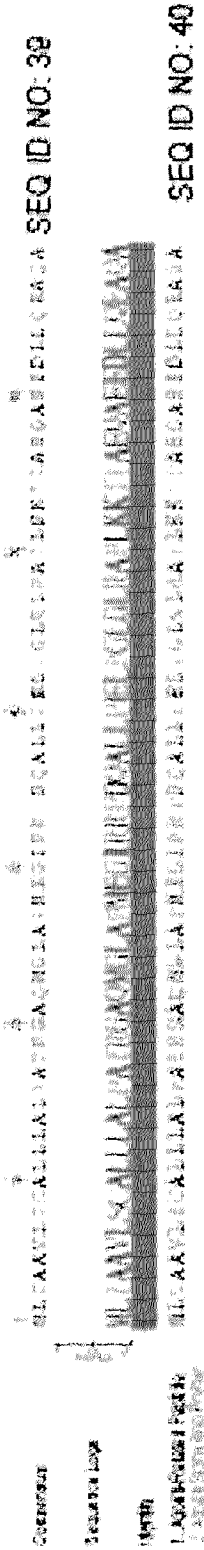


Figure 12: Alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and agouti-related peptide and agouti signalling protein. Sequences were aligned, and sequence logo generated using Geneious Software. A conserved C-C-CC-C-C-C is observed in all sequences.

FIG 13

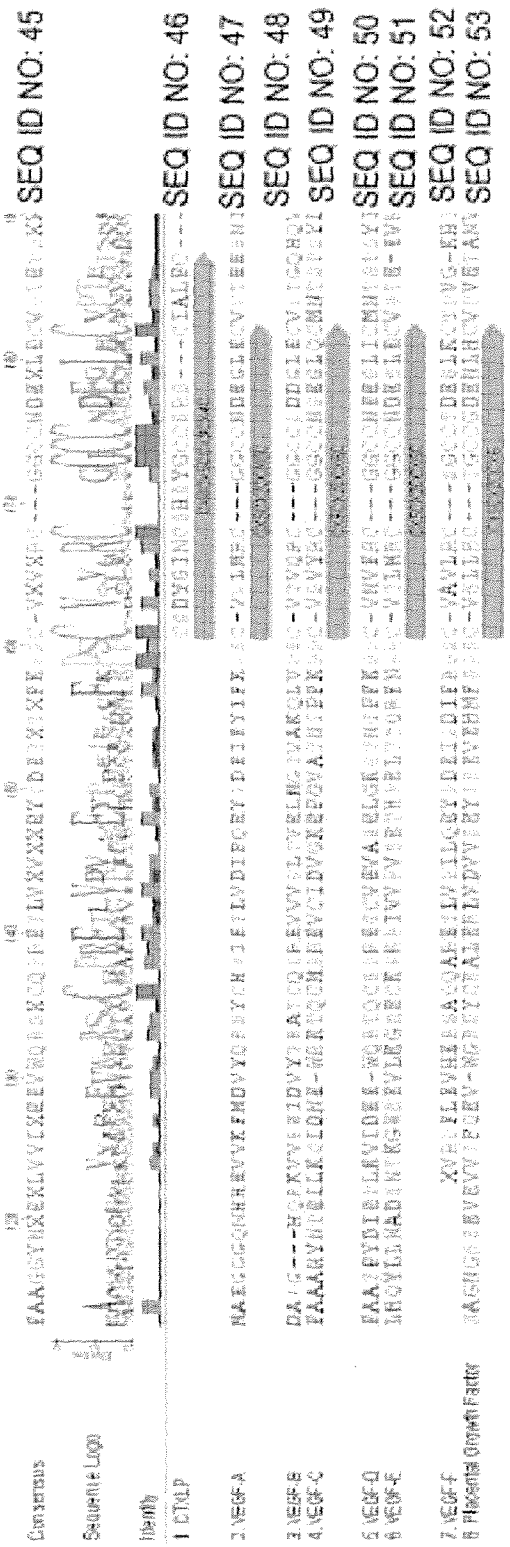


Figure 13: Alignment and sequence logo of ERVK CTXLP cysteine-rich peptide to 7 VEGF Proteins. Sequences were aligned, and sequence logo generated using Geneious Software. No significant conservation was identified between the VEGF proteins and CTXLP, due to the differences in spacing and total number of cysteine residues, as indicated by the grey bar cysteine spacing motif.

FIG 14

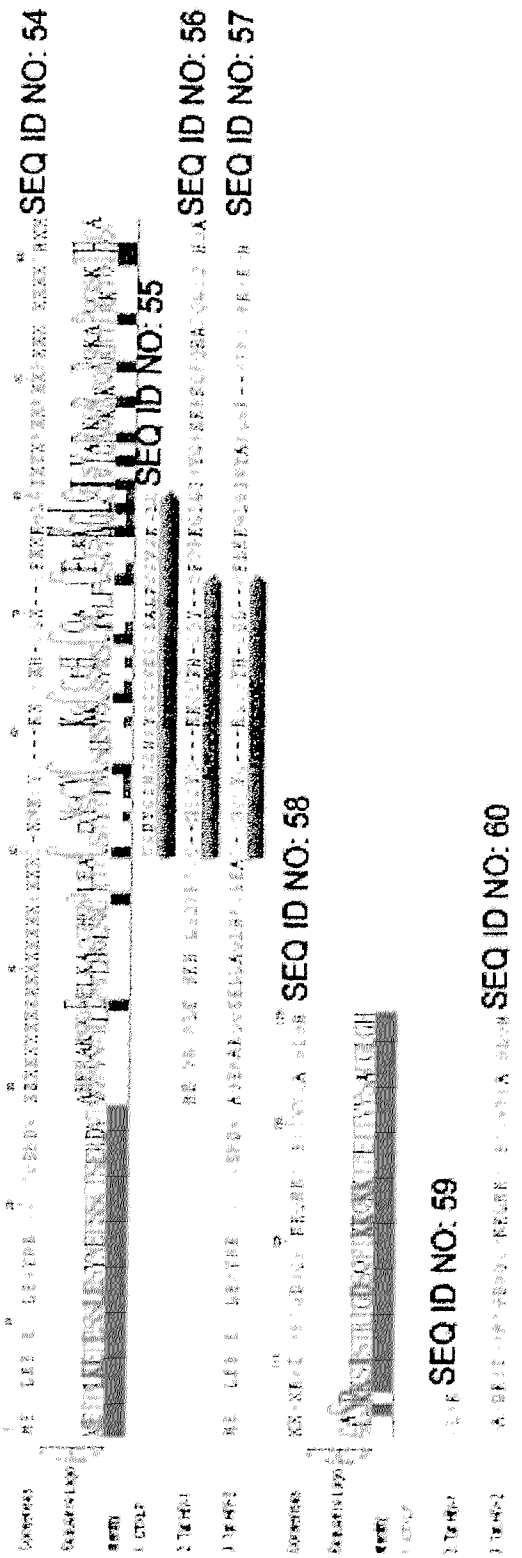


Figure 14: Alignment and sequence logo of ERVK CTXLP cysteine-rich peptide and HIV-1 and HIV-2 Tat proteins. Sequences were aligned, and sequence logo generated using Geneious Software. Conservation of 6 of the 7 CTXLP cysteine residues are found in HIV Tat, as well as 1 lysine and 1 leucine residue in the C terminus, between amino acids 75 and 80.

FIG 15

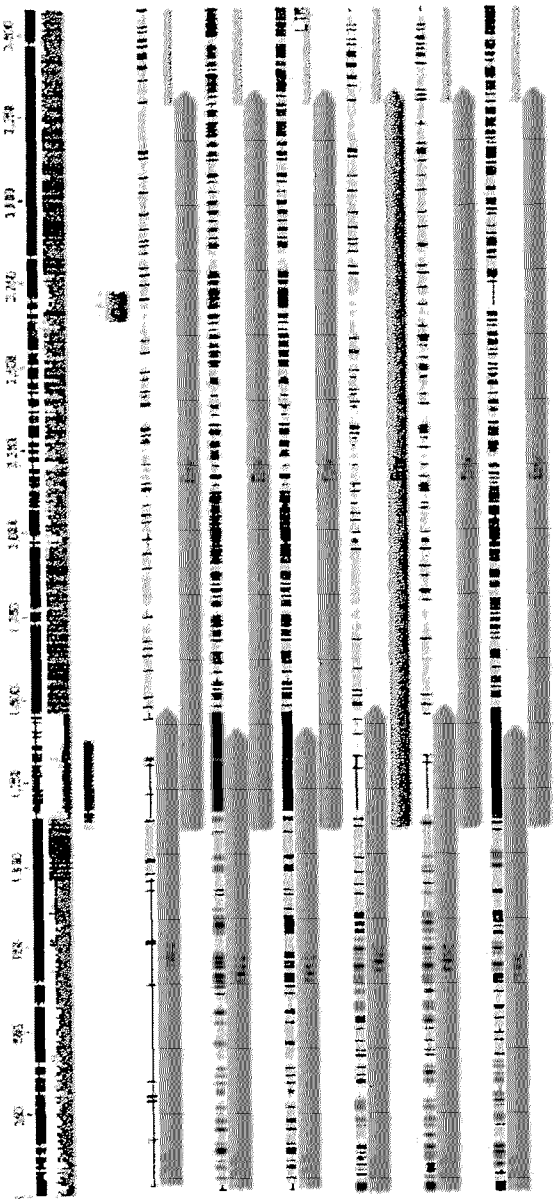


Figure 15: Example Alignment of Chromosome 1 ERVK HML-2 insertions with the DNA sequences for Rec exon 1 and CTXLP. Sequences were aligned, and sequence logo generated using Geneious software. Rec exon 1 aligned with bp 1 to 261 of *env*. CTXLP aligned with bp 1413 to 1505 of *env*.

FIG 16

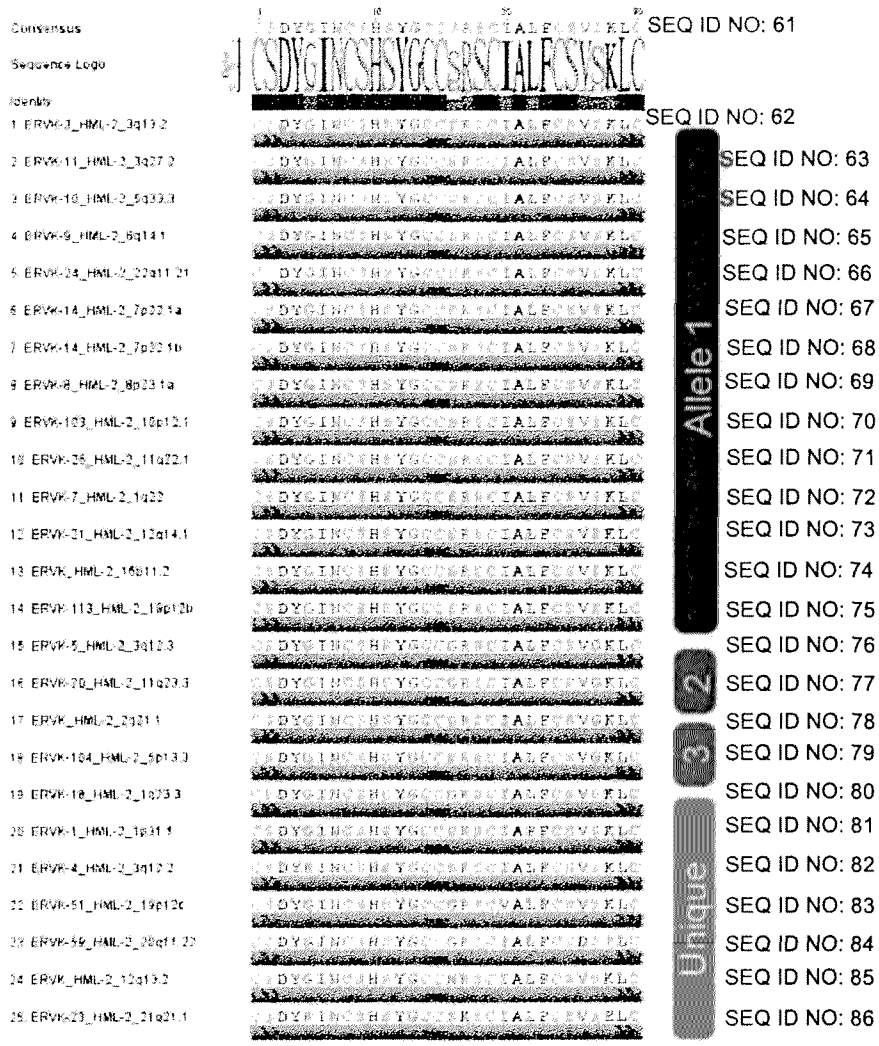


Figure 16: Alignment of conotoxin-like peptides of 25 human ERVK HML-2 insertions. Alignment and sequence logo generated using Geneious Software. CTXLP-peptides showed variability in the *amino acid sequence. Three distinct alleles were identified, as well as several unique sequences.*

FIG 17

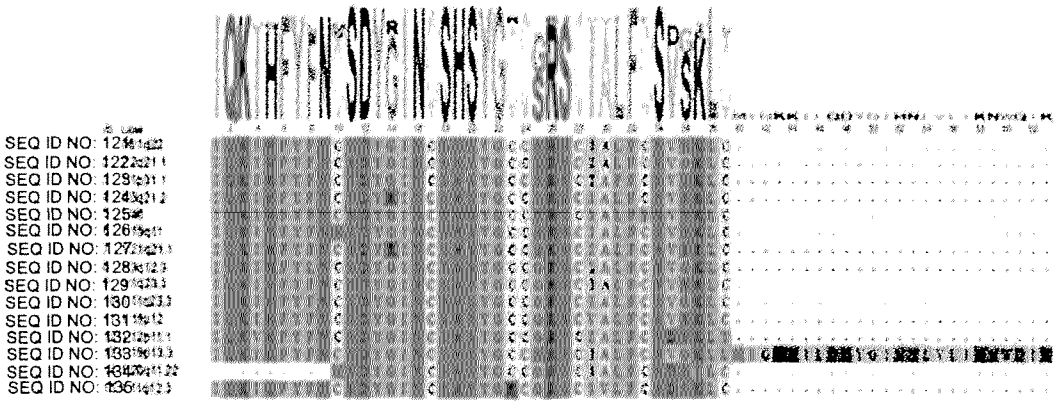


Figure 17: CTXLP variants in the humans, based on genome build GRCh38. Translated open reading frames longer than 59 bp identified in human ERVK loci annotated by RepeatMasker were searched for PF08087 by HMMER. The matching sequences were clustered using cd-hit. In the alignment above we see that the first variant is the most common, with 16 copies, followed by the second with 2. All the following ones are present as unique copies. The last three sequences with a deletion or a mutated cysteine residue are likely to be non-functional.

FIG 18

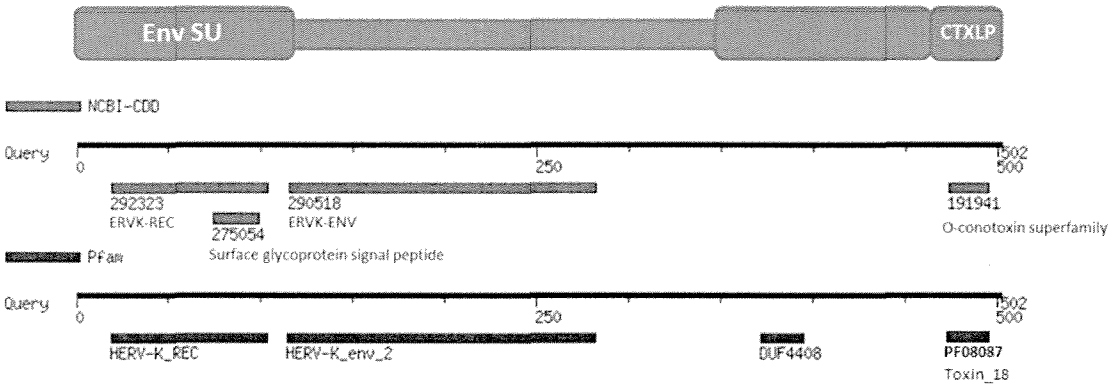


Figure 18: Schematic representation of CTXLP and amino acid sequence similarities found using NCBI-CDD and Pfam databases. The SU subunit of CTXLP is red and the omega conotoxin domain is in green. The wider portions of the diagram represent the ordered regions of the proteins and the narrow region represent the disordered region as predicted by ELM resource[5].

FIG 19

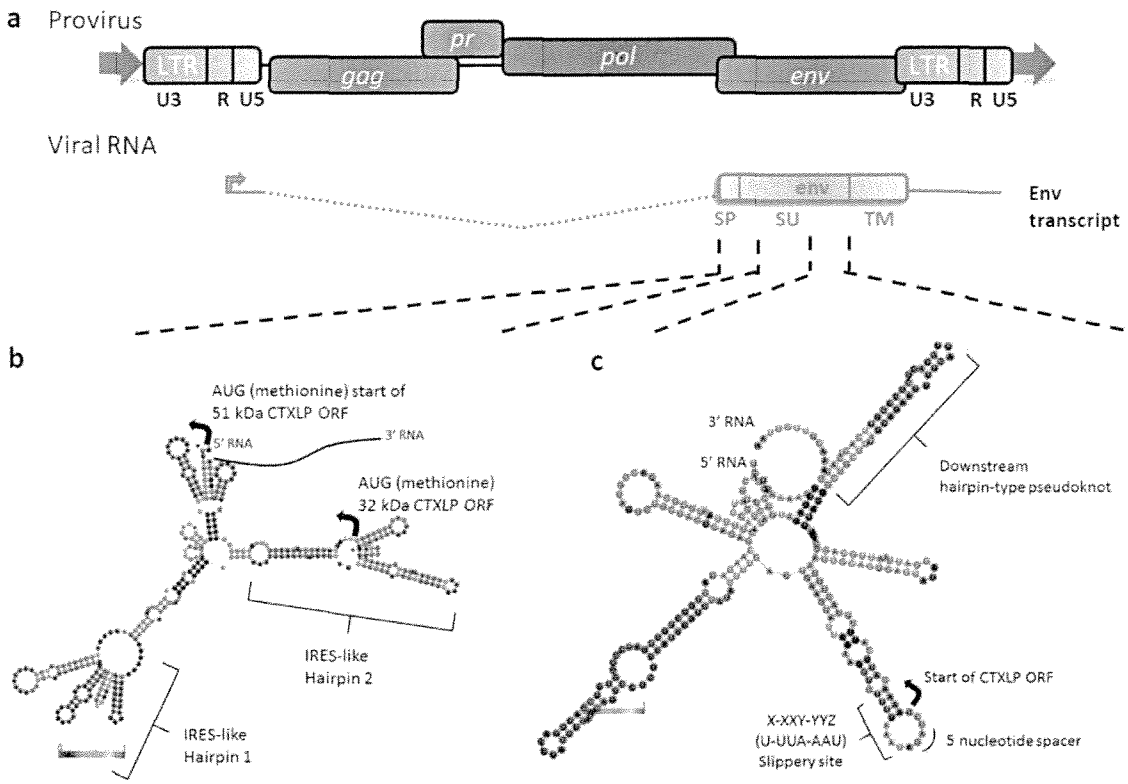


Figure 19: Analysis of the ERVK envelope transcript reveals prototypic RNA secondary structures. A) Formation of the ERVK envelope (Env) transcript from an ERVK provirus. The ERVK Env transcript encodes the viral surface unit (SU) and transmembrane (TM) proteins. **B) Predicted IRES-like RNA hairpin structures in the ERVK *env* transcript.** The first 350 bp of ERVK Env-encoding RNA contains numerous AUG (methionine) translational start sites. Two distinct IRES-like hairpins are identified at nucleotides 84-187 and 213-318. tRNA can potentially bind at the AUG start site identified in IRES-like hairpin to produce a smaller isoform of ERVK Env or CTXLP. **C) Predicted RNA secondary structure for ERVK-4 *env* transcript upstream and including the CTXLP ORF.** Directly upstream of the CTXLP ORF translational start is a conserved -1 programmed ribosomal frameshifting sequence, which contains three elements i) a slippery site containing an X-XXY-YYZ motif which after frameshifting by -1 results in XXX-YYY reading, ii) a 5 to 10 nucleotide spacer sequence, and iii) a downstream hairpin-type pseudoknot. The ERVK *env* transcript slippery site is encoded by a U-UUA-AAU sequence, followed by a 5 nucleotide spacer. Structure prediction formed with RNAfold Software shows a strong probability of hairpin-loops forming within the CTXLP-coding region, likely providing a downstream hairpin-type pseudoknot.

FIG 20

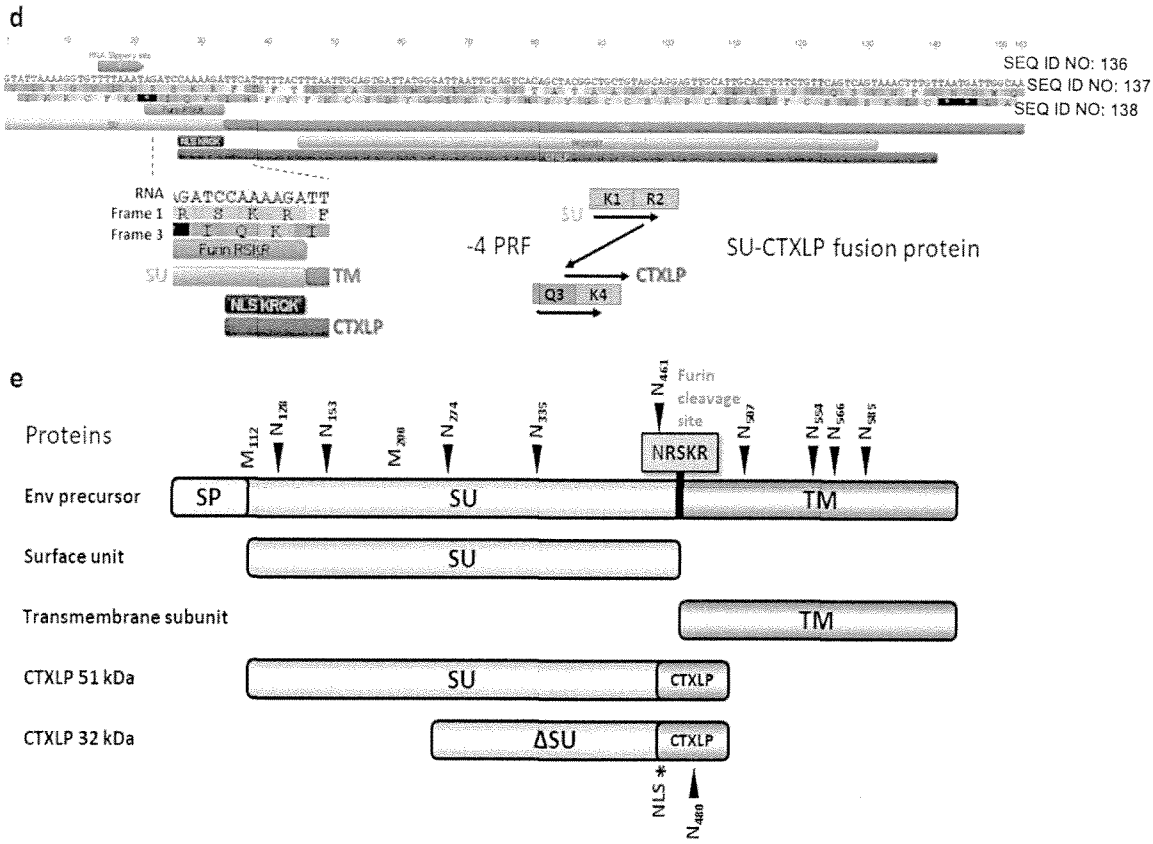


Figure 20: Predicted CTXLP isoforms derived from the ERVK envelope transcript. D) Ribosomal frameshifting event resulting in formation of CTXLP peptide fused to surface unit protein. At the RNA slippery site within the ERVK Env transcript, the ribosome translating the RNA may bounce back by 4 nucleotides and begin reading in an alternate frame. This introduces a canonical KRQK nuclear localization sequence before proceeding into the CTXLP peptide. The resulting protein is a modified SU-CTXLP fusion protein. **E) Examination of the known and predicted protein products derived from the ERVK Env transcript.** It is well established that ERVK can produce canonical retroviral surface unit (SU) and transmembrane (TM) proteins. We show that ERVK can also produce SU-CTXLP fusion proteins. These CTXLP isoforms contain a nuclear localization sequence (NLS) and an additional N-linked glycosylation site at position 480.

FIG 21

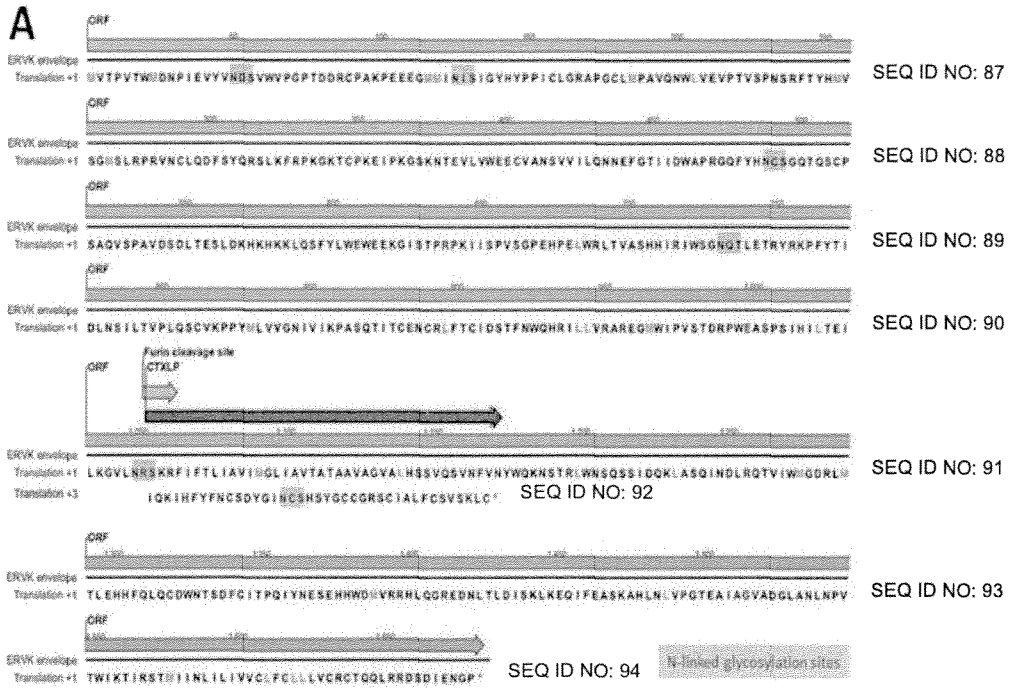
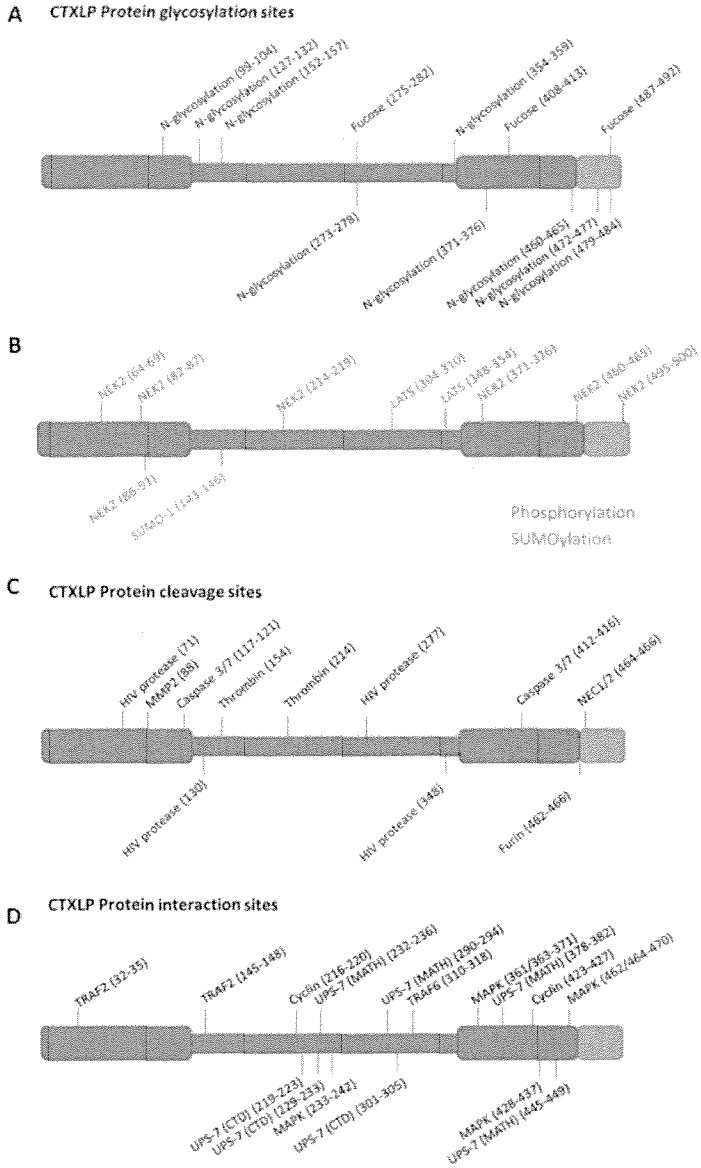


Figure 21: Bioinformatic identification of CTXLP in the genome of endogenous retrovirus-K. ERVK113 was used as a template for the CTXLP domain in the ERVK envelope gene. The ERVK envelope polyprotein is cleaved by the cellular protease furin downstream of the R-X-R/K-R site. This splits the ERVK Env polyprotein into the surface unit (SU) and transmembrane (TM) proteins which interact to form the viral spike protein on the surface of virions. A -4 programmed ribosomal frameshift (PRF) allows for the translation of the CTXLP cysteine-rich motif (red) at the C-terminal end of the SU protein. Post-translational modification of ERVK SU protein includes glycosylation. N-linked N-X-S/T glycosylation sites are identified in red boxes. Alternative methionine (M) start sites are indicated in green.

FIG 22

Figure 22: Predicted post-translational modifications and protein interactions of CTXLP. (A)



Schematic diagram of predicted glycosylation sites. (B)Schematic diagram of predicted phosphorylation and SUMOylation sites. (C) Schematic diagram of predicted protein cleavage sites. (D) Schematic diagram of predicted protein interaction sites.

FIG 23

Antigen Profiler

Custom Antigen Information	
Designed For:	Renee Douville_O-conotoxin
Sequence:	Length: 39 amino acids.

Protein Sequence and Features

Antigen Peptide Candidates

Name	Peptide	Position	Length
OBS-3:8	KIHFYF SEQ ID NO: 96	3	6
OBS-5:10	HFYFNC SEQ ID NO: 97	5	6
OBS-10:15	CSDYGI SEQ ID NO: 98	10	6
OBS-11:16	SDYGIN SEQ ID NO: 99	11	6
OBS-11:17	SDYGINC SEQ ID NO: 100	11	7
OBS-12:17	DYGINC SEQ ID NO: 101	12	6
OBS-18:23	SHSYGC SEQ ID NO: 102	18	6
OBS-18:24	SHSYGCC SEQ ID NO: 103	18	7
OBS-18:25	SHSYGCCS SEQ ID NO: 104	18	8
OBS-19:24	HSYGCC SEQ ID NO: 105	19	6
OBS-19:25	HSYGCCS SEQ ID NO: 106	19	7
OBS-20:25	SYGCCS SEQ ID NO: 107	20	6
OBS-22:27	GCCSRS SEQ ID NO: 108	22	6
OBS-25:30	SRSCIA SEQ ID NO: 109	25	6
OBS-26:31	RSCIAL SEQ ID NO: 110	26	6
OBS-27:32	SCIALF SEQ ID NO: 111	27	6
OBS-31:36	LFCSVS SEQ ID NO: 112	31	6
OBS-31:39	LFCSVSKLC SEQ ID NO: 113	31	9
OBS-34:39	SVSKLC SEQ ID NO: 114	34	6

Figure 23: Antigenic profile of the ERVK CTXLP domain, and predicted epitopes.

FIG 24

Fisher Catalog #	Description	Qty	Unit	Unit Price	Lot Price
SPECIAL #1	**Thermo Scientific Pierce 2 Rabbit 90-Day Protocol, SYN PEP** Thermo Scientific Pierce No. HAB2010 • Project #: FOF1260 • Peptide name: O-Conotoxin • Conjugation: 4MG, KLH, EG • Estimated project time: 12 weeks • Elisa titers: bleeds • Hold bulk bleed: yes • Affinity purification: yes • Special instructions: Provided seq. Linear pep, not oxid. • Peptide sequence: IQKIHFYFNCSDYGINCSHSYGCCSRSCIALFCSVSK LC	1	EA	$1,912.60	$1,912.60

Figure 24: Rabbit immunization protocol for generation of a polyclonal antibody against the ERVK CTXLP domain.

FIG 25

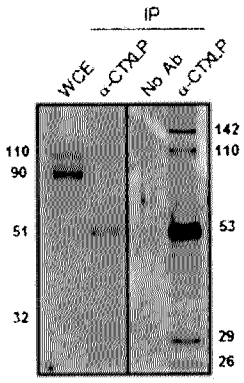


Figure 25: NCCIT spontaneously express ERVK CTXLP protein. NCCIT whole cell extract and immunoprecipitated CTXLP-enriched fractions were run on Western blot CTXLP was detected with custom rabbit anti-CTXLP antibody. The far-right lane in the image, is an image of the left α-CTXLP lane that was over-exposed upon blot development to bring out the details in the bands.

FIG 26

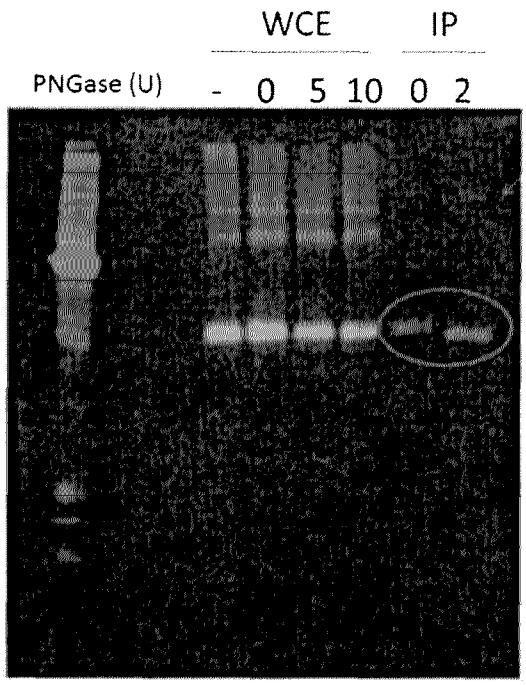


Figure 26: PNGase treatment of IP-purified CTXLP protein results in a decrease in western blot band size associated with removal of N-linked glycosylation moieties. Whole cell extracts and CTXLP enriched by immunoprecipitation were treated with the deglycosylation enzyme PNGase for 1 hour at 37°C. Treated proteins were analysed using western blot. PNGase-treated proteins exhibit a decrease in band size as compared to control indicative of deglycosylation of CTXLP protein isoforms.

FIG 27

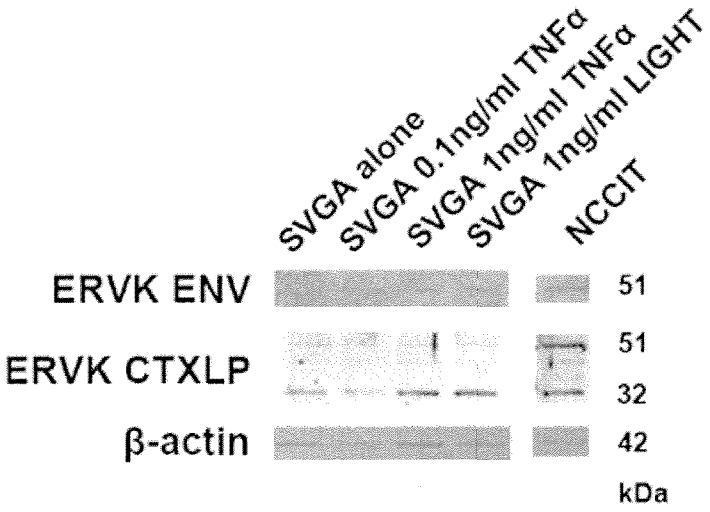


Figure 27: CTXLP expression is inducible in SVGA cells upon exposure to pro-inflammatory stimuli. SVGA cells were cultured in medium alone, TNFα (0.1 ng/ml or 1 ng/ml) or LIGHT (1 ng/ml) for 24 hours. Untreated NCCIT cells were used as a positive culture control. Cell lysates were run on Western blot and incubated with anti-CTXLP antibody. The blot was also incubated anti-β-actin as the loading control (One representative of three independent experiments shown).

FIG 28

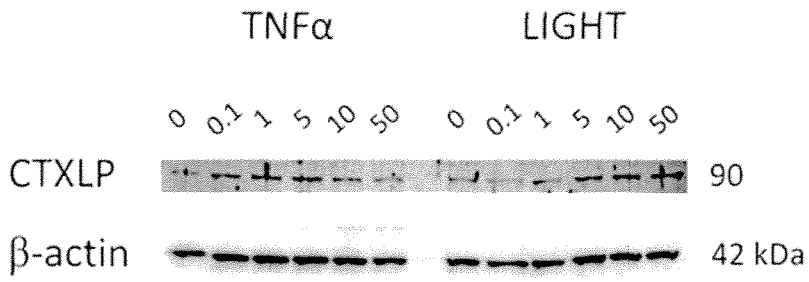


Figure 28: ERVK CTXLP is inducible in human neurons. ReNcell-derived neurons were treated with increasing doses of pro-inflammatory cytokines TNFα or LIGHT for 24 hours. CTXLP expression (90kDa form) was enhanced optimally with 1ng/ml TNFα and 10ng/ml LIGHT treatment, n=1.

FIG 29

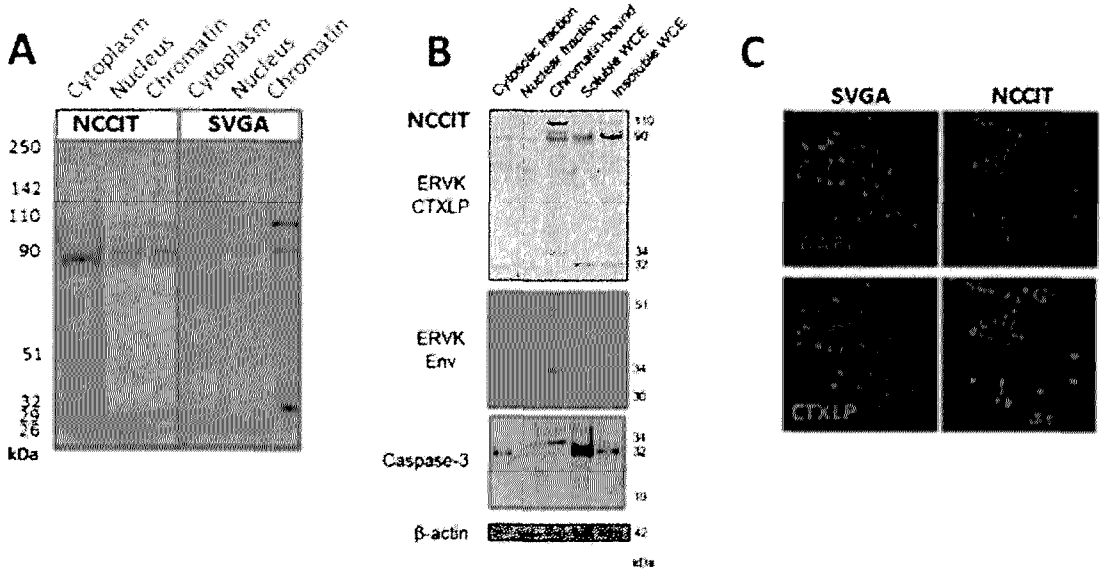


Figure 29: CTXLP protein expression predominantly localizes with chromatin in NCCIT and SVGA cells. NCCIT and SVGA cells in the absence of stimuli, were subject to cytoplasmic, nuclear and chromatin cell fractionation. (**A**) NCCIT cells exhibited CTXLP expression in all cell lysate fractions (n=4). (**B**) Additional NCCIT cells cultures were subject to further cell fractionation. CTXLP expression was again identified within all cell fractions, along within soluble and insoluble whole cell lysates. Whereas, ERVK Env 51 kDa isoform was identified in whole cell lysates but not cell fractions, with the exception of a 34kDa band in the chromatin fraction (coinciding with the observed 34 kDa CTXLP band). Caspase-3 expression was predominantly in the cytosolic fraction and the soluble whole cell lysates. β-actin was used as a loading control. (**C**) Moreover, CTXLP appeared dispersed throughout NCCIT cells in confocal imaging. In contrast, SVGA cells exhibited expression of the small (32kDa) and large (90-110kDa) isoforms of CTXLP in association with the chromatin (**A**). This was supported by nuclear localization of CTXLP in SVGA cells as shown by confocal microscopy (**C**).

FIG 30

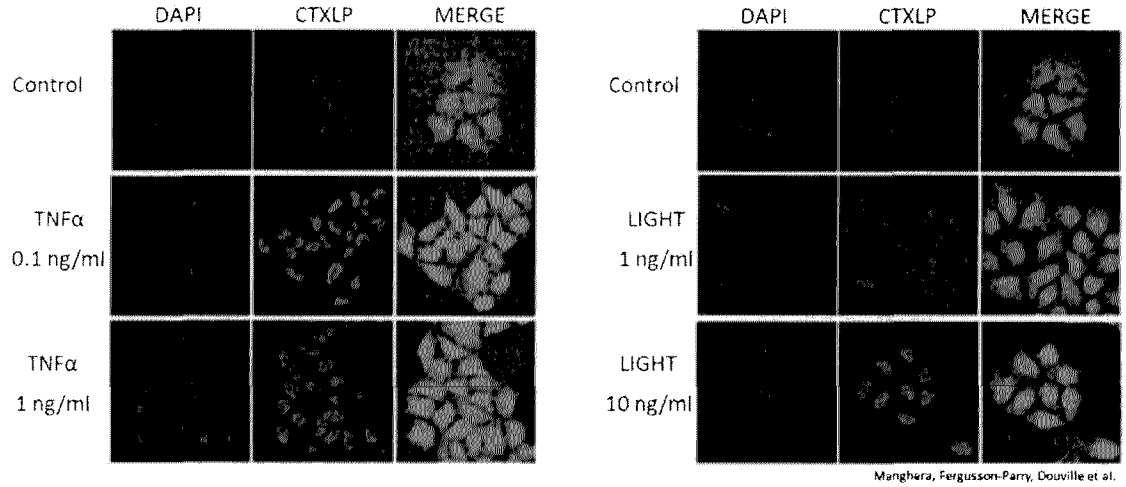


Figure 30: ERVK CTXLP is inducible in astrocytes with pro-inflammatory cytokines. Confocal analysis of protein expression for ERVK CTXLP (red) and ERVK reverse transcriptase (RT, green) in cells treated with or without TNFα and LIGHT, n=2. Enhanced CTXLP expression precedes increases in RT expression. DAPI stain indicates nuclei (blue).

FIG 31

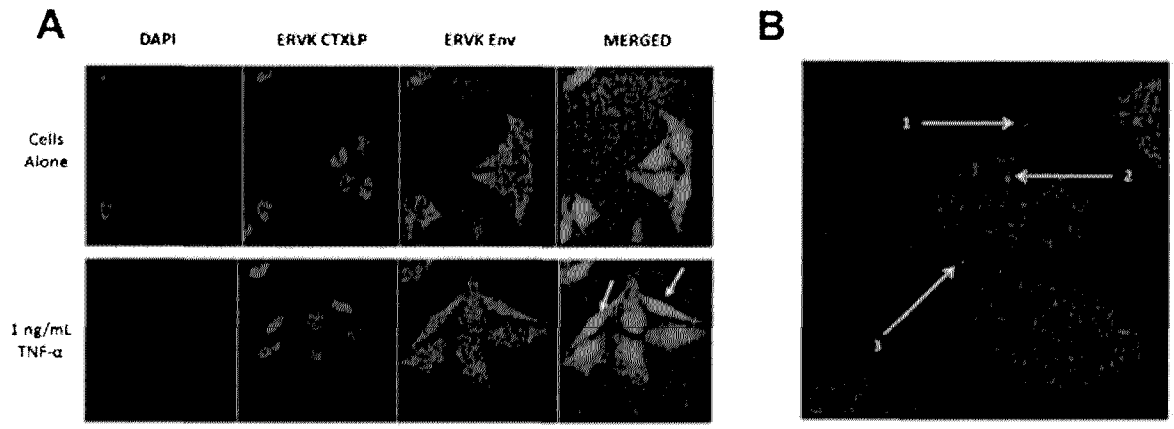


Figure 31: Cellular localization of ERVK CTXLP and SU proteins in human astrocytes. (**A**) SVGA cells were cultured with medium alone or in the presence of 0.1 ng/mL TNFα for 24 h and analyzed by confocal (400x magnification). Increased CTXLP protein (indicated by arrows) and Env expression were identified (but not co-localized) in TNFα-treated cells compared to controls. (**B**) SVGA cells were cultured with medium alone (not shown) or in the presence of 10 ng/mL LIGHT for 24 h and analyzed by confocal (600x magnification). Distinct puncta (indicated by arrows) were identified within the 1) cytoplasm, 2) nucleus, and 3) surface membrane of the cell.

FIG 32

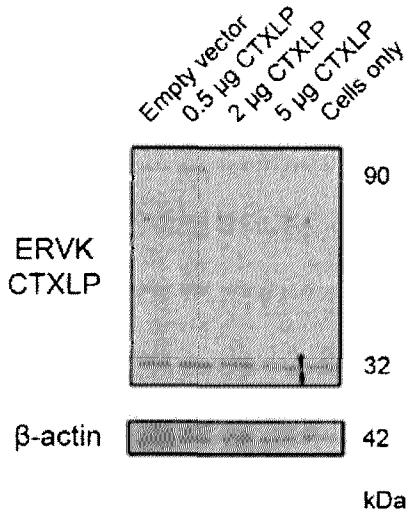


Figure 32: CTXLP overexpression in soluble fractions of SVGA cells enhances the 32 kDa and 90 kDa forms of endogenous CTXLP. SVGA cells were transfected using Lipofectamine LTX with 5 µg Empty Vector, or 0.5, 2 and 5 µg CTXLP cysteine-rich construct for 48 hr. β-actin was used as a loading control. Representative image of n=3.

FIG 33

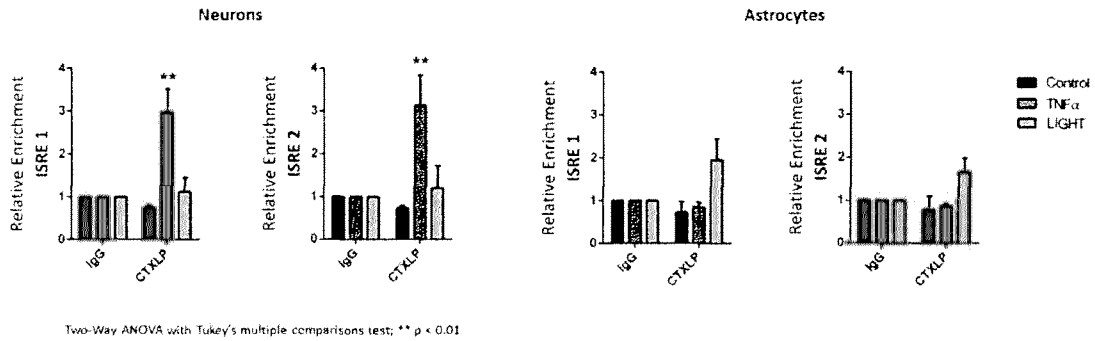


Figure 33: CTXLP binds interferon response elements (ISREs) within the ERVK promoter (5' LTR). CTXLP may regulate ERVK gene expression, as well as other genes containing ISREs. Chromatin immunoprecipitation (ChIP) following 8 hours of 10 ng/ml TNFα or LIGHT treatment in human ReNcell-derived neurons (n=2) and human astrocytic cell line (SVGA) (n=3). Notable increase in CTXLP chromatin binding in neurons upon pro-inflammatory stimulation with the cytokine TNFα.

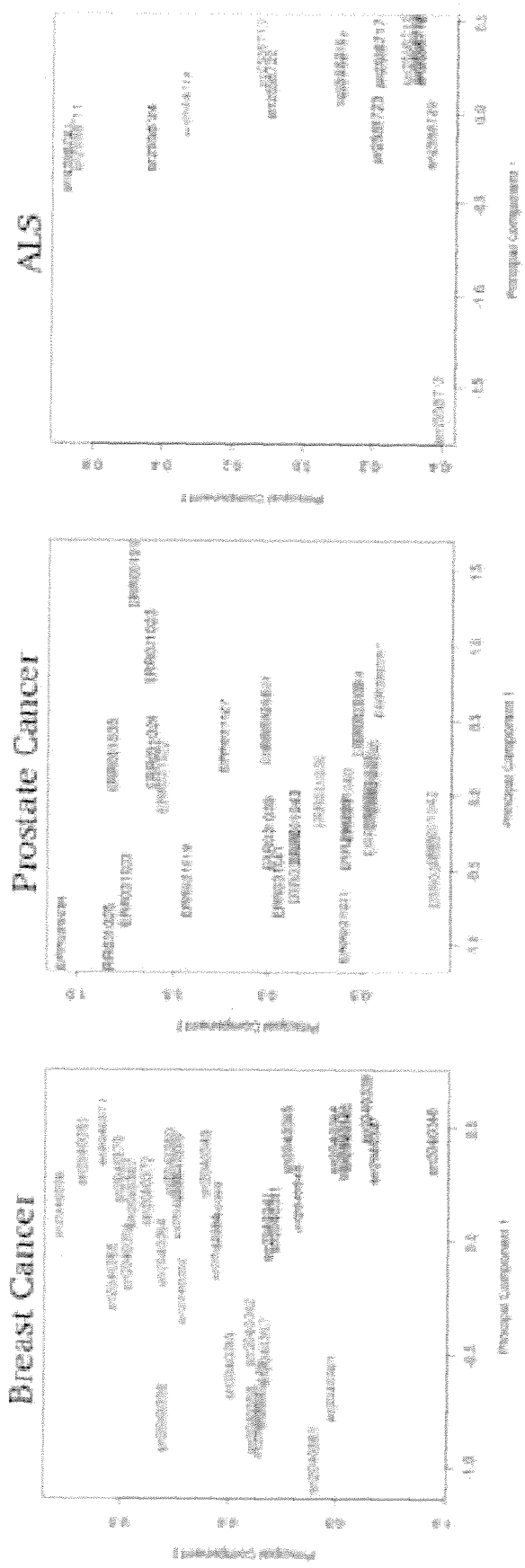


FIG. 34

FIG. 34 CONT'D

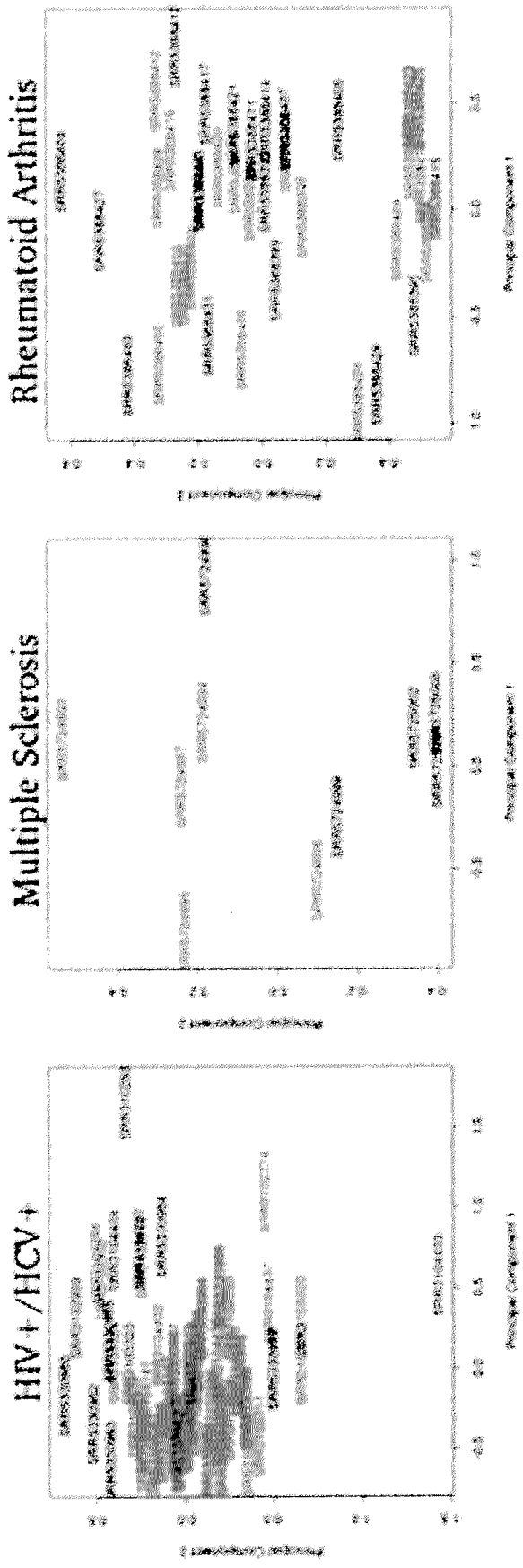
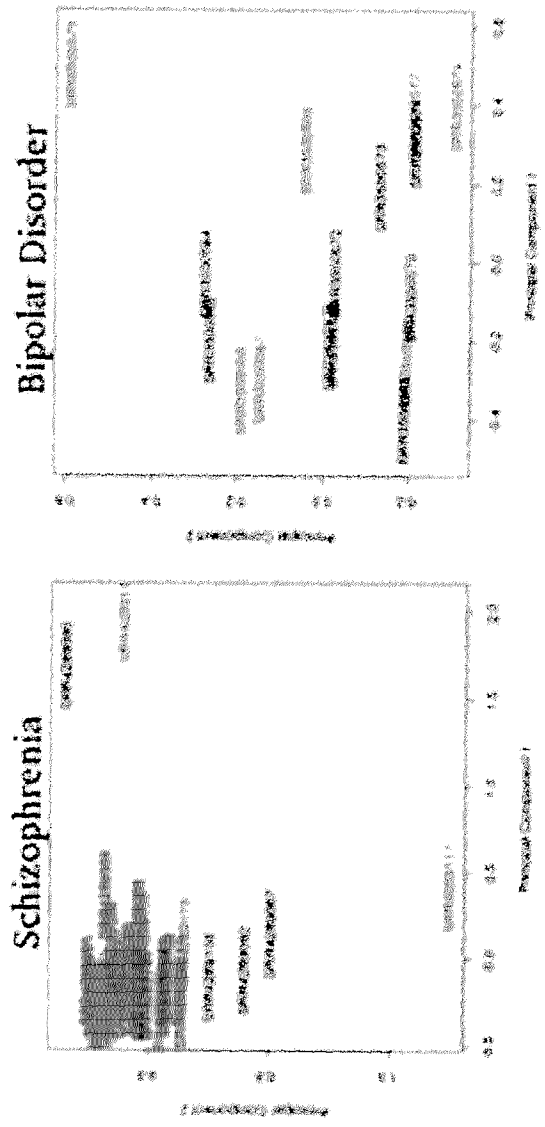


FIG. 34 CONT'D



Expression principal component plots produced by EdgeR from frequently expressed PFI38604 ORFs.
Each point is the SRA accession for the RNA-Seq data used to produce it. Red labels are + for the condition listed, black are controls. The blue labels for HIV were treated with INFα for 24 hrs

These plots show if there is a common set of genes whose expression is different between the two conditions. If so, you can see that they are non-overlapping on the plot.

FIG. 34 CONT'D

Figure 34: ERVK Expression Multidimensional Scaling. Two dimensional plots were produced by reduction of the high dimensional RNA-seq expression data derived from the Sequence Read Archive (SRA) using the R package EdgeR. The plot labels correspond to SRA studies as follows: ALS (SRP064478), Bipolar Disorder (SRP074904), Breast Cancer (SRP058722), HIV/HCV (SRP068424), Multiple Sclerosis (SRP110016), Prostate Cancer (ERP000550), Rheumatoid Arthritis (SRP102685), and Schizophrenia (SRP090259). In these plots each axis represents the leading log-2-scaled fold-change at one particular ERVK locus; the two loci chosen are those with the most extreme values in the majority of the clinical group's samples. Since these represent the biggest difference between samples, if no separation is apparent in these plots, there is no clear difference in transcript profiles. Each sample is represented by its SRA accession number, coloured red for disease-associated samples and black are the control samples. The blue HIV/HCV samples were treated with interferon alpha for 24 hours before sample collection. *Methods:* The expression of each ERVK locus identified in hg38 by RepeatMasker was measured in reads per kilobase per million mapped reads. This measure corresponds to the number of reads aligning to each region normalized by the region length and the number of reads mapped to the genome overall. This number was computed using the R package EdgeR. Read counts were obtained from alignments using samtools. Alignments were produced using bowtie2 to place raw reads onto an unmasked copy of hg38 with appropriate trimming parameters. Raw reads were obtained from the SRA using fastq-dump and trimming parameters were determined using FastQC.

FIG. 35

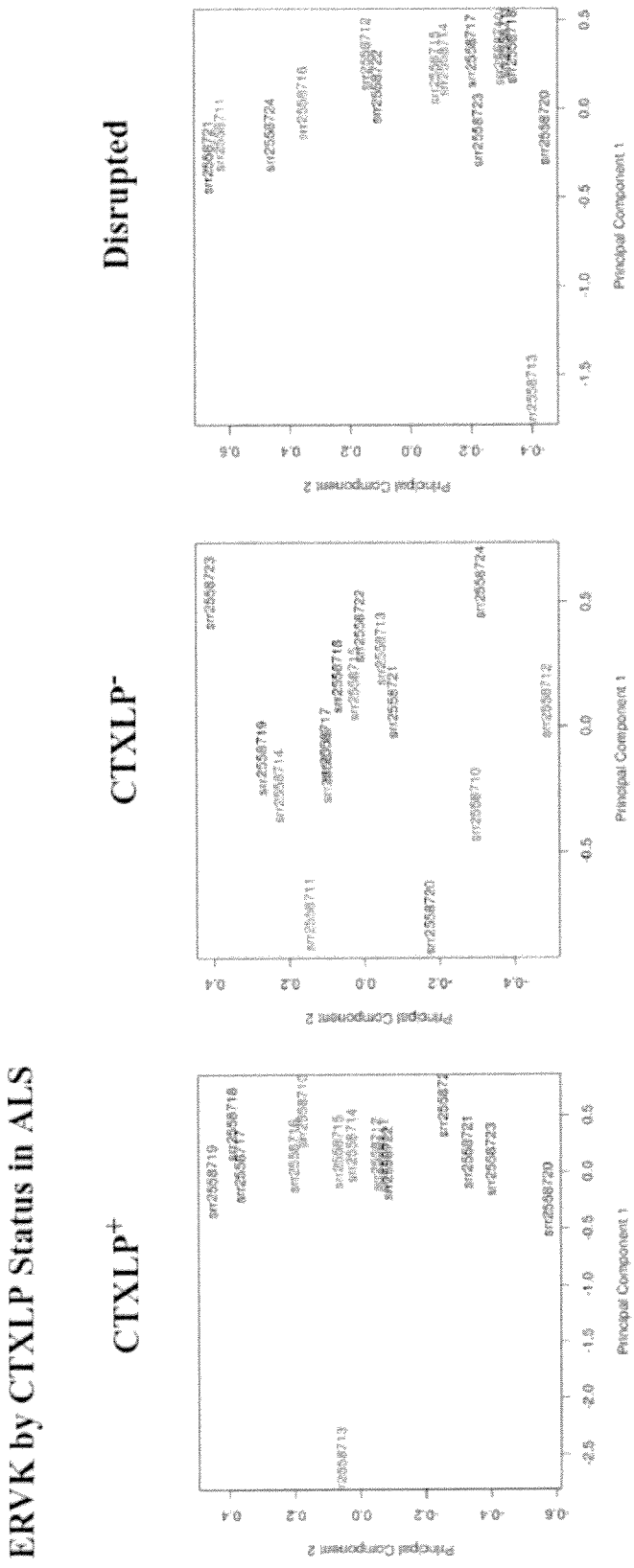

ERVK by CTXLP Status in ALS
CTXLP+
CTXLP-
Disrupted
Principal Component 2
Principal Component 1

FIG. 35 CONT'D

ERVK by CTXLP Status in Bipolar Disorder

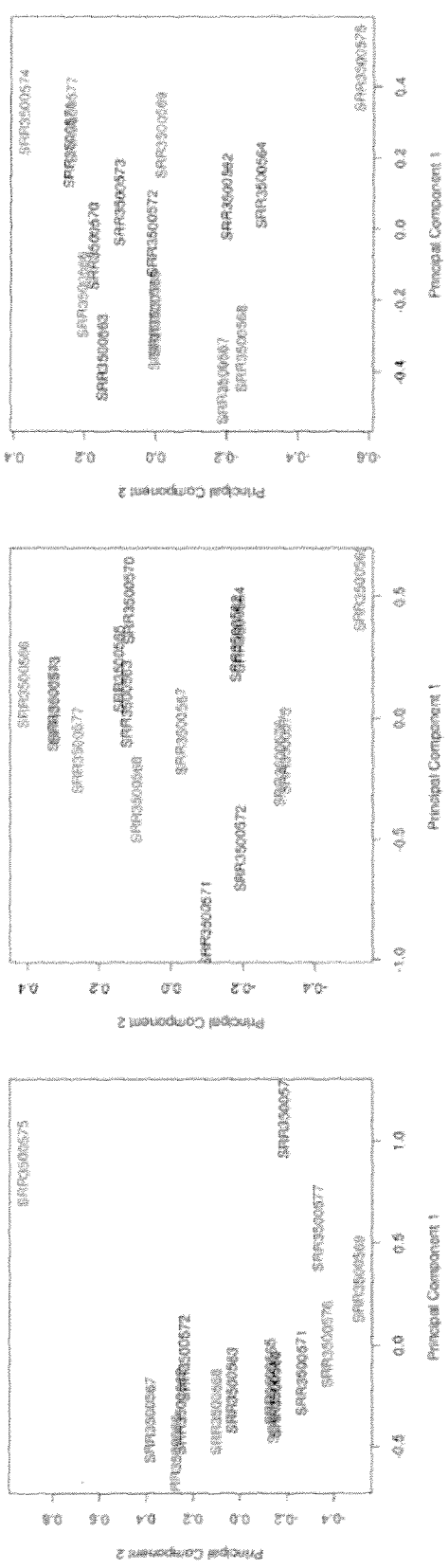


FIG. 35 CONT'D

ERVK by CTXLP Status in Breast Cancer

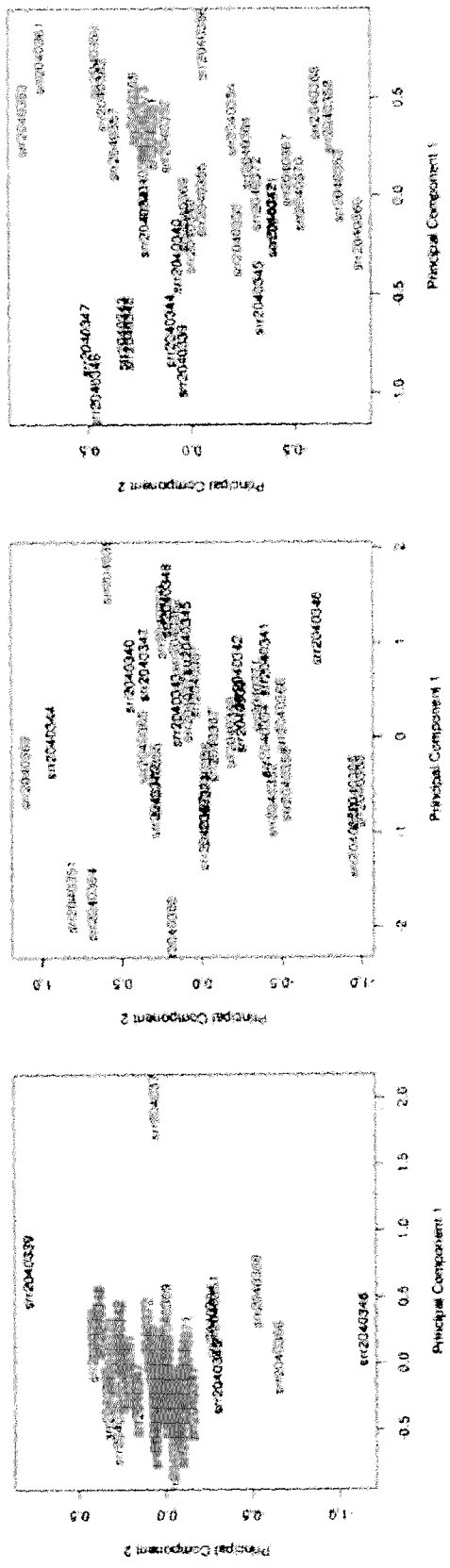


FIG. 35 CONT'D

ERVK by CTXLP Status in HIV-1 / HCV Co-infection

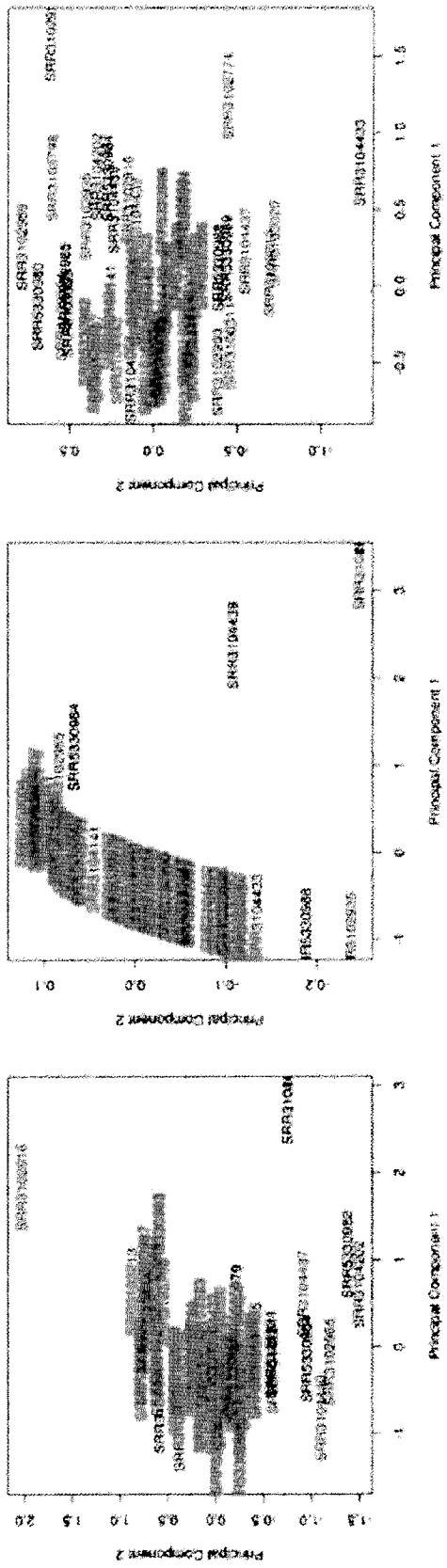


FIG. 35 CONT'D

ERVK by CTXLP Status in Multiple Sclerosis

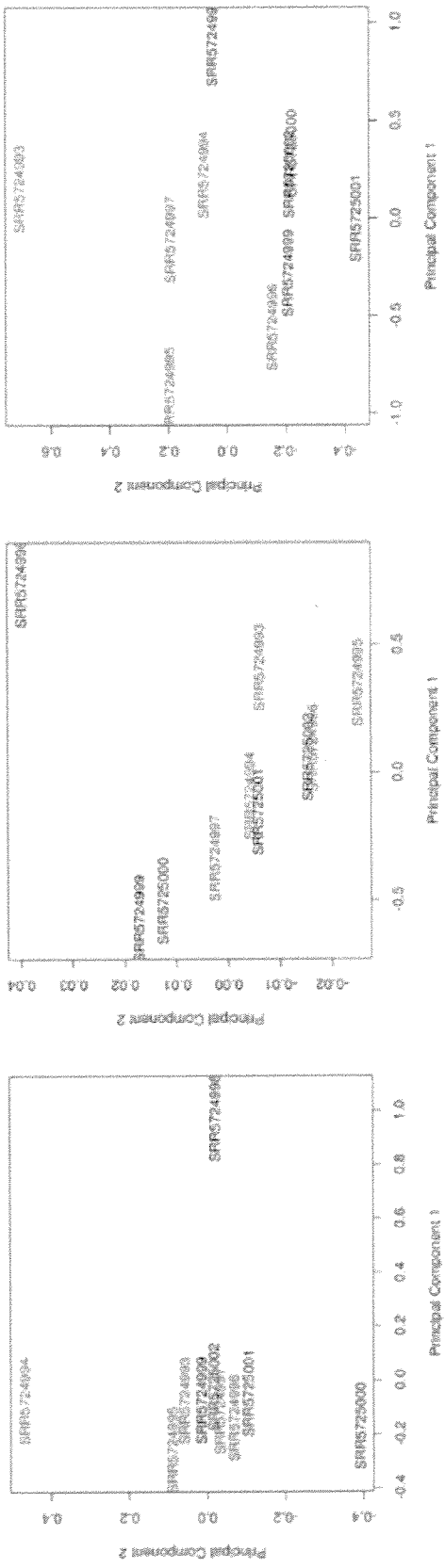


FIG. 35 CONT'D

ERVK by CTXLP Status in Prostate Cancer

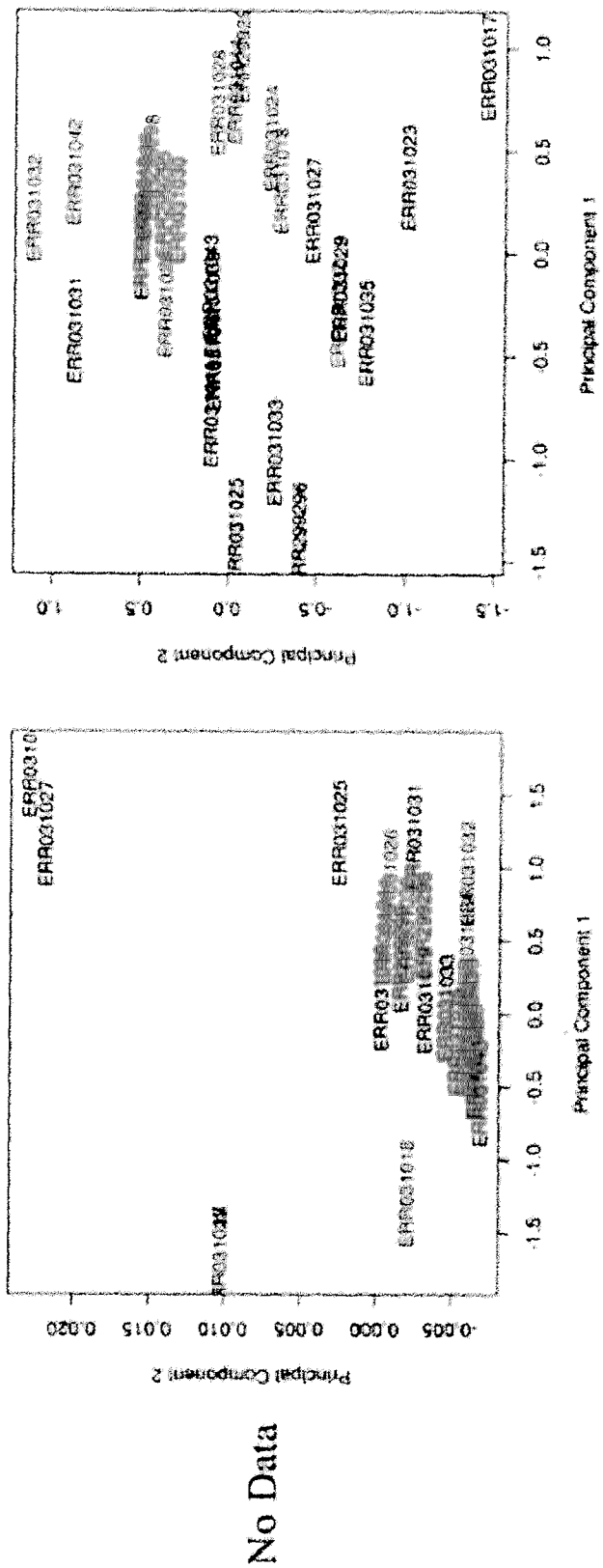


FIG. 35 CONT'D

ERVK by CTXLP Status in Rheumatoid Arthritis

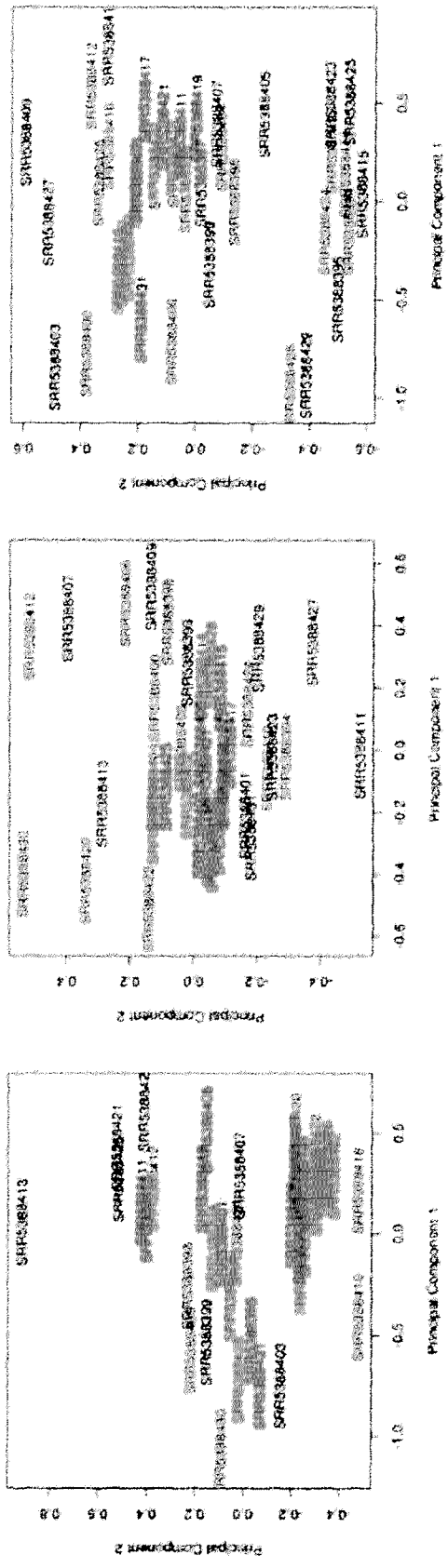


FIG. 35 CONT'D

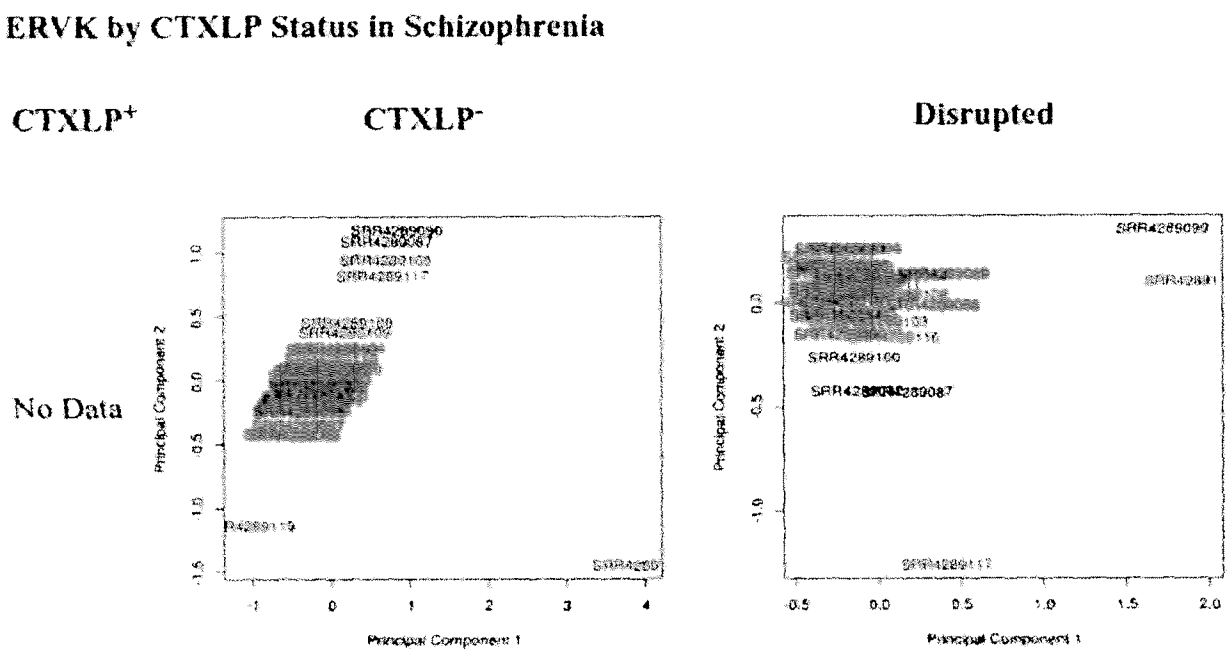


Figure 35: ERVK Multidimensional Scaling by CTXLP status in Human Disease. Each plot was produced as described in Figure 34, except with the modification that ERVK loci were subsetted into three states: "CTXLP+" which could produce CTXLP, "Disrupted" which cannot produce CTXLP but may have had an ancestral loci that produced CTXLP, and finally "CTXLP-" loci which do not and likely never did produce CTXLP from the ERVK *env* gene. In all cases, samples from the ALS, Bipolar Disorder, Breast Cancer, HIV-1/HCV co-infection, Multiple Sclerosis and Rheumatoid Arthritis cohorts expressed all three types of ERVK *env* transcripts. In Prostate Cancer and Schizophrenia, there were no loci expressed which were CTXLP+, and so no data is shown for that plot.

FIG. 36

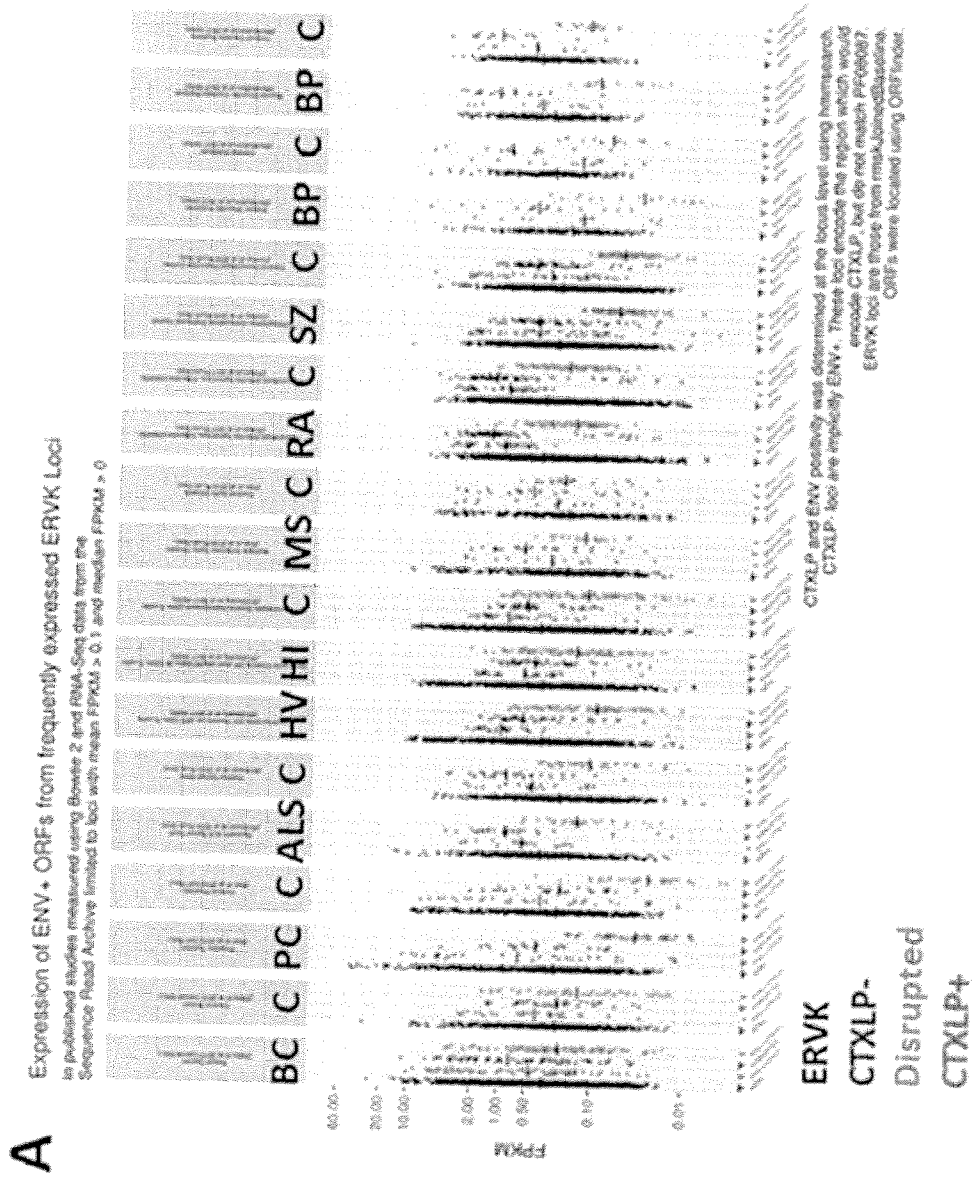

A
Expression of ENV+ ORFs from frequently expressed ERVK Loci
BC C PC C ALS C HV HI C MS C RA C SZ C BP C BP C
FPKM
ERVK
CTXLP-
Disrupted
CTXLP+

FIG. 36 CONT'D

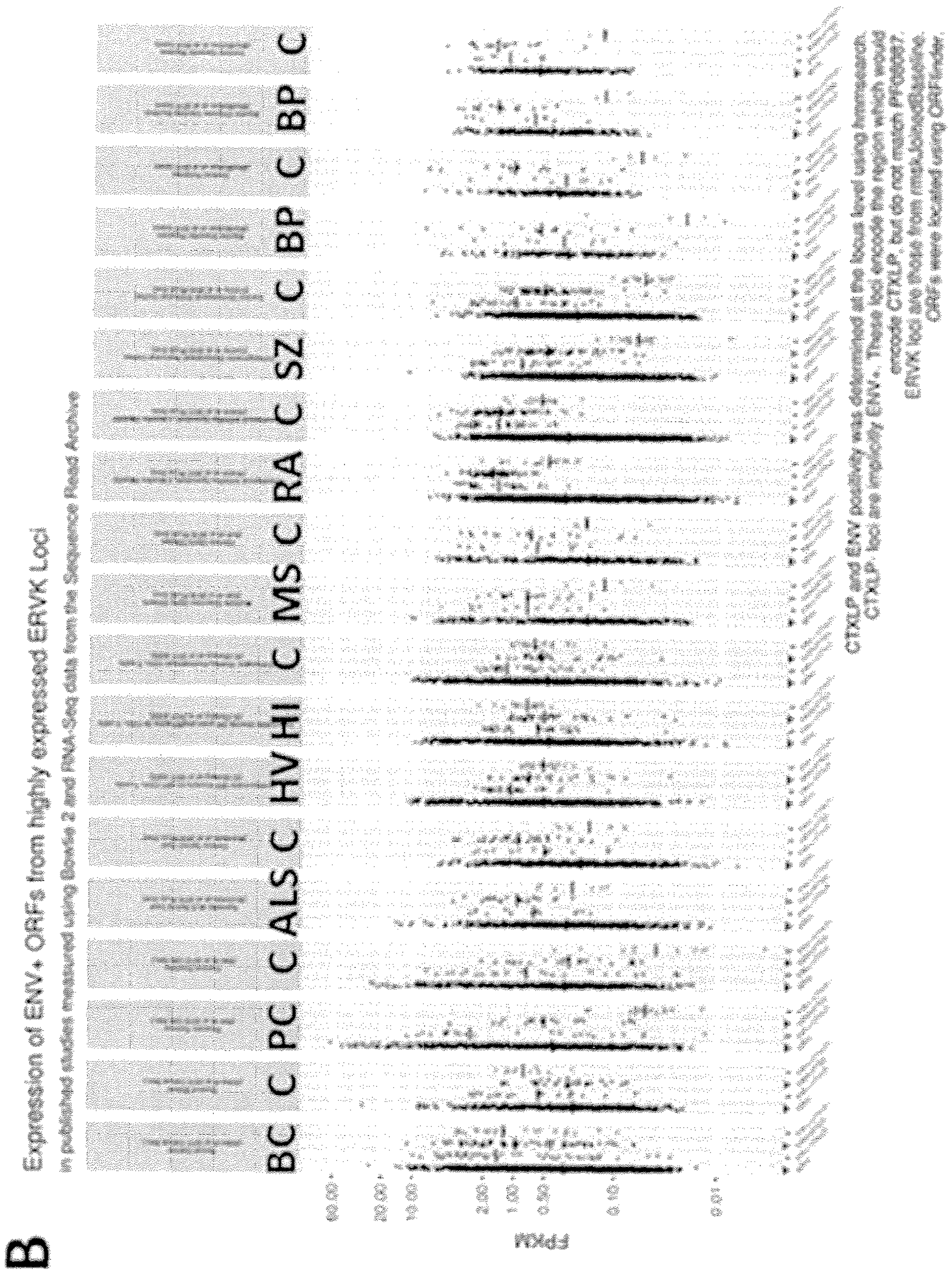

B
Expression of ENV+ ORFs from highly expressed ERVK Loci
in published studies measured using Bowtie 2 and RNA-Seq data from the Sequence Read Archive
BC
C
PC
C
ALS
C
HV
HI
C
MS
C
RA
C
SZ
C
BP
C
BP
C
FPKM
50.00
20.00
10.00
2.00
1.00
0.50
0.10
0.01
CTXLP and ENV positivity was determined at the locus level using hmmsearch.
CTXLP- loci are implicitly ENV+. These loci encode the region which would
encode CTXLP, but do not match PF08087.
ERVK loci are those from rmsk.JoinedBaseline.
ORFs were located using ORFfinder.

FIG. 36 CONT'D

Figure 36: Per-Locus Differential ERVK Expression. Each panel shows in detail the expression of individual ERVK loci encoding envelope in each human disease condition. ERVK loci (black) are plotted against CTXLP+ (red), CTXLP- (blue) and disrupted (grey) loci. The plot labels correspond to SRA studies as follows: ALS (ALS: SRP064478), Bipolar Disorder (BP: SRP074904), Breast Cancer (BC: SRP058722), HIV/HCV (HV; HIV/HCV + interferon (HI): SRP068424), Multiple Sclerosis (MS: SRP110016), Prostate Cancer (PC: ERP000550), Rheumatoid Arthritis (RA: SRP102685), and Schizophrenia (SZ: SRP090259). For each study, controls (C) and indicated to the right of cases. Panel A exclude ERVK loci with very low expression; only loci with a median expression greater than 0 and a mean expression greater than 0.1 are plotted. Panel B shows only loci which were highly expressed; only ERVK loci which had a maximum expression higher than 2 are plotted. *Methods:* FASTQ files for each run from each study were downloaded from the SRA using fastq-dump. Trimming positions for each study were chosen based on examining the output of FastQC. Reads were mapped to an unmasked copy of hg38 obtained from UCSC using bowtie2. All the above were performed on the Compute Canada cluster Orcinus. Mapped reads (SAM files) were downloaded and SAM files were converted to indexed BAM files. Expression of ERVK loci was measured using samtools view to count reads aligning to each locus. Expression data were interpreted using EdgeR.

FIG. 37

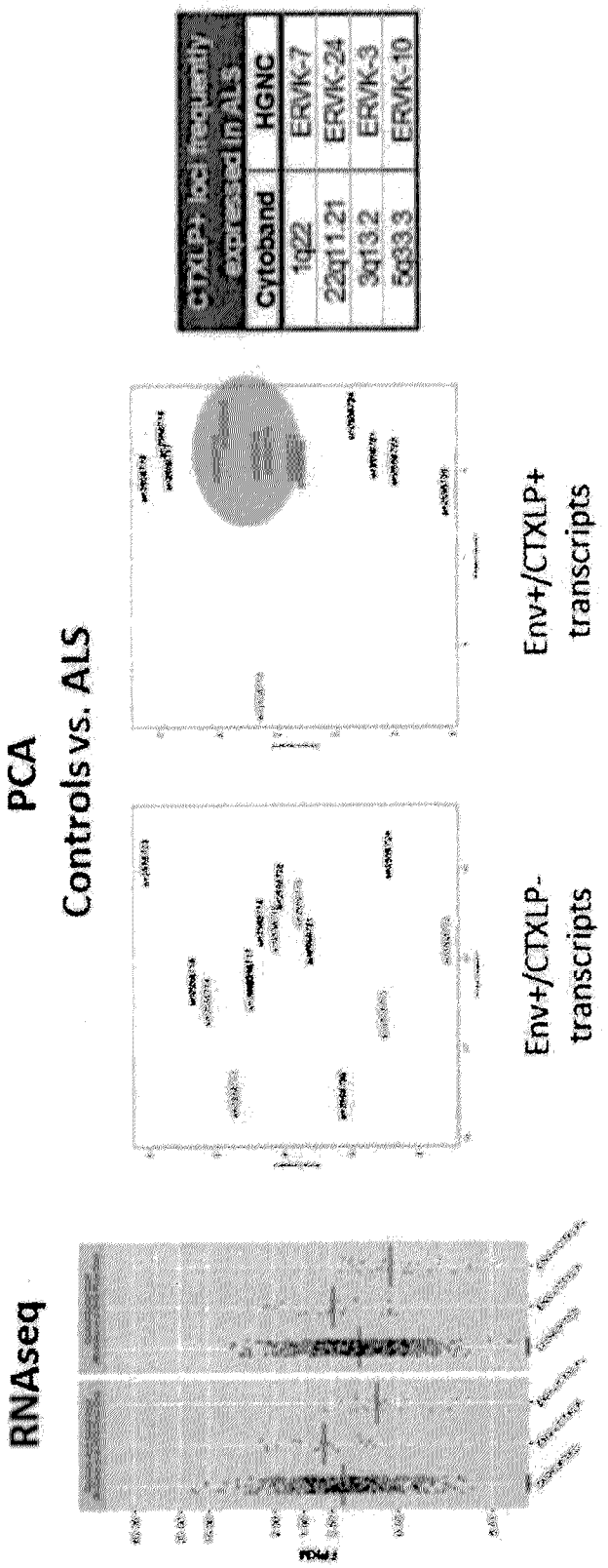


Figure 37: ERVK CTXLP encoding transcripts and CTXLP protein are present in Amyotrophic Lateral Sclerosis (ALS). (A) Re-analysis of RNAseq data[6] in control (right) and sporadic ALS (left) spinal cords for expression of disrupted non-coding (black), Env+/CTXLP- (blue) and Env+/CTXLP+ (red) *env* transcripts. Principle component analysis (PCA) reveals ALS patient clustering in terms of CTXLP+ transcript expression, with most frequently expressed CTXLP encoding loci indicated.

FIG 38

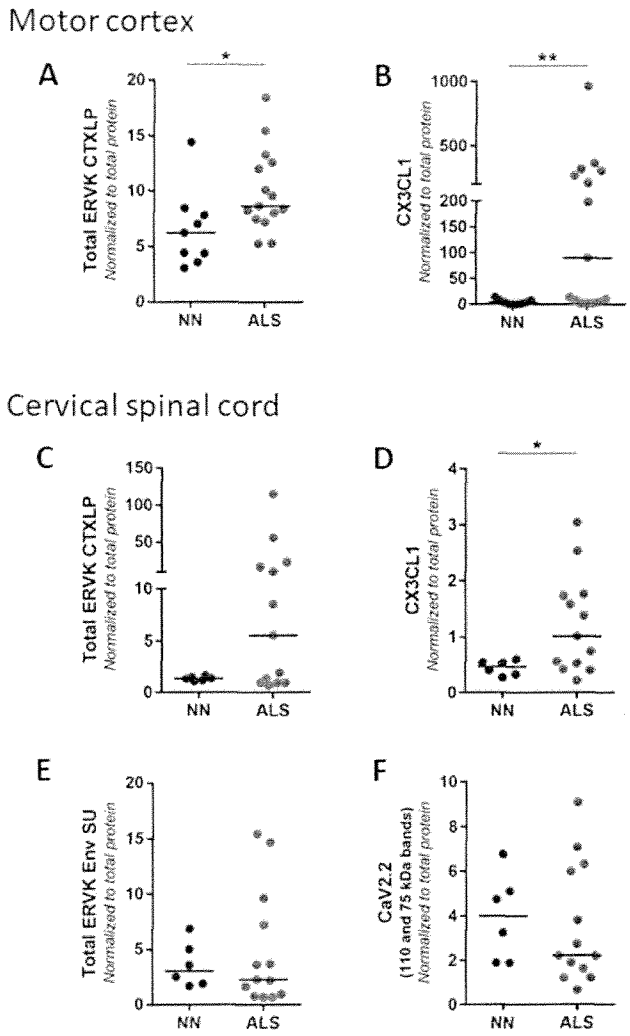


Figure 38: ERVK CTXLP levels are enhanced in autopsy spinal cord and brain tissues of patients with ALS. Graph represents total CTXLP (**A**) and CX3CL1 (**B**) quantification in neuronormal (NN) (n=9) and ALS (n=15) motor cortex specimens, as measured by western blot. Graph represents total CTXLP (**C**), CX3CL1 (**D**), ERVK Env SU (**E**) and CaV2.2 (**F**) quantification in NN (n=6) and ALS (n=13) cervical spinal cord specimens, as measured by western blot. Statistical test with unpaired two-tailed t-tests (*p<0.05, **p<0.01, black bars are medians).

FIG 39

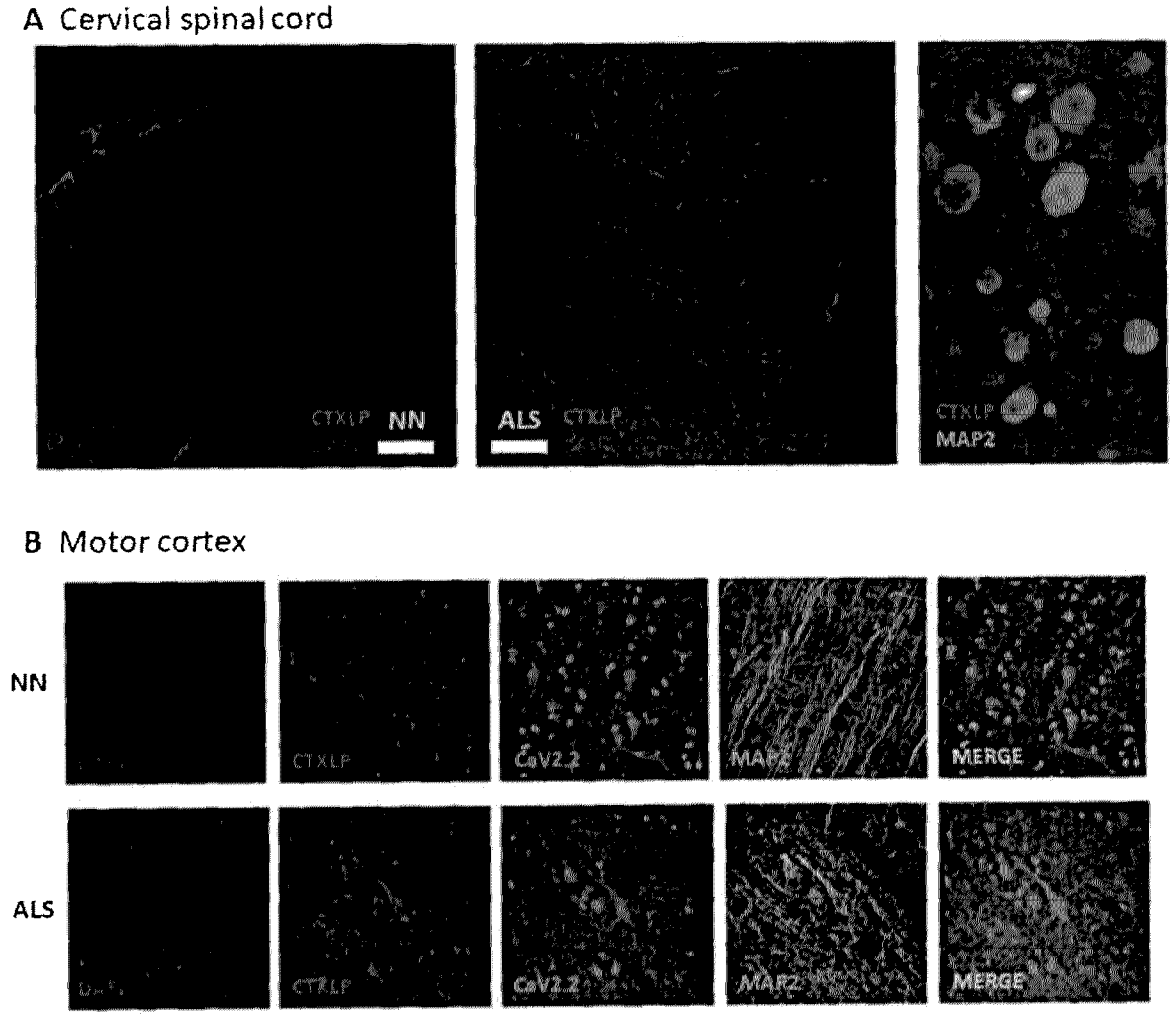
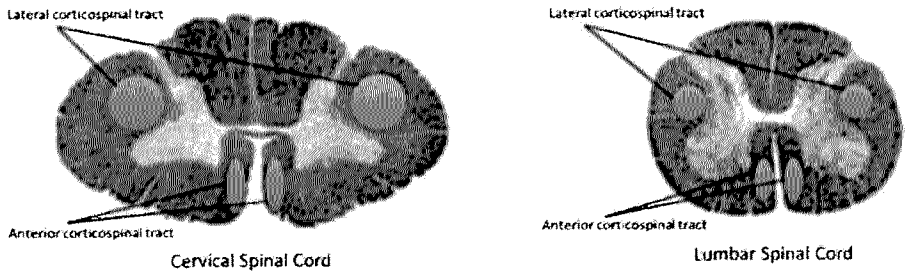


Figure 39: ERVK CTXLP levels are enhanced in autopsy spinal cord and brain tissues of patients with ALS, as measured by confocal microscopy. (A) Representative 10x confocal micrographs of ERVK CTXLP expression in *ex vivo* cervical spinal cord of a neuronormal control (NN, n=3) and patient with ALS (n=3). DAPI stain depicts nuclei. High magnification reveals staining in cells surrounding MAP2+ axons, suggesting CTXLP+ oligodendrocytes. CTXLP+ rings ranged from 6-16μM in diameter. **(B)** Representative 40x confocal micrographs of ERVK CTXLP, voltage-gated calcium channel CaV2.2 (C-terminal antibody) and neuronal MAP2 expression in Brodmann area 6 (BA6) motor cortex tissue of a NN control (n=3) and patient with ALS (n=3). Note the translocation of nuclear CTXLP to cytoplasmic aggregates in neurons from an ALS patient, as well as an overall decrease in CaV2.2 expression. DAPI stain depicts nuclei.

FIG 40


Lateral corticospinal tract
Anterior corticospinal tract
Cervical Spinal Cord
Lateral corticospinal tract
Anterior corticospinal tract
Lumbar Spinal Cord

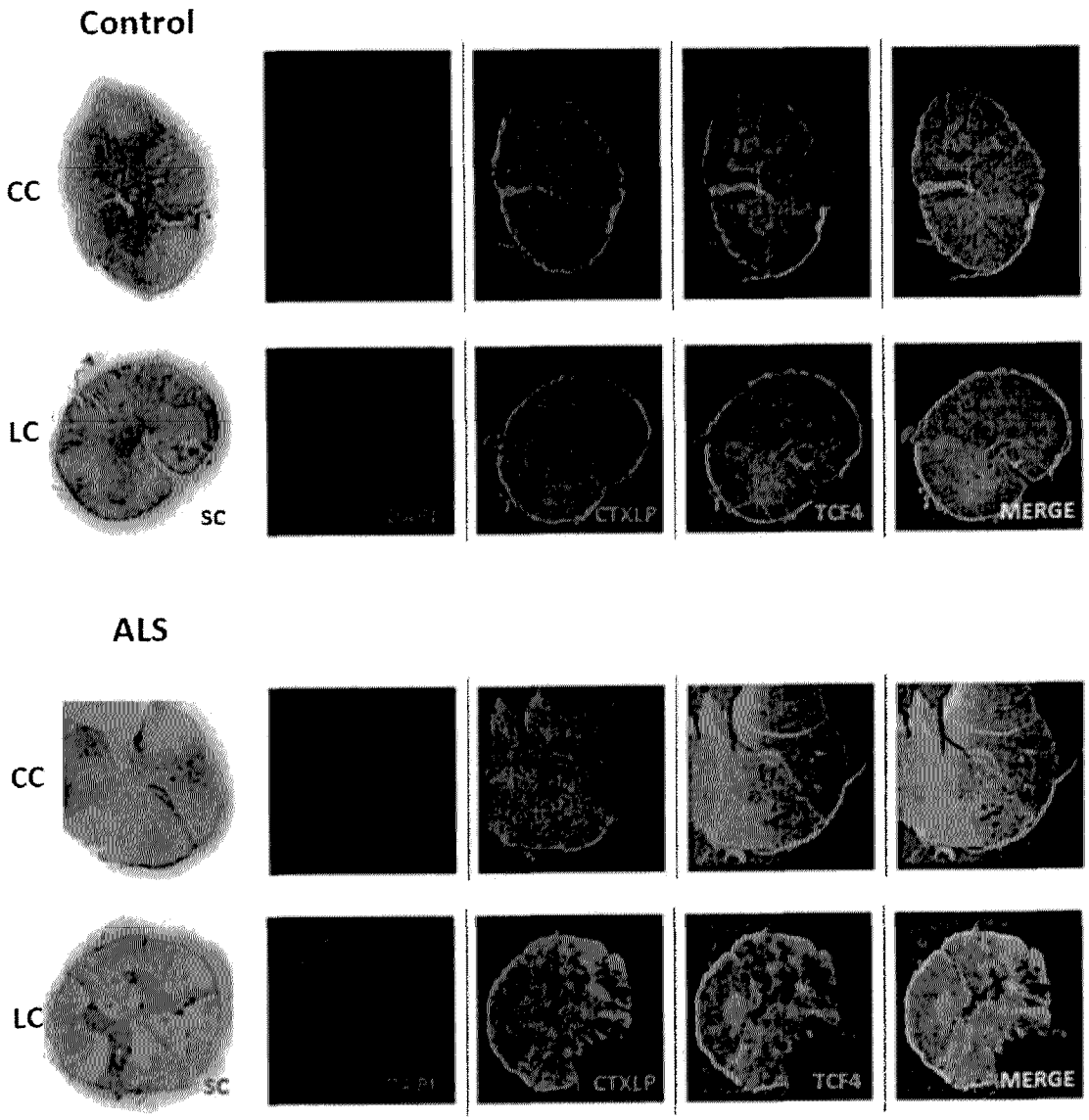

Control
CC
LC
SC
CTXLP
TCF4
MERGE
ALS
CC
LC
SC
CTXLP
TCF4
MERGE

FIG 40 CONT'D

Figure 40: ERVK CTXLP levels are enhanced in autopsy cervical and lumbar spinal cord tissues from patients with ALS, as measured by light and confocal microscopy. Representative 10x confocal micrographs of ERVK CTXLP expression in *ex vivo* cervical (CC) and lumbar (LC) spinal cord of a neuronormal control (NN, n=3) and patients with ALS (n=3). Solochrome cyanine (SC) stain (purple) with eosin counterstain (pink) depicts tissue myelination; pale lesions appear in ALS tissues. These lesioned areas exhibit increased CTXLP expression is in red. Oligodendrocyte precursor marker TCF4 is in green. DAPI stain depicts cellular nuclei. Note that CTXLP expression occurs in either lateral and/or anterior cortical spinal tracts.

FIG 41

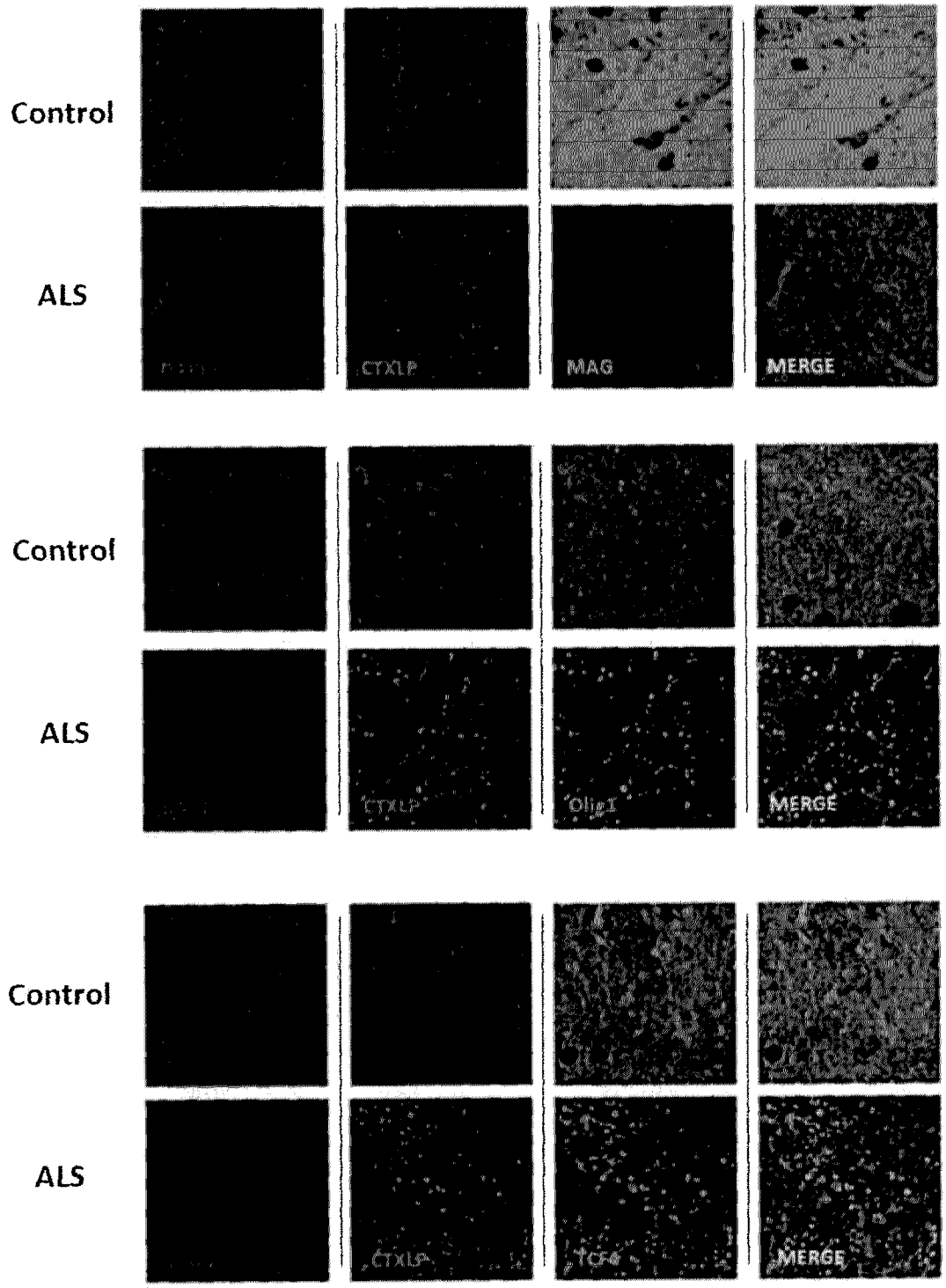

Control
ALS
CTXLP
MAG
MERGE
Control
ALS
CTXLP
Olig1
MERGE
Control
ALS
CTXLP
TCF4
MERGE

FIG 41 CONT'D

Figure 41: ERVK CTXLP levels are associated with demyelination and CTXLP is enhanced in *TCF4+Olig1+ oligodendrocyte precursors* in cervical spinal cord tissues from patients with ALS. Representative 20x confocal micrographs of ERVK CTXLP expression in *ex vivo* cervical (CC) spinal cord of a neuronormal control (NN, n=3) and a patient with ALS (n=3). Notably decreased myelin stain as measure my MAG protein expression (green) is evident in ALS tissue as compared to control. In ALS tissue, CTXLP expression (red) co-localizes with TCF4 (green) or Olig1 (green) markers indicative of oligodendrocyte precursor cells. DAPI stain depicts cellular nuclei.

FIG 42

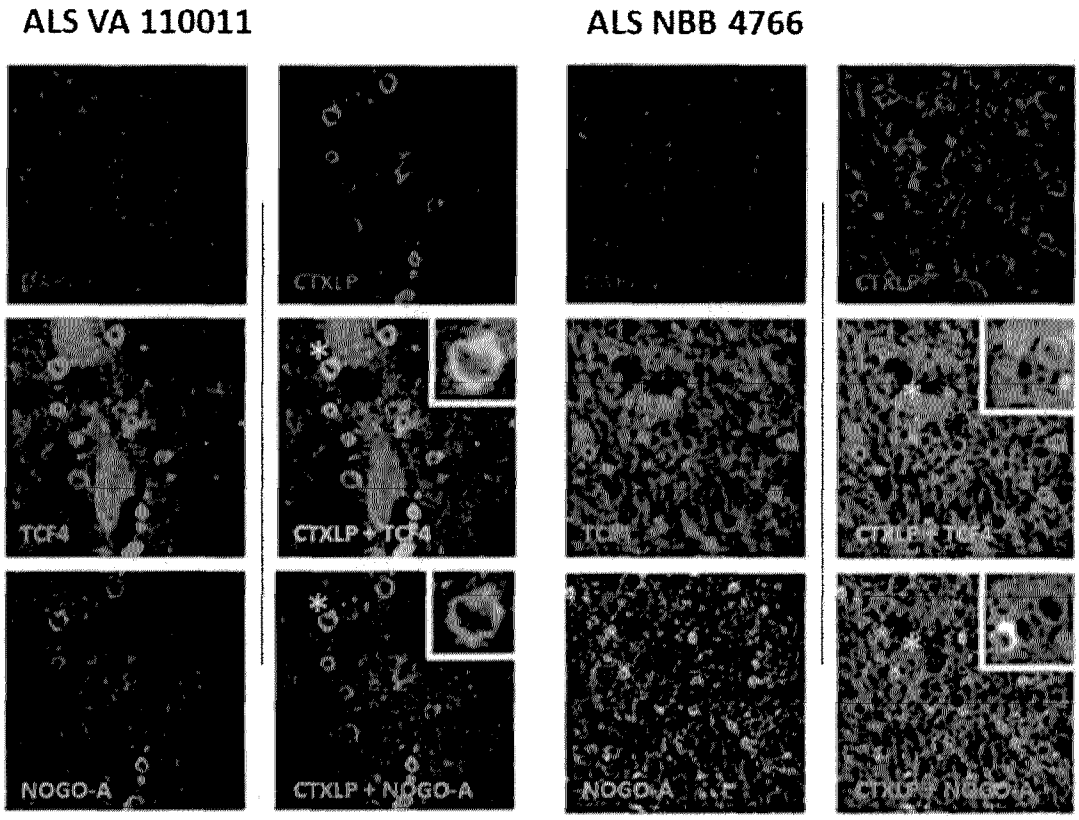


Figure 42: ERVK CTXLP+ oligodendrocyte precursors express myelin inhibitory protein Nogo-A, or are in close proximity to Nogo-A positive cells in spinal cord tissues of patients with ALS. Human *ex vivo* cervical spinal cord tissues were stained for ERVK CTXLP (red), TCF4 (green), Nogo-A (grey) and nuclei (blue) in neuro-normal controls (n=3) and patients with ALS (n=3). Image merging for CTXLP and TCF4 indicate that oligodendrocyte precursors express CTXLP in ALS. Image merging for CTXLP and Nogo-A indicate that oligodendrocyte precursors can express myelin inhibitor protein Nogo-A (left panel) or alternately are in proximity to Nogo-A expressing cells in ALS (right panel). White stars indicate areas that are magnified to depict overlapping protein expression in CTXLP+ rings.

FIG 43

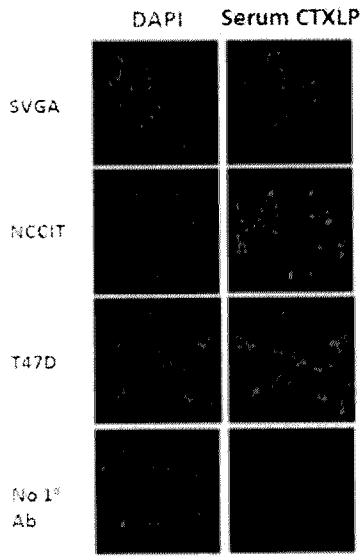


Figure 43: Cancer cells express greater levels of CTXLP as compared to non-cancer cells. Prototypic cell lines for teratocarcinoma (NCCIT) and breast cancer (T47D) were examined for CTXLP expression as compared to astrocytic SVGA cells using confocal microscopy. No antibody negative control is to show that specificity of CTXLP (red) staining requires an antibody targeting ERVK CTXLP. Nuclei are shown in blue using a DAPI stain.

FIG 44

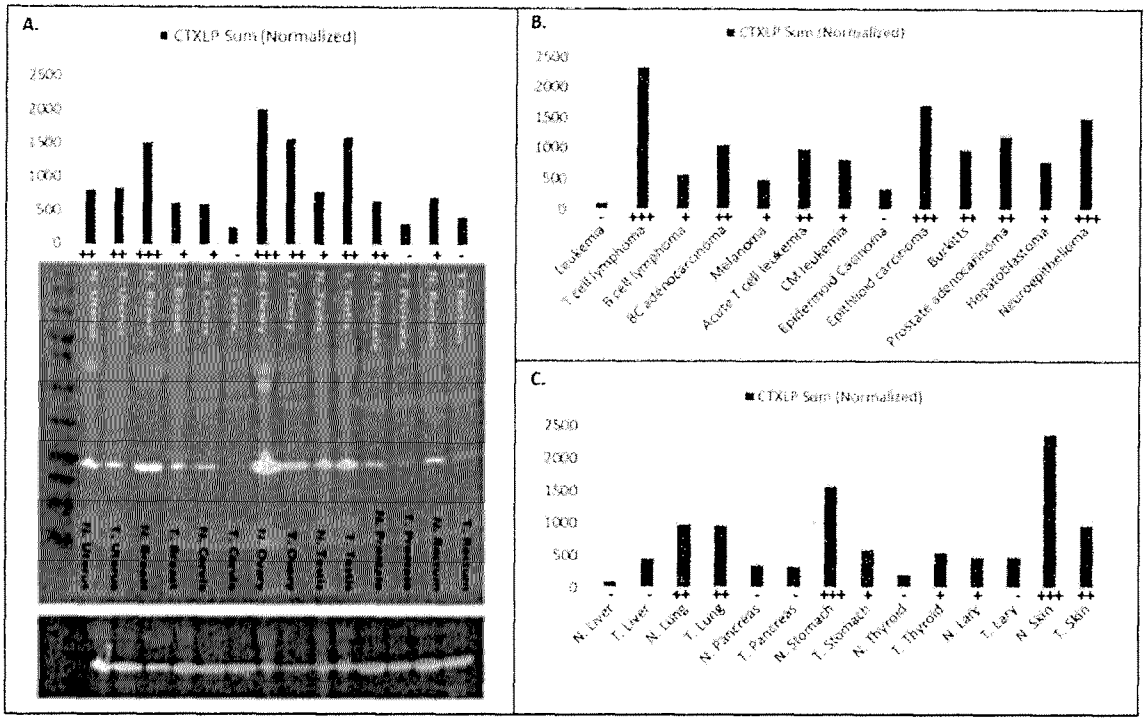


Figure 44: CTXLP expression in G-Bioscience Ready-to-screen cancer tissue and cell line blots. TB56-I (A), TB55 (B) and TB56-II (C) blots were screen for CTXLP expression (blue bars, top blot) normalized to β-actin loading control (lower blot). Enhanced CTXLP expression in noted is several cancer types, including T cell lymphoma, neuroepithelioma, prostate, ovary, testis and skin cancers.

FIG 45

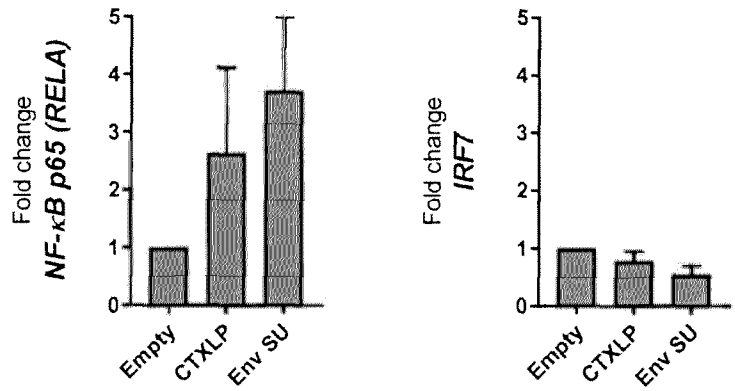


Figure 45: Changes in the gene expression of pro-inflammatory NF-κB p65 and anti-viral IRF7 in response to CTXLP and SU expression. 293T cells were transfected with plasmids encoding empty vector, ERVK CTXLP or ERVK SU for 24 hours. Cell pellets were collected, RNA extracted, and cDNA produced for use in Q-PCR experiments to evaluate relative gene expression of RELA and IRF7. Analysis performed using ΔΔCt method and 18S RNA as a calibrator.

FIG 46

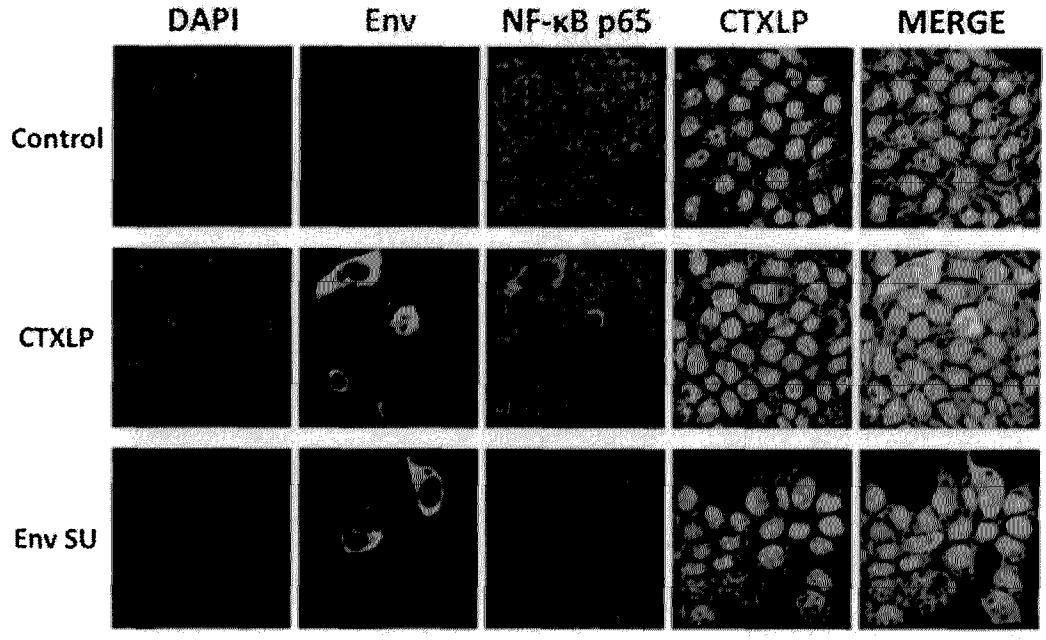


Figure 46: Confocal images of control, CTXLP-expressing and SU-expressing 293T cells. *Cells* were stained with DAPI nuclear stain, and antibodies against Env SU (green) fluorescent dye, NF-κB p65 (red) fluorescent dye and CTXLP (grey / white) fluorescent dye. Only CTXLP expressing cells exhibit upregulated NF-κB p65 expression. Representative micrographs of n=3 experiments.

FIG 47

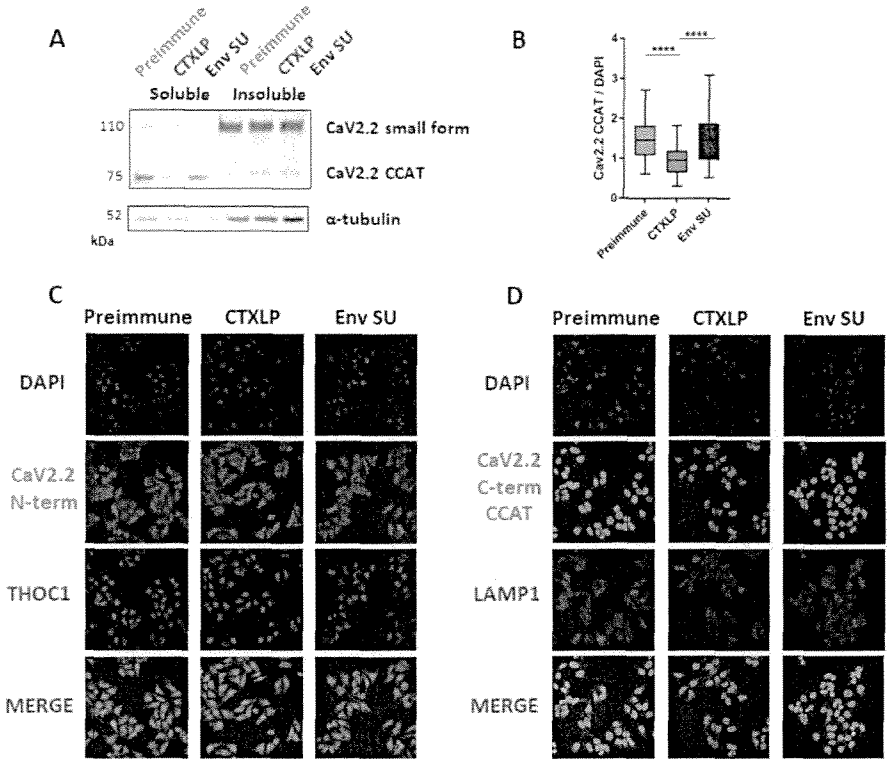


Figure 47: ERVK CTXLP, but not ERVK Env SU, depletes CaV2.2 calcium channel-associated transcription regulator (CCAT). **(A)** CaV2.2 expression in SVGA cells treated with 5µl of immunoprecipitation (IP) products from NCCIT cell lysates extracted using rabbit pre-immune serum, custom rabbit anti-CTXLP antibody or custom rabbit anti-ERVK Env SU antibody for 24 hours. Membrane-bound forms of CaV2.2 are 240/210 kDa (not shown) and 110 kDa respectively, whereas the CaV2.2 CCAT is 75 kDa. **(B)** Quantification of CaV2.2 depletion in IP-product treated astrocytes (2hrs), based on confocal quantification (****p<0.0001, 80-100 cells per condition quantified). **(C & D)** Confocal imaging of CaV2.2 (N-terminal antibody) and CaV2.2 CCAT (C-terminal antibody) illustrates that CTXLP depletes nuclear CaV2.2 CCAT within 2 hours, but not the membrane-associated CaV2.2 channel (n=2).

FIG 48

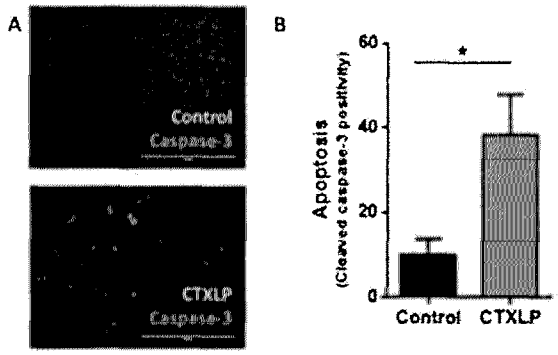


Figure 48: CTXLP administration enhances caspase-3 expression in SVGA cells. (A) Images of live cells after 24 hours taken using EVOS imager, top image depicts untreated control cells and bottom image depicts cells treated with 5µl of CTXLP immunoprecipitated solution. (B) Graphical depiction of caspase-3 expression in control and CTXLP-treated cell after 24 hours.

FIG 49

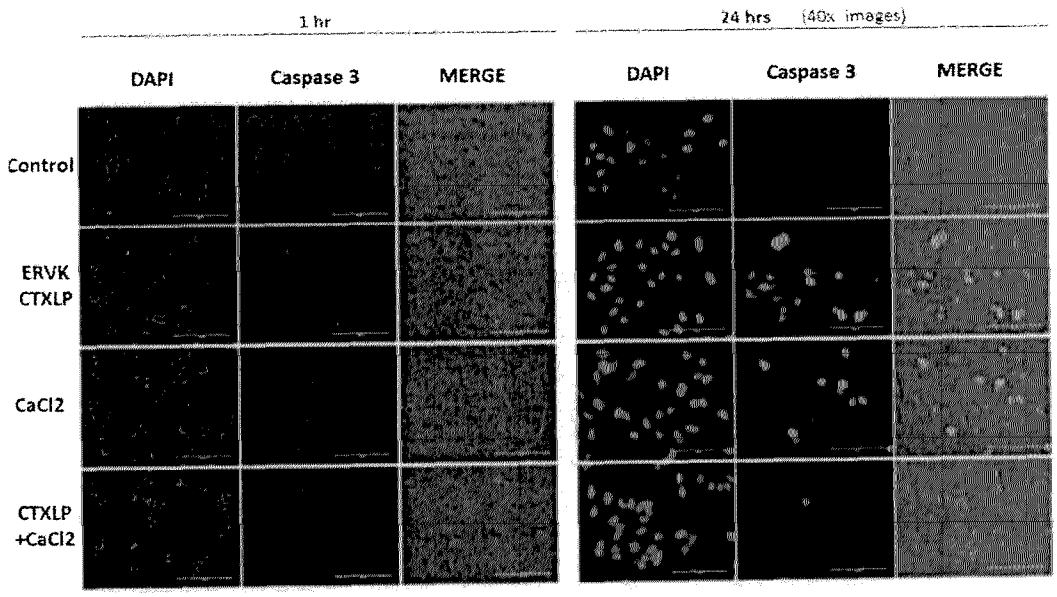


Figure 49: CTXLP induces caspase-3 activation and apoptosis, which can be blocked by excess extracellular calcium. Cell survival 1 hour and 24 hours post treatment with 5µl of buffer, calcium chloride, CTXLP or CTXLP and calcium chloride. Cells were examined using an EVOS microscopy for caspase-3 activation (green) or nuclei (blue), n=2). Excess calcium chloride is known to block the cellular effects of conotoxin proteins[7].

FIG 50

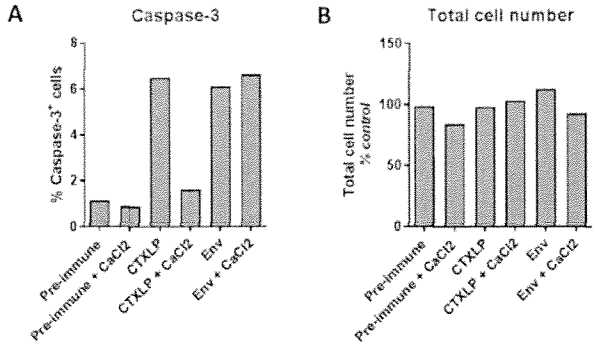


Figure 50: Cell survival and confluency 24 hours post treatment with 5µl of pre-immune serum, pre-immune serum and calcium chloride, CTXLP, CTXLP and calcium chloride, SU, or SU and calcium chloride. (A) Graphical depiction of caspase-3 expression in SVGA cells 24 hours post treatment. (B) Graphical depiction of cell confluency of SVGA cells 24 hours post-treatment. Note that pre-immune serum was used as a negative control as this is the component of immunoprecipitation product.

FIG 51

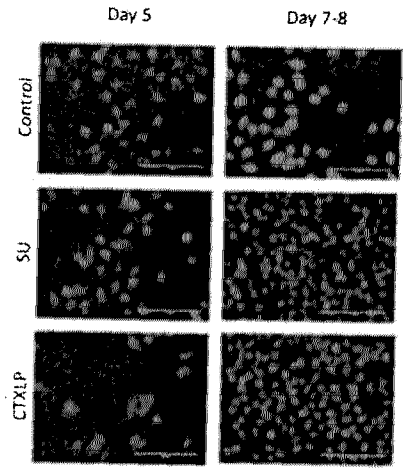


Figure 51: CTXLP induced caspase-3 does not result in cell death. Live cell images of SVGA cells in control, SU-treated and CTXLP-treated conditions stained for caspase-3 using EVOS live cell imaging. Cells in each condition were imaged after 5 days. Control cells were imaged at 7 days and SU and CTXLP-treated cells were imaged after 8 days. Controls cells express greater amounts of the apoptotic marker caspase-3, and have not proliferated to the extent observed in CTXLP and Env treatments.

FIG 52

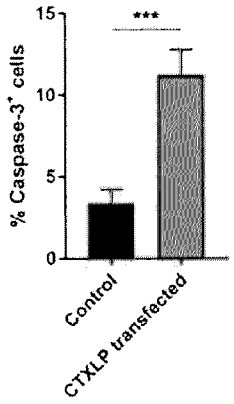


Figure 52: CTXLP-transfection markedly enhances capase-3 expression. Percentage of control and CTXLP-transfected SVGA cells expressing caspase-3 after 24 hours. SVGA cells were transfected with empty vector and lipofectamine LTX (control) or a pcDNA3.1 CTXLP-expressing vector. Graphical depiction of apoptosis marker caspase-3 in control and CTXLP-transfected SVGA cell after 24 hours, as measured by confocal microscopy.

FIG 53

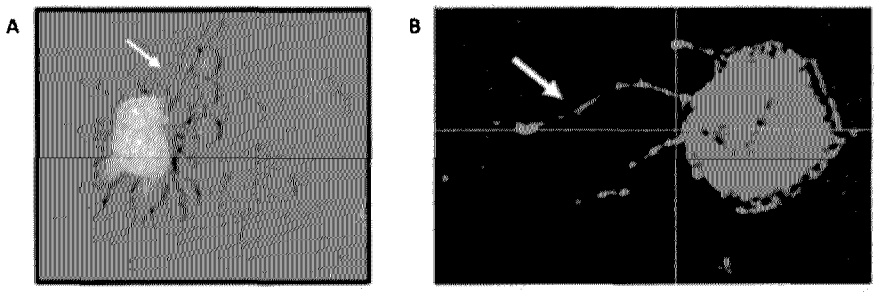


Figure 53: Abnormal cellular morphology in cells exposed to CTXLP. (A) Fluorescent image of CTXLP-treated SVGA cell expressing caspase-3 (green) and stained with DAPI nuclear stain (blue). (B) Confocal image of Env SU expression (green) in a CTXLP-expressing 293T cell.

FIG 54

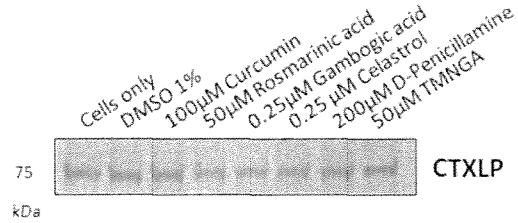


Figure 54: CTXLP-limiting drug screen in human NCCIT teratocarcinoma cells. The ERVK-expressing NCCIT teratocarcinoma cell line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 (data not shown) and 48 hours with known IC_{50} concentrations of drugs (NCCIT cells alone, 1% DMSO as drug carrier control, 100μM Curcumin, 50μM Rosmarinic acid, 0.25μM Gambogic acid, 0.25μM Celastrol, 200μM D-Penicillamine and 50μM Tetramethyl Nordihydroguaiaretic acid/TMNGA). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, as compared to β-actin loading control. Results demonstrate that select MAEs can reduce CTXLP expression in NCCIT cells. n = 2.

FIG 55

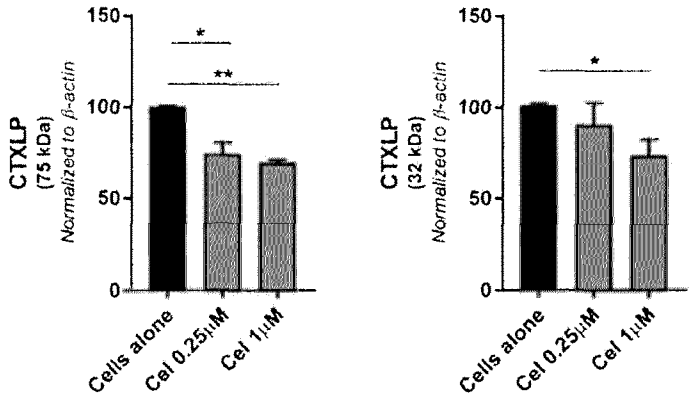


***Figure 55*: MAE drug Celastrol abrogates CTXLP expression in human NCCIT teratocarcinoma cells.** The ERVK-expressing NCCIT teratocarcinoma cell line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 hours with increasing doses of celastrol (Cel, 0.1, 0.25, 1 and 2.5μM). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, as compared to β-actin loading control. Results demonstrate that Cel dose-dependently reduces CTXLP expression in NCCIT cells. n = 4 independent experiments, *p<0.05, **p<0.01, two-tailed paired t-test.

FIG 56

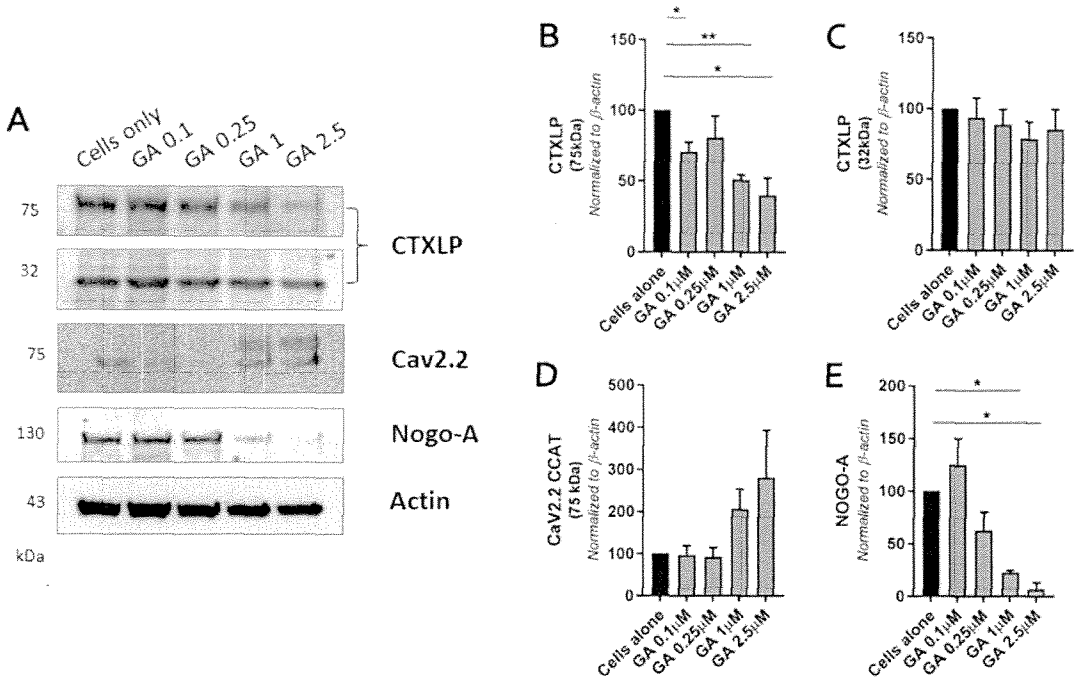


Figure 56: MAE drug Gambogic acid abrogates CTXLP expression in human NCCIT teratocarcinoma cells. The ERVK-expressing NCCIT teratocarcinoma cell line was grown in a monolayer in RMPI 1640 media supplemented with FetalGro. Cells were treated for 24 hours with increasing doses of gambogic acid (GA, 0.1, 0.25, 1 and 2.5μM). Cells were collected, protein extracted, and western blot performed to measure the expression of ERVK CTXLP, CaV2.2 C-terminal calcium channel-associated transcriptional regulator (CaV2.2 CCAT) and NOGO-A, as compared to β-actin loading control. Results demonstrate that GA dose-dependently reduces CTXLP and NOGO-A expression in NCCIT cells. CaV2.2 CCAT is inversely correlated with the expression of ERVK CTXLP in this culture system. n = 3, *p<0.05, **p<0.01, two-tailed paired t-test.

FIG 57

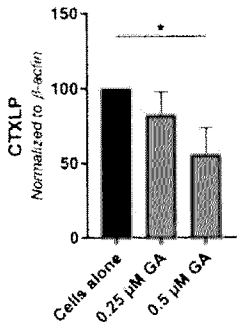


Figure 57: Endogenous levels of CTXLP in human astrocytes can be depleted in the presence of gambogic acid. Human astrocytic cell line SVGA were treated with increasing doses of gambogic acid in the low micromolar range (0.25 and 0.5µM). CTXLP expression was measured by western blot. n = 4, *p<0.05, one-tailed paired t-test.

FIG 58

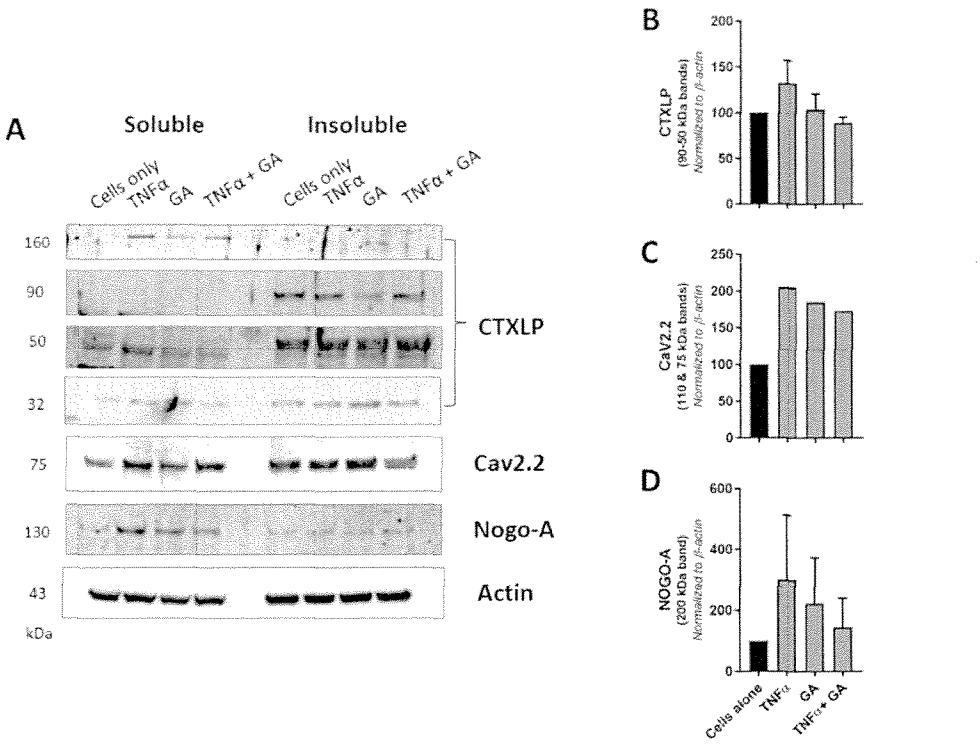


Figure 58: Gambogic acid blocks TNFα-induced CTXLP expression in human neurospheres. The ReNcell CX neuroprogenitor cell line was grown in suspension to produce human neurospheres (approximately 0.5mm diameter). Neurospheres were treated for 24 hours in neurobasal media alone, or with or without 1ng/ml TNFα and/or 0.5μM gambogic acid (GA). Cells were collected, protein extracted, and western blot performed to measure the expression of soluble (standard lysis) and insoluble (RIPA lysis) proteins for ERVK CTXLP, CaV2.2 C-terminal calcium channel-associated transcriptional regulator (CaV2.2 CCAT) and NOGO-A, as compared to β-actin loading control. Results demonstrate that TNFα enhances CTXLP and NOGO-A expression in neurons, whereas in the presence of GA this effect is blocked. CaV2.2 CCAT is inversely correlated with the expression of ERVK CTXLP in this neuronal culture system. n = 2.

FIG. 59

SEQ ID NOs:
Orthologues
Human
SEQ ID NOs:
Chimpanzee
SEQ ID NOs:
Gorilla
Mangabey

FIG. 59 CONT'D

FIG. 59 CONT'D

Representative Sequences from Each Species			SEQ ID NOs:
gorilla	526948	IQKIHFYFNCSDYGINCSHSYGCCGRSCIALFCSVGKLC	315
mangabey	1353842	IQKIHFYFNCSDYGINCSHSYSLCGRSCIALFCSDSKLC	316
chimpanzee	13529	IQKIHFYFNCSDYGINCSHSYGCCGRSCIALFCSVSKLC	317
human	1227	IQKIHFYFNCSDYGINCSHSYGCCSRSCIALFCSVGKLC	318

Figure 59: tBLASTx results from orthologous and non-orthologous sets with at least one Toxin_18+ ORF. Examination of CTXLP encoding loci in three non-human primate genomes, *Pan troglodytes* (Common chimpanzee), *Gorilla gorilla gorilla* (Western lowland gorilla), and *Cercocebus atys* (Sooty Mangabey), as well as humans reveals orthologous loci and conservation of the Toxin_18 cysteine motif (yellow). The name of each sequence is the Douville lab Retroexplorer database retroelement identification number of the locus encoding it. Entries with overlapping genomic regions were treated as one single entry during the tBLASTx search and merged accordingly. Orthology was determined by pairwise best BLAST matches of whole Retroexplorer retroelement entries and their flanking 1000bp, which mostly correspond to entire ERVs, but which sometimes were fragments. Three and four-way orthology was determined from pairwise orthology. Some loci do not have any species-specific sequence in the alignment, as the orthologous region in that species did not return a tBLASTx result. Non-orthologous sequences represent entries for which no orthologue was identified (possible paralogues or unique insertions). Representative sequences from each from each primate species highlights the degree of conservation in the Toxin_18 cysteine motif (yellow).

FIG 60

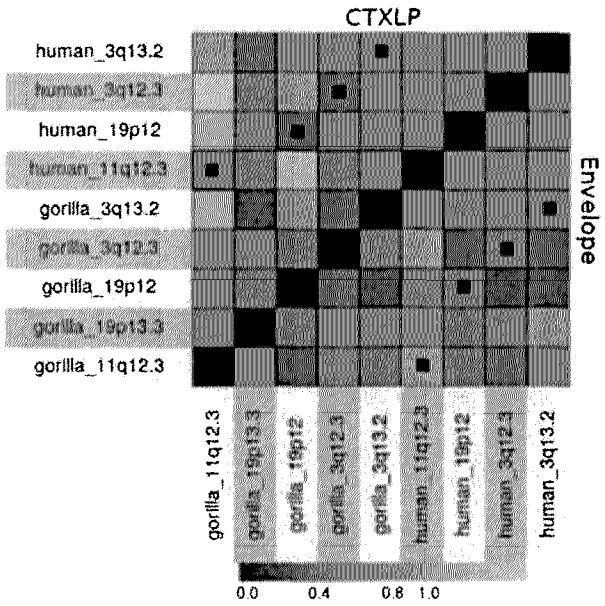


Figure 60: Different mutational patterns between orthologues and paralogues of ERVK env genes. Represented is a combined set of heatmaps generated by superheat from frames 0 (CTXLP) and 1 (Envelope) of the human and gorilla orthologues ORFs, which where both are positive for Toxin_18 positive (CTXLP). The sequences which are orthologues are indicated by black squares in the center of each space (paralogues do not have black squares). Blue is an ω (dN/dS ratio) less than 1, indicating purifying selection and similarity between the sequences. Yellow is an ω more than 1, indicating diversifying selection and dissimilarity between the sequences. Grey indicates that ω could not be computed (in all cases dS = 0, indicating no synonymous differences between the two sequences). Sequences which are identical along the diagonal are blacked out. The cytological bands are based on the human genome, with *Gorilla designations indicating their respective human* orthologue/paralogue coordinate. It is notable that orthologous sequences have a low ω when it is defined (when undefined it still has a low dN value). This suggests that there is more conservation between orthologues in different species than between paralogues from the same species. This pattern *is much more apparent for CTXLP than for Env reading frame, where differences are smaller if present* at all.

FIG. 61

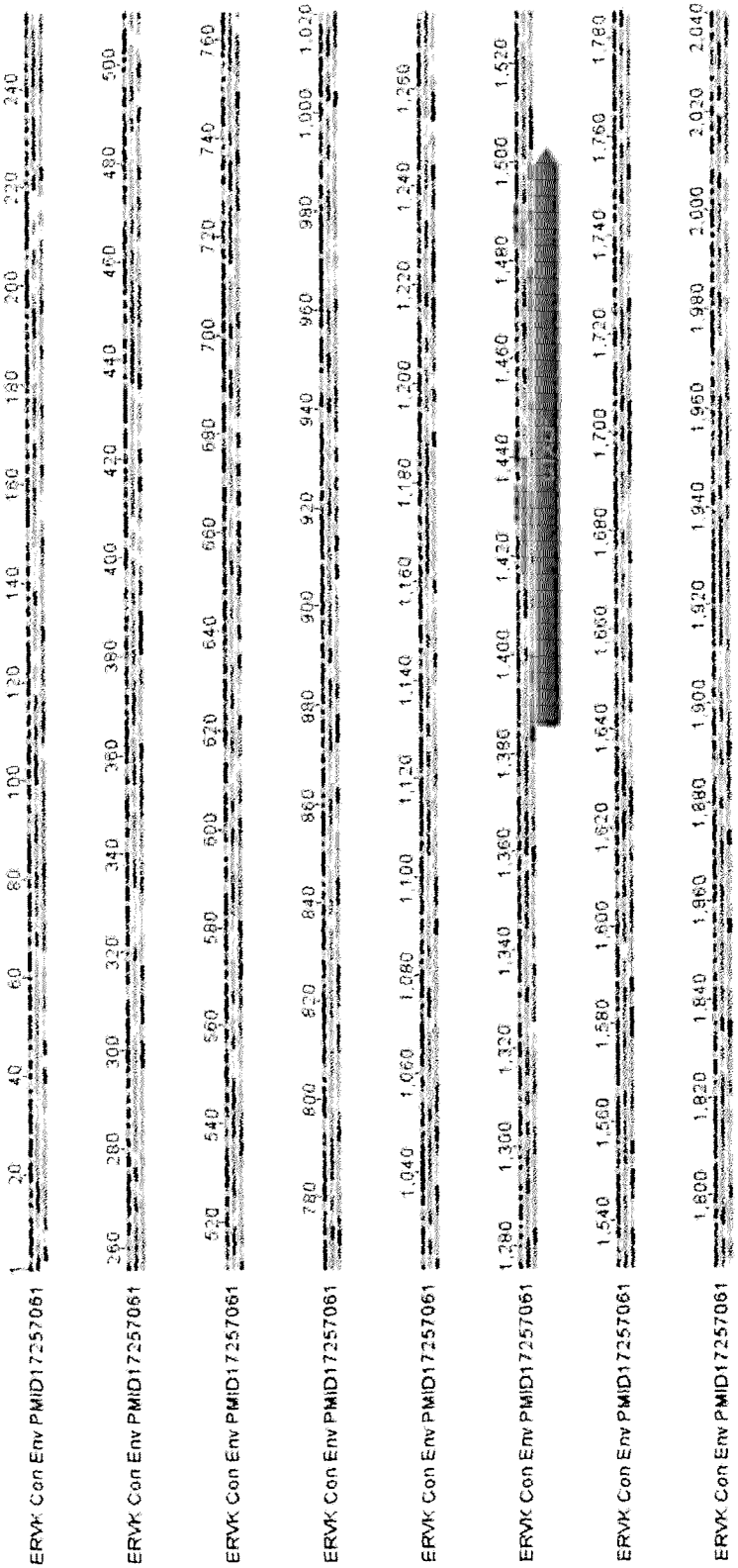


Figure 61: The transgene employed in the generation of ERVK envelope transgenic mice encodes CTXLP. The viral gene insert for the vector used in the generation of ERVK envelope transgenic mice[8], was analysed for the potential to encode and produce ERVK CTXLP. The red annotation indicates the location of CTXLP in the transgene insert.

FIG 62

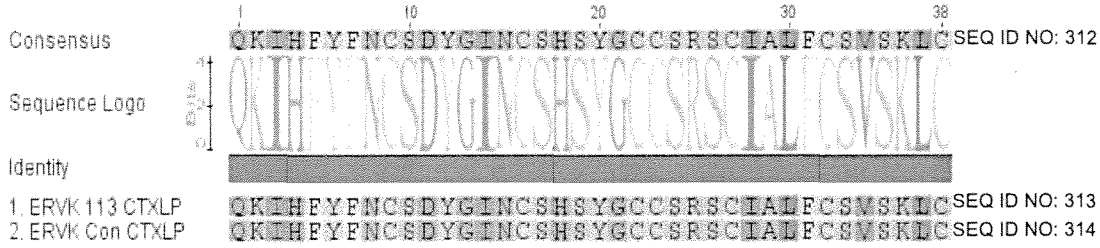


Figure 62: Alignment of ERVK113 CTXLP sequence and the ERVK Consensus sequence found in the ERVK envelope transgenic mice[8]. Note the similarity and retention of key cysteine motif.

FIG 63

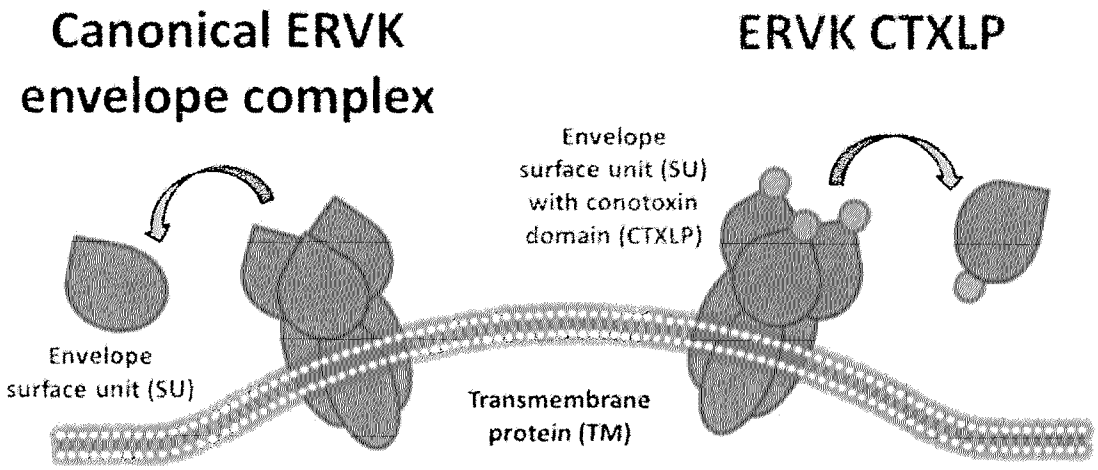


Figure 63: The ERVK Env and CTXLP proteins. Env is composed of the SU and TM subunits and is the prototypical gene product of the *env* gene. CTXLP is composed of the SU subunit and a C-terminal omega conotoxin domain and is encoded in an alternate open reading frame and may be produced due to ribosomal frameshifting.

FIG 64

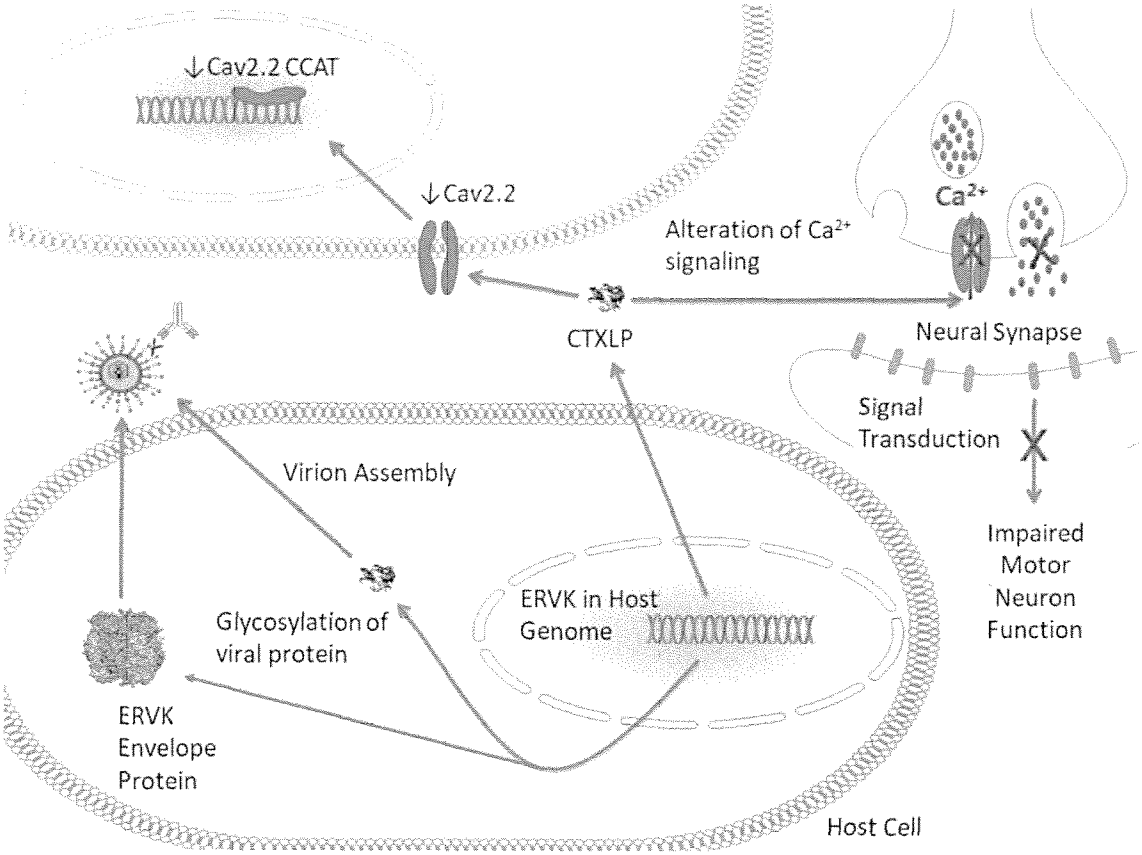


Figure 64: Illustration of the disruption of voltage-gated calcium channel CaV2.2 by ERVK CTXLP. Both canonical endogenous retrovirus-K (ERVK) envelope protein and conotoxin-like protein (CTXLP) can be produced from the ERVK *env* gene. ERVK CTXLP disrupts CaV2.2 on multiple levels, by decreasing CaV2.2 channel expression, as well as depleting the CaV2.2 calcium channel-associated transcription regulator (CCAT) in the nucleus. CTXLP inhibition of voltage gated calcium ion channel CaV2.2 may preventing neurotransmitter release and the continuation of signal transduction in the postsynaptic neurons.

FIG 65

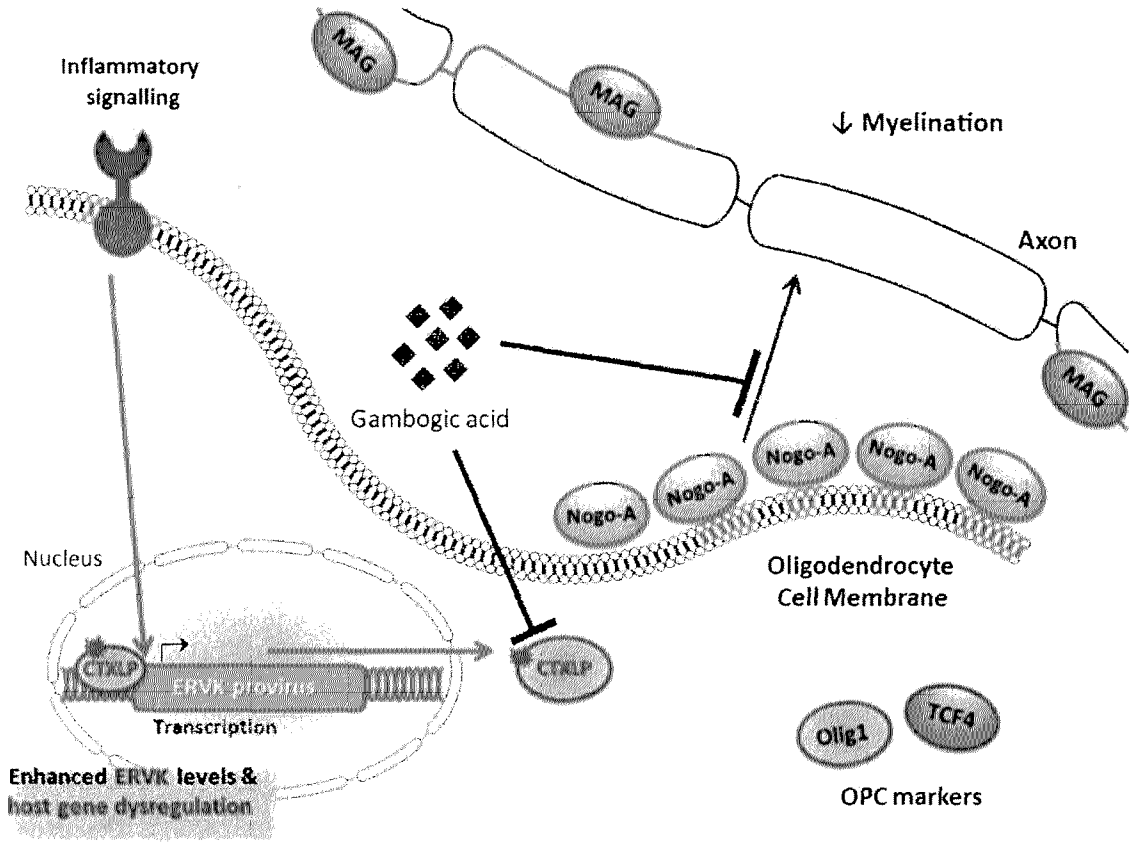


Figure 65: Pathological implications of CTXLP expression in oligodendrocyte precursor cells. CTXLP expression is enhanced in the presence of pro-inflammatory cytokines, including TNFα. Increased CTXLP protein levels in *ex vivo* human spinal cord tissue coincides with an increase in oligodendrocyte precursor cell (OPC) markers, transcription factor 4 (TCF4) and Olig1, along with elevated neurite outgrowth inhibitor A (Nogo-A) levels in adjacent cells and decreased myelin associated glycoprotein (MAG) axonal/myelin expression, which all suggest oligodendrocyte (OL) pathology. Nogo-A is a regulator of OPC differentiation, inhibitor of OL myelination, and axonal growth cone collapse during axon regeneration in the CNS. MAG is a marker for differentiated OLs and highly expressed in myelinating OLs. Based on expression patterns of OPC and OL markers, this suggests that heightened CTXLP expression in ALS is associated with OPCs being arrested in an immature state and inhibition of OL myelination. Treatment with CTXLP inhibitors, including gambogic acid, may reduce inflammatory signalling and decrease OPC and OL pathology by inhibiting increased Nogo-A expression.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2018/051306

A. CLASSIFICATION OF SUBJECT MATTER
IPC: ***C07K 14/15*** (2006.01) , ***A61K 31/35*** (2006.01) , ***A61K 31/575*** (2006.01) , ***A61K 31/7088*** (2006.01) , ***A61K 39/395*** (2006.01) , ***A61P 25/00*** (2006.01) (more IPCs on the last page)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPCs across all classifications

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

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)

Genomequest searched for SEQ ID NO:1, **Google:** Keywords: Douville, Mamneet and University of Winnipeg; **PubMed:** Keywords: Endogenous retrovirus K, conotoxin, omega and virus; **ORBIT:** Keywords: conotoxin, endogenous, retrovirus and omega

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,	US20050228172 A9 (DAVID WANG) 13 October 2005 (13-10-2005) (abstract) *see entire document*	1-11, 53 and 54; 12-14, 52, 73 and 74
X,	US7205146B1 (KEITH et al.) 17 April 2007 (17-04-2007) (abstract) *see entire document*	1-11, 52-54 and 73-76
X,Y	US7964341B2 (LADEROUTE et al.) 21 June 2011 (21-06-2011) (abstract) *see entire document*	1-11, 53, 54, 73 and 74; 13-16, 19, 20, 23, 24, 32-35, 38, 39, 42, 43, 51, 52-54, 55, 56, 59, 60, 64, 72, 75 and 76

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent but published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 January 2019 (11-01-2019)	Date of mailing of the international search report 22 January 2019 (22-01-2019)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 819-953-2476	Authorized officer Stephanie Etchells (819) 639-7796

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,	US7745391 (MINTZ et al.) 29 June 2010 (29-06-2010) (abstract) *see entire document*	1-6
X,	US7776523 (GARCIA et al.) 17 August 2010 (17-08-2010) (abstract) *see entire document*	1-6
Y	MAMNEET MANGHERA, "*Re-activation of Human Endogenous Retrovirus-K in Neuroinflammatory Disease*". Univeristy of Winnipeg Master's Thesis, 31 July 2015 (31-07-2015), Vol. 1, pp. 1-199, (abstract) [online] [retrieved on 10 January 2019 (10-01-2019)]. *see pages vii and 16*	13-16, 19, 20, 23, 24, 32-35, 38, 39, 42, 43, 51, 52-54, 55, 56, 59, 60, 64, 72, 75 and 76

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US2005228172A9	13 October 2005 (13-10-2005)	US2004181048A1	16 September 2004 (16-09-2004)
		US2005228172A9	13 October 2005 (13-10-2005)
		USH2191H	05 June 2007 (05-06-2007)
US7205146B1	17 April 2007 (17-04-2007)	US7205146B1	17 April 2007 (17-04-2007)
		AU5351201A	30 October 2001 (30-10-2001)
		AU2002307369A1	28 October 2002 (28-10-2002)
		CA2405078A1	25 October 2001 (25-10-2001)
		EP1317532A2	11 June 2003 (11-06-2003)
		US6683165B1	27 January 2004 (27-01-2004)
		US6737519B1	18 May 2004 (18-05-2004)
		US7407804B1	05 August 2008 (05-08-2008)
		US2004077011A1	22 April 2004 (22-04-2004)
		US7501118B2	10 March 2009 (10-03-2009)
		US7928200B1	19 April 2011 (19-04-2011)
		US2009292002A1	26 November 2009 (26-11-2009)
		US8105826B2	31 January 2012 (31-01-2012)
		US2011201782A1	18 August 2011 (18-08-2011)
		US8124734B2	28 February 2012 (28-02-2012)
		US2012100146A1	26 April 2012 (26-04-2012)
		US8907067B2	09 December 2014 (09-12-2014)
		US2003138925A1	24 July 2003 (24-07-2003)
		US2004002470A1	01 January 2004 (01-01-2004)
		US2004023215A1	05 February 2004 (05-02-2004)
		US2013039914A1	14 February 2013 (14-02-2013)
		US2014248271A1	04 September 2014 (04-09-2014)
		US2015274816A1	01 October 2015 (01-10-2015)
		US2015361162A1	17 December 2015 (17-12-2015)
		US2016222098A1	04 August 2016 (04-08-2016)
		US2016326239A1	10 November 2016 (10-11-2016)
		US2017066818A1	09 March 2017 (09-03-2017)
		WO0178894A2	25 October 2001 (25-10-2001)
		WO0178894A9	27 December 2001 (27-12-2001)
		WO0178894A3	20 March 2003 (20-03-2003)
		WO02083077A2	24 October 2002 (24-10-2002)
		WO02083077A3	12 September 2003 (12-09-2003)
US7964341B2	21 June 2011 (21-06-2011)	US2009017447A1	15 January 2009 (15-01-2009)
		US7964341B2	21 June 2011 (21-06-2011)
		CA2501301A1	18 September 2006 (18-09-2006)
		CA2673395A1	21 September 2006 (21-09-2006)
		CA2673395C	22 October 2013 (22-10-2013)
		US2012021403A1	26 January 2012 (26-01-2012)
		WO2006096985A1	21 September 2006 (21-09-2006)

Continued on the next page

Continued from the previous page

US7745391B2	29 June 2010 (29-06-2010)	US2007083334A1	12 April 2007 (12-04-2007)
		US7745391B2	29 June 2010 (29-06-2010)
		AU1410801A	30 May 2001 (30-05-2001)
		AU2005206388A1	04 August 2005 (04-08-2005)
		AU2005206388A2	04 August 2005 (04-08-2005)
		AU2005206389A1	04 August 2005 (04-08-2005)
		AU2005207625A1	04 August 2005 (04-08-2005)
		AU2005207882A1	11 August 2005 (11-08-2005)
		AU2005207883A1	11 August 2005 (11-08-2005)
		AU2005207886A1	11 August 2005 (11-08-2005)
		AU2005248530A1	08 December 2005 (08-12-2005)
		AU2005276208A1	02 March 2006 (02-03-2006)
		AU2005288710A1	06 April 2006 (06-04-2006)
		AU2005288710A2	06 April 2006 (06-04-2006)
		AU2005320352A1	31 August 2006 (31-08-2006)
		AU2005320352A2	31 August 2006 (31-08-2006)
		AU2005320352A8	30 April 2009 (30-04-2009)
		AU2006256374A1	14 December 2006 (14-12-2006)
		AU2006296171A1	05 April 2007 (05-04-2007)
		CA2554440A1	11 August 2005 (11-08-2005)
		CA2554585A1	04 August 2005 (04-08-2005)
		CA2554599A1	11 August 2005 (11-08-2005)
		CA2554623A1	11 August 2005 (11-08-2005)
		CA2554703A1	08 December 2005 (08-12-2005)
		CA2554707A1	02 March 2006 (02-03-2006)
		CA2554718A1	06 April 2006 (06-04-2006)
		CA2555509A1	27 July 2005 (27-07-2005)
		CA2611409A1	14 December 2006 (14-12-2006)
		CA2624535A1	05 April 2007 (05-04-2007)
		EP1713827A2	25 October 2006 (25-10-2006)
		EP1713900A2	25 October 2006 (25-10-2006)
		EP1713900A4	17 June 2009 (17-06-2009)
		EP1716227A2	02 November 2006 (02-11-2006)
		EP1716227A4	06 January 2010 (06-01-2010)
		EP1716256A2	02 November 2006 (02-11-2006)
		EP1721257A2	15 November 2006 (15-11-2006)
		EP1730181A2	13 December 2006 (13-12-2006)
		EP1730183A2	13 December 2006 (13-12-2006)
		EP1732943A2	20 December 2006 (20-12-2006)
		EP1732943A4	12 January 2011 (12-01-2011)
		EP1735342A2	27 December 2006 (27-12-2006)
		EP1735468A2	27 December 2006 (27-12-2006)
		EP1749025A2	07 February 2007 (07-02-2007)
		EP1774046A2	18 April 2007 (18-04-2007)
		EP1774046A4	15 September 2010 (15-09-2010)
		EP1851543A2	07 November 2007 (07-11-2007)
		EP1891218A2	27 February 2008 (27-02-2008)
		EP1934256A2	25 June 2008 (25-06-2008)
		EP2816351A2	24 December 2014 (24-12-2014)
		EP2816351A3	25 March 2015 (25-03-2015)
		IL132978D0	19 March 2001 (19-03-2001)
		IL133455D0	30 April 2001 (30-04-2001)
		IL135619D0	20 May 2001 (20-05-2001)
		IL136776D0	14 June 2001 (14-06-2001)
		IL177186D0	10 December 2006 (10-12-2006)
		IL185501D0	20 January 2008 (20-01-2008)
		IL187946D0	20 March 2008 (20-03-2008)
		IL190408D0	03 November 2008 (03-11-2008)

Continued on the next page

Continued from the previous page

JP2007520217A	26 July 2007 (26-07-2007)
JP2007525213A	06 September 2007 (06-09-2007)
JP2007526763A	20 September 2007 (20-09-2007)
JP2008507261A	13 March 2008 (13-03-2008)
US2002110811A1	15 August 2002 (15-08-2002)
US6936450B2	30 August 2005 (30-08-2005)
US2005277156A1	15 December 2005 (15-12-2005)
US7332569B2	19 February 2008 (19-02-2008)
US2006172311A1	03 August 2006 (03-08-2006)
US7345142B2	18 March 2008 (18-03-2008)
US2006051774A1	09 March 2006 (09-03-2006)
US7368548B2	06 May 2008 (06-05-2008)
US2006257896A1	16 November 2006 (16-11-2006)
US7488813B2	10 February 2009 (10-02-2009)
US2007259386A1	08 November 2007 (08-11-2007)
US7528243B2	05 May 2009 (05-05-2009)
US2009004648A1	01 January 2009 (01-01-2009)
US7553948B2	30 June 2009 (30-06-2009)
US2006046257A1	02 March 2006 (02-03-2006)
US7569662B2	04 August 2009 (04-08-2009)
US2007092484A1	26 April 2007 (26-04-2007)
US7601692B2	13 October 2009 (13-10-2009)
US7667001B1	23 February 2010 (23-02-2010)
US2008159992A1	03 July 2008 (03-07-2008)
US7678769B2	16 March 2010 (16-03-2010)
US2008193941A1	14 August 2008 (14-08-2008)
US7714100B2	11 May 2010 (11-05-2010)
US2009202991A1	13 August 2009 (13-08-2009)
US7741433B2	22 June 2010 (22-06-2010)
US2009036374A1	05 February 2009 (05-02-2009)
US7758862B2	20 July 2010 (20-07-2010)
US2010120022A1	13 May 2010 (13-05-2010)
US7842459B2	30 November 2010 (30-11-2010)
US2010124786A1	20 May 2010 (20-05-2010)
US7906635B2	15 March 2011 (15-03-2011)
US2010075891A1	25 March 2010 (25-03-2010)
US7939634B2	10 May 2011 (10-05-2011)
US2004101876A1	27 May 2004 (27-05-2004)
US2004142325A1	22 July 2004 (22-07-2004)
US2004248157A1	09 December 2004 (09-12-2004)
US2006014166A1	19 January 2006 (19-01-2006)
US2006040278A1	23 February 2006 (23-02-2006)
US2006068405A1	30 March 2006 (30-03-2006)
US2006147946A1	06 July 2006 (06-07-2006)
US2006172378A1	03 August 2006 (03-08-2006)
US2006183131A1	17 August 2006 (17-08-2006)
US2006263786A1	23 November 2006 (23-11-2006)
US2007082337A1	12 April 2007 (12-04-2007)
US2007219125A1	20 September 2007 (20-09-2007)
US2008014590A1	17 January 2008 (17-01-2008)
US2008182299A1	31 July 2008 (31-07-2008)
US2009075257A1	19 March 2009 (19-03-2009)
US2009215042A1	27 August 2009 (27-08-2009)
US2009215046A1	27 August 2009 (27-08-2009)
US2009286957A1	19 November 2009 (19-11-2009)
US2010068736A1	18 March 2010 (18-03-2010)
US2010183573A1	22 July 2010 (22-07-2010)

Continued on the next page

Continued from the previous page

US2010189652A1	29 July 2010 (29-07-2010)
US2010248270A1	30 September 2010 (30-09-2010)
US2011091454A1	21 April 2011 (21-04-2011)
US2011183924A1	28 July 2011 (28-07-2011)
US2011212051A1	01 September 2011 (01-09-2011)
WO0136632A2	25 May 2001 (25-05-2001)
WO0136632A3	03 January 2002 (03-01-2002)
WO2004096979A2	11 November 2004 (11-11-2004)
WO2004096979A3	10 August 2006 (10-08-2006)
WO2004096980A2	11 November 2004 (11-11-2004)
WO2004096980A3	03 August 2006 (03-08-2006)
WO2005069724A2	04 August 2005 (04-08-2005)
WO2005069724A3	10 August 2006 (10-08-2006)
WO2005071058A2	04 August 2005 (04-08-2005)
WO2005071058A3	08 November 2007 (08-11-2007)
WO2005071059A2	04 August 2005 (04-08-2005)
WO2005071059A3	12 February 2009 (12-02-2009)
WO2005072049A2	11 August 2005 (11-08-2005)
WO2005072049A3	28 December 2006 (28-12-2006)
WO2005072050A2	11 August 2005 (11-08-2005)
WO2005072050A9	14 January 2010 (14-01-2010)
WO2005072053A2	11 August 2005 (11-08-2005)
WO2005072053A9	04 March 2010 (04-03-2010)
WO2005072055A2	11 August 2005 (11-08-2005)
WO2005072055A9	17 February 2011 (17-02-2011)
WO2005072340A2	11 August 2005 (11-08-2005)
WO2005072340A3	02 October 2008 (02-10-2008)
WO2005084116A2	15 September 2005 (15-09-2005)
WO2005084116A8	01 December 2005 (01-12-2005)
WO2005084116A3	16 April 2009 (16-04-2009)
WO2005107364A2	17 November 2005 (17-11-2005)
WO2005107364A9	11 May 2006 (11-05-2006)
WO2005107364A3	16 April 2009 (16-04-2009)
WO2005116850A2	08 December 2005 (08-12-2005)
WO2005116850A9	07 December 2006 (07-12-2006)
WO2005116850A3	17 June 2010 (17-06-2010)
WO2006021874A2	02 March 2006 (02-03-2006)
WO2006021874A8	15 March 2007 (15-03-2007)
WO2006021874A3	09 April 2009 (09-04-2009)
WO2006035273A2	06 April 2006 (06-04-2006)
WO2006035273A3	16 April 2009 (16-04-2009)
WO2006043271A1	27 April 2006 (27-04-2006)
WO2006054297A2	26 May 2006 (26-05-2006)
WO2006054297A3	30 April 2009 (30-04-2009)
WO2006072954A2	13 July 2006 (13-07-2006)
WO2006072954A3	22 May 2009 (22-05-2009)
WO2006090389A2	31 August 2006 (31-08-2006)
WO2006090389A3	29 November 2007 (29-11-2007)
WO2006131783A2	14 December 2006 (14-12-2006)
WO2006131783A3	27 August 2009 (27-08-2009)
WO2006131928A2	14 December 2006 (14-12-2006)
WO2006131928A3	07 June 2007 (07-06-2007)
WO2006131928A8	19 July 2007 (19-07-2007)
WO2007036945A2	05 April 2007 (05-04-2007)
WO2007036945A3	13 September 2007 (13-09-2007)

Continued on the next page

Continued from the previous page

US7776523B2	17 August 2010 (17-08-2010)	US2010136522A1	03 June 2010 (03-06-2010)
		US7776523B2	17 August 2010 (17-08-2010)
		AT458003T	15 March 2010 (15-03-2010)
		AU2736502A	18 June 2002 (18-06-2002)
		AU2002351332A1	23 June 2003 (23-06-2003)
		AU2002351332A8	23 June 2003 (23-06-2003)
		AU2002368276A1	13 May 2004 (13-05-2004)
		AU2002368276A8	13 May 2004 (13-05-2004)
		CA2430379A1	13 June 2002 (13-06-2002)
		CA2468903A1	06 May 2004 (06-05-2004)
		CA2469049A1	19 June 2003 (19-06-2003)
		DE60235413D1	01 April 2010 (01-04-2010)
		EP1415005A2	06 May 2004 (06-05-2004)
		EP1415005B1	21 November 2012 (21-11-2012)
		EP1521594A2	13 April 2005 (13-04-2005)
		EP1521594A4	21 March 2007 (21-03-2007)
		EP1521594B1	02 October 2013 (02-10-2013)
		EP1578919A2	28 September 2005 (28-09-2005)
		EP1578919A4	25 July 2007 (25-07-2007)
		EP1578919B1	17 February 2010 (17-02-2010)
		EP2208736A2	21 July 2010 (21-07-2010)
		EP2208736A3	27 October 2010 (27-10-2010)
		EP2336368A1	22 June 2011 (22-06-2011)
		EP2339035A1	29 June 2011 (29-06-2011)
		ES2339433T3	20 May 2010 (20-05-2010)
		ES2397627T3	08 March 2013 (08-03-2013)
		JP2006508690A	16 March 2006 (16-03-2006)
		JP4646315B2	09 March 2011 (09-03-2011)
		JP2005512525A	12 May 2005 (12-05-2005)
		JP4822490B2	24 November 2011 (24-11-2011)
		JP2004535765A	02 December 2004 (02-12-2004)
		JP2008212152A	18 September 2008 (18-09-2008)
		JP2009100768A	14 May 2009 (14-05-2009)
		JP2009131287A	18 June 2009 (18-06-2009)
		JP2012125260A	05 July 2012 (05-07-2012)
		JP2013165722A	29 August 2013 (29-08-2013)
		US2006275747A1	07 December 2006 (07-12-2006)
		US2007037147A1	15 February 2007 (15-02-2007)
		US2011020352A1	27 January 2011 (27-01-2011)
		US2014135384A1	15 May 2014 (15-05-2014)
		WO0246477A2	13 June 2002 (13-06-2002)
		WO0246477A3	19 February 2004 (19-02-2004)
		WO03050258A2	19 June 2003 (19-06-2003)
		WO03050258A3	17 February 2005 (17-02-2005)
		WO2004037972A2	06 May 2004 (06-05-2004)
		WO2004037972A3	04 May 2006 (04-05-2006)

A61P 31/00 (2006.01) , ***A61P 31/12*** (2006.01) , ***A61P 35/00*** (2006.01) , ***C07D 493/04*** (2006.01) , ***C07K 16/10*** (2006.01) , ***C12N 15/48*** (2006.01) , ***C12N 5/10*** (2006.01) , and ***C12P 21/02*** (2006.01)