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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 January 2019 (31.01.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/023291 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US2018/043588

(22) International Filing Date:
25 July 2018 (25.07.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/536,870 25 July 2017 (25.07.2017) US

(71) Applicant: **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 450 Brookline Avenue, Office of General Counsel 10BP-372A, Boston, Massachusetts 02215 (US).

(72) Inventors: **LIU, Xiaole**; 26 Overlook Road, Wayland, Massachusetts 01778 (US). **BROWN, Myles**; 450 Brookline Avenue, Boston, Massachusetts 02215 (US). **PENG, Jingyu**; 450 Brookline Avenue, Boston, Massachusetts 02215 (US). **XIAO, Tengfei**; 450 Brookline Avenue, Boston, Massachusetts 02215 (US). **LI, Wei**; 450 Brookline Avenue, Boston, Massachusetts 02215 (US).

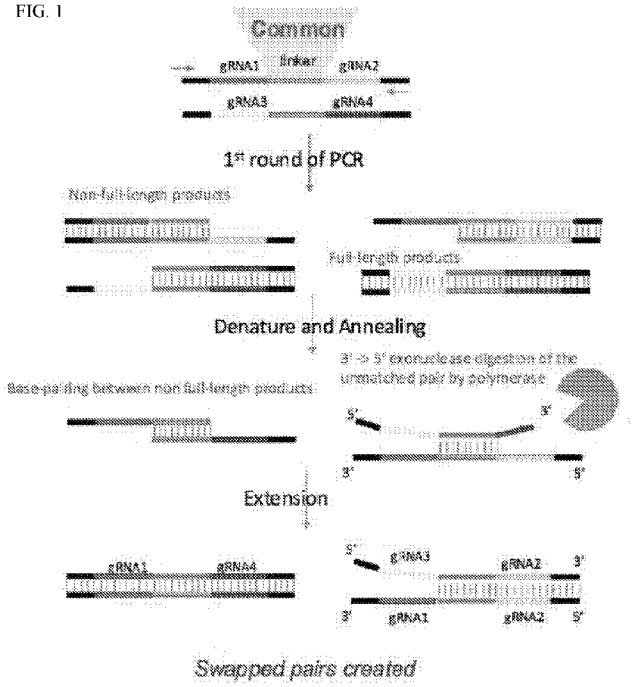
(74) Agent: **CLARKE, Daniel W.** et al.; BURNS & LEVINSON, LLP, 125 Summer Street, Boston, Massachusetts 02110 (US).

(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: COMPOSITIONS AND METHODS FOR MAKING AND DECODING PAIRED-GUIDE RNA LIBRARIES AND USES THEREOF

FIG. 1



(57) Abstract: The present disclosure relates to compositions and methods for making and decoding paired-guide RNA (pgRNA) libraries using the Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) system, and using the resulting pgRNA/CRISPR libraries to identify genetic interactions or functional non-coding elements.

[Continued on next page]

WO 2019/023291 A2

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

COMPOSITIONS AND METHODS FOR MAKING AND DECODING PAIRED-GUIDE RNA LIBRARIES AND USES THEREOF

RELATED APPLICATIONS

This application claims the benefit of priority under 35 U.S.C. § 119(e) to U.S. Provisional Application No: 62/536,870, filed July 25, 2017, which is incorporated herein by reference in its entirety.

FIELD OF THE DISCLOSURE

The disclosure relates to compositions and methods for making and decoding paired-guide RNA (pgRNA) libraries using the Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) system, and using the pgRNA/CRISPR libraries to identify synthetic lethal genetic interactions (SLGIs) and functional cis-elements (e.g., enhancers).

BACKGROUND OF THE DISCLOSURE

Cancer is a disease in which abnormal cells divide without control and can invade nearby tissues (*i.e.*, metastasize). According to the World Health Organization, cancer is one of the leading causes of morbidity and mortality worldwide, and was responsible for 8.8 million deaths in 2015. Globally, cancer is responsible for nearly 1 in 6 deaths. In 2015, the most common cancer deaths occurred from the following types of cancer: lung cancer (1.69 million deaths), liver cancer (788,000 deaths), colorectal cancer (774,000 deaths), stomach cancer (754,000 deaths), and breast cancer (571,000 deaths).

Cancer is typically treated by any of a variety of methods such as surgery, chemotherapy, radiation therapy, immunotherapy, etc. Unfortunately, many of these methods have toxic/undesirable side effects. For example, standard chemotherapies for cancer were initially developed based on their ability to kill rapidly dividing cells, and many of their common side effects (*e.g.*, immunosuppression, nausea, hair loss, and the like) are due to their toxic effects on normal tissues that include cell types that undergo rapid division. A central goal of cancer research over the past two decades has been to identify new therapies having great

efficaciousness and fewer side effects. To this end, cancer research has focused on discovering tumor-specific traits that may be exploited for selective targeting.

One such approach involves screening for synthetic lethal genetic interactions (SLGIs), which occur when inhibition of two non-essential genes results in a lethal phenotype. The presence of a mutation that inhibits one of the non-essential genes in cancer cells, but not in normal cells, therefore creates an opportunity to selectively kill the cancer cells with a targeted therapy that reduces or eliminates expression of the second non-essential gene. The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system is a revolutionary approach for genome editing and functional genomics research in mammalian systems that may be used to knockout (KO) any pair of genes separately or simultaneously to identify SLGIs, and/or non-coding elements that is essential for cancer cell growth. The development of lentiviral delivery of a genome-scale CRISPR/Cas9 KO library targeting all genes enables both negative and positive selection screening on mammalian cell lines in a cost-effective manner. Unfortunately, prior art paired-guide RNA (pgRNA) CRISPR/Cas9 KO libraries suffer from the significant disadvantage that they are prone to recombination during construction that creates undesirable constructs, and such libraries are therefore not amenable to scaling. Accordingly, there remains an urgent unmet need for the construction of high-quality, recombination-free pgRNA/CRISPR libraries that allow for reliable, scalable functional genomics studies to identify SLGIs and non-coding elements that may be useful in the treatment of cancer.

SUMMARY OF THE DISCLOSURE

The present disclosure provides paired-guide RNA (pgRNA)/Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) libraries having reduced or eliminated rates of internal pgRNA swapping/recombination that may be constructed by using vectors that include two guide RNA (gRNA) cassettes, each having a general structure of promoter-gRNA-scaffold that are constructed from a synthesized oligonucleotide having a general structure of gRNA-1 cassette—unique linker—gRNA-2 cassette such that the unique linker is removed from the final vector containing the two gRNA cassettes. The promoter used in each gRNA cassette may be different, for example, a gRNA-1 cassette may use a human U6 promoter while a paired gRNA-2 cassette may use a mouse U6 promoter. Additionally, the scaffold sequence in each gRNA cassette will typically be different. The present disclosure provides compositions and methods

for making and decoding pgRNA libraries using the CRISPR system. Advantageously, the pgRNA/CRISPR libraries disclosed herein may be used to identify synthetic lethal genetic interactions (SLGI) and functional non-coding elements. The techniques provided herein are important because identifying and characterizing SLGIs that occur in combination with tumor suppressor genes may provide novel therapies with which to treat cancer.

In one aspect, the present disclosure provides a paired-guide ribonucleic acid (pgRNA) vector that includes a first guide RNA (gRNA) cassette, a second gRNA cassette; and a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated protein 9 (Cas9) expression cassette in which the second gRNA cassette is positioned between the first gRNA cassette and the Cas9 expression cassette.

In one aspect, the disclosure provides an intermediate paired-guide RNA (pgRNA) nucleic acid that includes a first guide RNA (gRNA); a unique linker; and a second gRNA configured so that the unique linker is positioned between the first gRNA and the second gRNA.

In an exemplary embodiment, the first gRNA cassette may include a first nucleic acid sequence including, in 5' to 3' order, a first gRNA promoter, a first gRNA, and a first gRNA scaffold, and the second gRNA cassette may include a second nucleic acid sequence including, in 5' to 3' order, a second gRNA promoter, a second gRNA, and a second gRNA scaffold.

In an exemplary embodiment, the first gRNA promoter may be selected from a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and/or a modified bovine 7SK promoter.

In an exemplary embodiment, the second gRNA promoter may be selected from the group consisting of a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and/or a modified bovine 7SK promoter.

In an exemplary embodiment, the second gRNA promoter may be different than the first gRNA promoter.

In an exemplary embodiment, the first gRNA and the second gRNA may each be between about 17 and 27 nucleotides in length. In an exemplary embodiment, the first gRNA and the second gRNA are each about 19 nucleotides in length.

In an exemplary embodiment, the pgRNA vector may be constructed by using an intermediate pgRNA nucleic acid that includes a first gRNA cassette, a unique linker, and a second gRNA cassette in which the unique linker is positioned between the first gRNA cassette and the second gRNA cassette.

In an exemplary embodiment, the unique linker may be between about 10 and 30 nucleotides in length. In an exemplary embodiment, the unique linker may be about 16 nucleotides in length.

In an exemplary embodiment, the Cas9 cassette may include a promoter, a Cas9 coding sequence, and a P2A sequence. In an exemplary embodiment, the promoter may be an EF-1α or a CMV promoter.

In an exemplary embodiment the unique linker may have a GC content of less than or equal to 40%.

In one aspect, the present disclosure provides a method of making a paired-guide RNA (pgRNA) library vector that may include the steps of: obtaining a first nucleic acid sequence including, in 5' to 3' order, a first guide RNA (gRNA) cassette promoter, a vector linker, and a second gRNA cassette scaffold; removing the vector linker to create a double strand break (DSB) between a 3' end of the first gRNA cassette promoter and a 5' end of the second gRNA cassette scaffold; inserting into the DSB a second nucleic acid sequence including, in 5' to 3' order, a first guide RNA (gRNA) sequence, a unique linker, and a second gRNA sequence to create an intermediate nucleic acid sequence; removing the unique linker to create a DSB in the intermediate nucleic acid sequence between a 3' end of the first gRNA sequence and a 5' end of the second gRNA sequence; and inserting into the DSB in the intermediate nucleic acid sequence a third nucleic acid sequence including, in 5' to 3' order, a first gRNA cassette scaffold, a spacer, and a second guide RNA (gRNA) cassette promoter, thereby creating the pgRNA vector.

In an exemplary embodiment, the first gRNA cassette promoter may be selected from a mouse U6 promoter and/or a human U6 promoter. In an exemplary embodiment, the second gRNA cassette promoter may be selected from the group consisting of a mouse U6 promoter and/or a human U6 promoter. In an exemplary embodiment, the second gRNA cassette promoter may be different than the first gRNA cassette promoter.

In an exemplary embodiment, the first gRNA sequence and the second gRNA sequence may each be between about 17 and 27 nucleotides in length. In an exemplary embodiment, the

first gRNA sequence and the second gRNA sequence may each be about 19 nucleotides in length.

In an exemplary embodiment, the unique linker may be between about 12 and 24 nucleotides in length. In an exemplary embodiment, the unique linker may be about 16 nucleotides in length.

In an exemplary embodiment, the first nucleic acid sequence further includes a Cas9 cassette. In an exemplary embodiment, the Cas9 cassette includes a promoter, a Cas9 coding sequence, and a P2A sequence.

In one aspect, the present disclosure provides a paired-guide RNA (pgRNA)/Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) library that includes: a plurality of pgRNA sequence pairs capable of targeting a plurality of target sequence pairs in a target genome via a CRISPR/Cas9 system to knockout function of a first target sequence and a second target sequence in the target sequence pair, and where the pgRNA vector is constructed by using an intermediate pgRNA nucleic acid, that includes a first guide RNA (gRNA) cassette; a unique linker; and a second gRNA cassette; wherein the unique linker is positioned between the first gRNA cassette and the second gRNA cassette.

In an exemplary embodiment, each of the plurality of pgRNA sequence pairs may include a first guide RNA (gRNA) cassette and a second gRNA cassette.

In an exemplary embodiment, the first gRNA cassette may include a first nucleic acid sequence including, in 5' to 3' order, a first gRNA promoter, a first gRNA sequence, and a first gRNA scaffold, and the second gRNA cassette includes a second nucleic acid sequence including, in 5' to 3' order, a second gRNA promoter, a second gRNA sequence, and a second gRNA scaffold.

In an exemplary embodiment, the first gRNA promoter may be selected from a mouse U6 promoter and/or a human U6 promoter. In an exemplary embodiment, the second gRNA promoter may be selected from a mouse U6 promoter and/or a human U6 promoter. In an exemplary embodiment, the second gRNA promoter may be different than the first gRNA promoter.

In an exemplary embodiment, the first gRNA sequence and the second gRNA sequence may each be between about 17 and 27 nucleotides in length. In an exemplary embodiment, the

first gRNA sequence and the second gRNA sequence may each be about 19 nucleotides in length.

In an exemplary embodiment, the unique linker is between about 12 and 24 nucleotides in length. In an exemplary embodiment, the unique linker may be about 16 nucleotides in length.

In one aspect, the present disclosure provides a method of identifying synthetic lethal genetic interaction (SLGI) within a genome that includes the steps of: contacting a population of cells with one or more of the above-described pgRNA vectors; selecting successfully transduced cells; culturing the population of cells for a plurality of population doubling times, wherein genomic DNA may be harvested on a first day of culture and on a last day of culture; deep sequencing the genomic DNA harvested on the first day of culture and on the last day of culture; quantifying abundance of a first guide RNA (gRNA) included in the first gRNA cassette and a second guide RNA (gRNA) included in the second gRNA cassette at the first day of culture and the last day of culture; analyzing an abundance fold change of the first gRNA and the second gRNA between the first day of culture and the last day of culture; and identifying, based on the abundance fold change; a SLGI.

In an exemplary embodiment, the analyzing step further includes a regression residual analysis. In an exemplary embodiment, the analyzing step further includes a BLISS independence model analysis.

In an exemplary embodiment, the plurality of population doubling times may be between about 8 and 16. In an exemplary embodiment, the plurality of population doubling times may be about 12.

In one aspect, the disclosure provides a tangible, non-transitory, computer-readable media having software encoded thereon, the software, when executed by a processor on a particular device, may be operable to: identify a plurality of gene pairs; determine a response variable; analyze, by a feature selection and regression model, the plurality of gene pairs; and determine, based on the response variable and the analysis, that one or more gene pairs within the plurality of gene pairs interact genetically.

Definitions

Unless specifically stated or obvious from context, as used herein, the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard

deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1 %, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein can be modified by the term about.

As used herein, the term "primer" and its derivatives refers generally to any polynucleotide that can hybridize or anneal to a target sequence of interest. In some embodiments, the primer can also serve to prime nucleic acid synthesis. Typically, the primer functions as a substrate onto which nucleotides can be polymerized by a polymerase; in some embodiments, however, the primer can become incorporated into the synthesized nucleic acid strand and provide a site to which another primer can hybridize to prime synthesis of a new strand that is complementary to the synthesized nucleic acid molecule. The primer may be comprised of any combination of nucleotides or analogs thereof, which may be optionally linked to form a linear polymer of any suitable length. In some embodiments, the primer is a single-stranded oligonucleotide or polynucleotide. (For purposes of this disclosure, the terms "polynucleotide" and "oligonucleotide" and "oligo" are used interchangeably herein). In some embodiments, the primer is single-stranded but it can also be double-stranded. The primer optionally occurs naturally, as in a purified restriction digest, or can be produced synthetically. In some embodiments, the primer acts as a point of initiation for amplification or synthesis when exposed to amplification or synthesis conditions; such amplification or synthesis can occur in a template-dependent fashion and optionally results in formation of a primer extension product that is complementary to at least a portion of the target sequence. Exemplary amplification or synthesis conditions can include contacting the primer with a polynucleotide template (e.g., a template including a target sequence), nucleotides and an inducing agent such as a polymerase at a suitable temperature and pH to induce polymerization of nucleotides onto an end of the target-specific primer. If double-stranded, the primer can optionally be treated to separate its strands before being used to prepare primer extension products. In some embodiments, the primer is an oligodeoxyribonucleotide or an oligoribonucleotide. In some embodiments, the primer can include one or more nucleotide analogs. The exact length and/or composition, including sequence, of the target-specific primer can influence many properties, including melting temperature (Tm), GC content, formation of secondary structures, repeat nucleotide motifs, length of predicted primer extension products, extent of coverage across a nucleic acid molecule of interest, number of primers present in a single amplification or synthesis reaction, presence of

nucleotide analogs or modified nucleotides within the primers, and the like. In some embodiments, a primer can be paired with a compatible primer within an amplification or synthesis reaction to form a primer pair consisting of a forward primer and a reverse primer. In some embodiments, the forward primer of the primer pair includes a sequence that is substantially complementary to at least a portion of a strand of a nucleic acid molecule, and the reverse primer of the primer of the primer pair includes a sequence that is substantially identical to at least of portion of the strand. In some embodiments, the forward primer and the reverse primer are capable of hybridizing to opposite strands of a nucleic acid duplex. Optionally, the forward primer primes synthesis of a first nucleic acid strand, and the reverse primer primes synthesis of a second nucleic acid strand, wherein the first and second strands are substantially complementary to each other, or can hybridize to form a double-stranded nucleic acid molecule. In some embodiments, one end of an amplification or synthesis product is defined by the forward primer and the other end of the amplification or synthesis product is defined by the reverse primer. In some embodiments, where the amplification or synthesis of long primer extension products is required, such as amplifying an exon, coding region, or gene, several primer pairs can be created than span the desired length to enable sufficient amplification of the region. In some embodiments, a primer can include one or more cleavable groups. In some embodiments, primer lengths are in the range of about 10 to about 60 nucleotides, about 12 to about 50 nucleotides and about 15 to about 40 nucleotides in length. Typically, a primer is capable of hybridizing to a corresponding target sequence and undergoing primer extension when exposed to amplification conditions in the presence of dNTPs and a polymerase. In some instances, the particular nucleotide sequence or a portion of the primer is known at the outset of the amplification reaction or can be determined by one or more of the methods disclosed herein. In some embodiments, the primer includes one or more cleavable groups at one or more locations within the primer.

As used herein, “polymerase” and its derivatives, generally refers to any enzyme that can catalyze the polymerization of nucleotides (including analogs thereof) into a nucleic acid strand. Typically, but not necessarily, such nucleotide polymerization can occur in a template-dependent fashion. Such polymerases can include without limitation naturally occurring polymerases and any subunits and truncations thereof, mutant polymerases, variant polymerases, recombinant, fusion or otherwise engineered polymerases, chemically modified polymerases, synthetic

molecules or assemblies, and any analogs, derivatives or fragments thereof that retain the ability to catalyze such polymerization. Optionally, the polymerase can be a mutant polymerase comprising one or more mutations involving the replacement of one or more amino acids with other amino acids, the insertion or deletion of one or more amino acids from the polymerase, or the linkage of parts of two or more polymerases. Typically, the polymerase comprises one or more active sites at which nucleotide binding and/or catalysis of nucleotide polymerization can occur. Some exemplary polymerases include without limitation DNA polymerases and RNA polymerases. The term "polymerase" and its variants, as used herein, also refers to fusion proteins comprising at least two portions linked to each other, where the first portion comprises a peptide that can catalyze the polymerization of nucleotides into a nucleic acid strand and is linked to a second portion that comprises a second polypeptide. In some embodiments, the second polypeptide can include a reporter enzyme or a processivity-enhancing domain. Optionally, the polymerase can possess 5' exonuclease activity or terminal transferase activity. In some embodiments, the polymerase can be optionally reactivated, for example through the use of heat, chemicals or re-addition of new amounts of polymerase into a reaction mixture. In some embodiments, the polymerase can include a hot-start polymerase or an aptamer based polymerase that optionally can be reactivated.

As used herein, "primer/probe set" refers to a grouping of a pair of oligonucleotide primers and an oligonucleotide probe that hybridize to a specific nucleotide sequence. The oligonucleotide set in certain embodiments may include: (a) a forward discriminatory primer that hybridizes to a first location of a nucleic acid sequence or adjacent a particular mutation portion; (b) a reverse discriminatory primer that hybridizes to a second location of the nucleic acid sequence downstream of the first location and (c) preferably a fluorescent probe labeled with a fluorophore and a quencher, which hybridizes to a location of the nucleic acid sequence between the primers. In other words, an oligonucleotide set in certain embodiments consists of a set of specific PCR primers capable of initiating synthesis of an amplicon specific to screening for synthetic lethal genetic interactions (SLGIs) such as, for example, indel or point mutations, and may also include a fluorescent probe that hybridizes to the amplicon. The set may also include in other embodiments a probe with binds to or reacts with one or both of the primers where each or at least one of the primers is modified to contain a marker moiety (e.g., ligand that can be detected with a labeled antibody).

As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K. B. Mullis U.S. Pat. Nos. 4,683,195 and 4,683,202, hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a polynucleotide of interest in a mixture of genomic DNA without cloning or purification. This process for amplifying the polynucleotide of interest consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired polynucleotide of interest, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded polynucleotide of interest. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the polynucleotide of interest molecule. Following annealing, the primers are extended with a polymerase to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (i.e., denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired polynucleotide of interest. The length of the amplified segment of the desired polynucleotide of interest (amplicon) is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of repeating the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the polynucleotide of interest become the predominant nucleic acid sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified." As defined herein, target nucleic acid molecules within a sample including a plurality of target nucleic acid molecules are amplified via PCR. In a modification to the method discussed above, the target nucleic acid molecules can be PCR amplified using a plurality of different primer pairs, in some cases, one or more primer pairs per target nucleic acid molecule of interest, thereby forming a multiplex PCR reaction. Using multiplex PCR, it is possible to simultaneously amplify multiple nucleic acid molecules of interest from a sample to form amplified target sequences. It is also possible to detect the amplified target sequences by several different methodologies (e.g., quantitation with a bioanalyzer or qPCR, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified target sequence). Any oligonucleotide sequence can be amplified with the appropriate set of primers,

thereby allowing for the amplification of target nucleic acid molecules from genomic DNA, cDNA, formalin-fixed paraffin-embedded DNA, fine-needle biopsies and various other sources. In particular, the amplified target sequences created by the multiplex PCR process as disclosed herein, are themselves efficient substrates for subsequent PCR amplification or various downstream assays or manipulations.

The methods disclosed herein also contemplate any other type of amplification reaction or modified PCR reaction known in the art, which may include, but are not limited to: Allele-specific PCR; Assembly PCR or Polymerase Cycling Assembly (PCA); Digital PCR (dPCR); Helicase-dependent amplification; Hot start PCR; In silico PCR; Intersequence-specific PCR (ISSR); Inverse PCR; Ligation-mediated PCR; Methylation-specific PCR (MSP); Miniprimer PCR; Multiplex Ligation-dependent Probe Amplification (MLPA); Multiplex-PCR; Nanoparticle-Assisted PCR (nanoPCR); Nested PCR; Overlap-extension PCR or Splicing by overlap extension (SOEing); PAN-AC (uses isothermal conditions for amplification and may be used in living cells); Quantitative PCR (qPCR); Reverse Transcription PCR (RT-PCR); Solid Phase PCR; Suicide PCR; Thermal asymmetric interlaced PCR (TAIL-PCR); Touchdown PCR (Step-down PCR); Universal Fast Walking; and the like.

As defined herein, the term “sample” and its derivatives, is used in its broadest sense and includes any specimen, culture and the like that is suspected of including a target. In some embodiments, the sample comprises DNA, RNA, PNA, LNA, chimeric, hybrid, or multiplex-forms of nucleic acids. The sample can include any biological, clinical, surgical, agricultural, atmospheric or aquatic-based specimen containing one or more nucleic acids. The term also includes any isolated nucleic acid sample such as genomic DNA, fresh-frozen or formalin-fixed paraffin-embedded nucleic acid specimen, and the like.

As used herein, “patient” or “subject” can mean either a human or non-human animal, preferably a mammal having a tumor, cancer, or otherwise a proliferative disorder. By “subject” is meant any animal, including horses, dogs, cats, pigs, goats, rabbits, hamsters, monkeys, guinea pigs, rats, mice, lizards, snakes, sheep, cattle, fish, and birds. A human subject may be referred to as a patient. It should be noted that clinical observations described herein were made with human subjects and, in at least some embodiments, the subjects are human.

As used herein, “kits” are understood to contain at least one non-standard laboratory reagent for use in the methods of the disclosure in appropriate packaging, optionally containing

instructions for use. The kit can further include any other components required to practice the method of the disclosure, as dry powders, concentrated solutions, or ready to use solutions. In some embodiments, the kit comprises one or more containers that contain reagents for use in the methods of the disclosure; such containers can be boxes, ampules, bottles, vials, tubes, bags, pouches, blister-packs, or other suitable container forms known in the art. Such containers can be made of plastic, glass, laminated paper, metal foil, or other materials suitable for holding reagents.

Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50, as well as all intervening decimal values between the aforementioned integers such as, for example, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, and 1.9. With respect to sub-ranges, "nested sub-ranges" that extend from either end point of the range are specifically contemplated. For example, a nested sub-range of an exemplary range of 1 to 50 may comprise 1 to 10, 1 to 20, 1 to 30, and 1 to 40 in one direction, or 50 to 40, 50 to 30, 50 to 20, and 50 to 10 in the other direction.

Where applicable or not specifically disclaimed, any one of the embodiments described herein are contemplated to be able to combine with any other one or more embodiments, even though the embodiments are described under different aspects of the disclosure.

These and other embodiments are disclosed and/or encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

The following detailed description, given by way of example, but not intended to limit the disclosure solely to the specific embodiments described, may best be understood in conjunction with the accompanying drawings, in which:

FIG. 1 depicts a paired-guide (pgRNA) library oligonucleotide design and the swapping pair issues that are generated from polymerase chain reaction (PCR). This design includes an oligonucleotide pool that contains a common linker between two guide RNA (gRNA) sequences. During each round of PCR, the non-full length PCR products base pair with each other through

the common linker, which serves as an anchor. The 3'->5' exonuclease activity of the polymerase may digest the unmatched gRNA sequence when two ssDNAs bind to each other through the common linker. After the extension step, recombination may occur between different gRNA pairs, leading to the creation of undesired gRNA pairs.

FIGS. 2A-2F depict the results of two rounds of CRISPR screens on T47D and MCF7 cell lines that revealed that ER-regulated C-Src Tyrosine Kinase (CSK) mediates hormone independent breast cancer cell growth and is synthetic lethal in combination with P21 (RAC1) Activated Kinase 2 (PAK2). FIG. 2A is a schematic that shows the experimental procedure for the first round of CRISPR screening. FIG. 2B is a graph that shows that CSK is positively selected in both T47D and MCF7 cells cultured in hormone depleted medium treated with vehicle conditions compared to Estradiol (E2). FIG. 2C is a graph that shows the frequency change of the CSK-targeting single-guide RNAs (sgRNAs) in both screens. FIG. 2D is a plate staining assay that depicts the effects on cell growth by knocking out CSK using three different gRNAs against CSK, and one gRNA against AAVS1 as a control. CSK function is rescued by the expression of gRNA-resistant CSK cDNAs in these CSK null cells. Cell growth was measured by crystal violet staining assays. FIG. 2E is a schematic that shows the experimental procedures of the second round of CRISPR screening in which T47D cells were first infected with lentiviral gCSK and gAAVS1. After blasticidin selection, T47D cells were generated with stable expression of gCSK and gAAVS1, respectively, and then the genome-wide CRISPR screens were performed in the same manner as the first round. FIG. 2F depicts a Western blot and bar graphs that validate the presence of a synthetic lethal interaction between PAK2 and CSK in T47D cells.

FIGS. 3A-3I depict the pgRNA CRISPR library construction and screening strategy according to an exemplary embodiment of the disclosure. FIG. 3A is a flowchart that depicts a two-step pgRNA cloning strategy. Briefly, a synthesized DNA oligo including the sequences of two gRNAs (represented in red and purple) with an identical linker (grey, in contrast to the unique linkers in the improved oligo design described herein to avoid swapping) was amplified using primers targeting flanking sequence to generate a double-stranded DNA molecule containing 40-80 bp homologies to the U6 promoter and the gRNA scaffold. A Gibson assembly reaction was performed between the amplified fragments and the BsmBI digested gRNA-expressing backbone, and then transformed into competent bacterial cells. This intermediate

construct was then digested by BsmBI and a ligation was performed with BsmBI digested tracrRNA-linker-U6 segment. FIG. 3B shows DNA sequences of the engineered oligo and linker between the two gRNAs of each pair (SEQ ID NO: 29). FIG. 3C shows a schematic of pgRNA cell library construction and screening procedures in which the pgRNA library was delivered into a Cas9-expressing cell line of interest by lentiviral infection with a MOI of about 0.3, and the infected cells were harvested by FACS for green fluorescence 3 days' post-infection. For screening, library cells were cultured for 30 days before genome DNA extraction and high-throughput sequencing analysis of the barcode gRNA regions. FIG. 3D shows an improved pgRNA vector including two gRNA cassettes and a Cas9 expression cassette according to an exemplary embodiment. FIG. 3E shows a method of making the improved pgRNA vector of FIG. 3D. FIG. 3F shows the design of the synthesized oligonucleotide including a first gRNA, a unique linker flanked by to restriction sites, and a second gRNA (SEQ ID NO: 16). FIG. 3G is a schematic showing how the method of FIG. 3E reduces frequencies of recombination/swapping of pgRNAs during library construction. FIG. 3H shows two graphs depicting the read count distribution of correct pgRNAs and swapped/recombined pgRNAs on the pgRNA plasmid library and the read count distribution on Day 0, Cell 1the cell library. FIG. 3I shows the table of colony PCR amplicons and sequencing analysis result.

FIG. 4 depicts a graph showing an exemplary regression residual approach to identify SLGI from a pgRNA screen. The Y-axis represents the logFC of pgRNA targeting a pair of TSG with partner, whereas the X-axis represents the logFC of pgRNA targeting a pair of AAVS1 with the same partner. Ideally, each SLGI of a gene should be supported by multiple pgRNAs. Under certain circumstances, synthetic rescue effect might be observed.

FIGS. 5A-D generally depict library design and gene calling for exemplary CRISPR screens. FIG. 5A is a schematic that shows a sequence logo illustrating the features that contribute to sgRNA efficiency. FIG. 5B includes a gel and a bar graph that shows that indel rates of the sgRNAs are predicted to be inefficient (predicted low) or efficient (predicted high). FIG. 5C is a table that shows an example design matrix of MAGeCK-MLE according to an exemplary embodiment of the disclosure in which 1 indicates the presence of a certain treatment such as, for example, adding a drug or chemical compound, removing a growth factor, etc., in a sample. FIG. 5D is a schematic that shows the initialization and iterative update of the EM model according to the MAGeCK algorithm.

FIG. 6 is a graph that depicts performance of a prediction algorithm with feature selection and a regression residual approach according to the techniques herein. The model was trained on known yeast SLGI pairs and TCGA colon cancer data, and tested on human SLGI pairs from a shRNA screen on HTC116 colon cancer cells. Using the 1204 identified GI pairs as true positives and randomly selected 1000 non-GI pairs as true negatives, the algorithm provides a clear separation of the two (p-value < 2.2e-16).

FIG. 7 is an equation that represents a weighted regression to combine different training datasets for SLGI prediction. For each data set, a weight score may be derived from cross-validation with a R2 metric, where R2 is the coefficient of determination (R^2) in regression. The final coefficient for each SLGI features may be solved through weighted least square method.

FIGS. 8A-C depict generally the characterization of the mechanisms of pan-cancer or cancer-specific SLGIs. FIG. 8A depicts a schematic demonstrating pan-cancer and cancer-specific SLGIs. FIG. 8B is a schematic that shows putative effects of pan-cancer SLGI on downstream gene expression. FIG. 8C is a schematic that shows putative effects of cancer-specific SLGI on cell number and downstream gene expression. In scenario 1, a downstream pathway is regulated similarly between different cancers but differentially required. In scenario 2, a downstream pathway is expressed differentially between cancers, which can be attributable to different expression of regulators.

FIGS. 9A-9B depict schematic overview of using an exemplary pgRNA library of the disclosure to conduct a functional enhancer screen (FIG. 9A) and a schematic of the screening protocol (FIG. 9B).

FIG. 10 shows six two schematics and two graphs providing data about the deletion of a CSK enhancer according to an exemplary embodiment of the disclosure. The upper portion of FIG. 10 presents a schematic that shows the location of one CSK enhancer (left schematic) and a schematic that shows the designed gRNA targeting loci around this enhancer (right schematic). The bottom portion of FIG. 10 shows CSK expression levels upon introduction of different pairs of gRNAs with indicated time of estrogen treatment (0, 1, 4 hours) in T47D (left graph) and MCF7 (right graph) cell lines.

FIG. 11 shows a schematic of the CSK enhancer tilling design in which more than 1,300 pgRNAs (black stick pairs in the second row) were designed in a tilling format to cover the CSK enhancer region with indicated DNaseI-, ER-, FoxA1-, GATA3- binding peaks.

FIG. 12 shows a schematic, a table, and a dot plot describing the analysis of the CSK enhancer tilling according to an exemplary embodiment of the disclosure. The top schematic shows the use of bins to convert overlapping pgRNA target regions into consecutive units on genomic DNA. The bottom left table shows the exemplary relationship between pgRNAs and bins, and the use of bins as genes to run MAGeCK to evaluate the change of each bin, while the bottom right dot plot is the MAGeCK result, showing the p-value distribution of the positively-selected bins.

FIG. 13 shows a schematic of a region with >1,300 pgRNAs and a similar schematic associated with dot plots of data derived from positive and negative selection experiments. The left schematic shows the location of the pgRNA-tilling covered enhancer region and CSK expression cassette, along with indicated DNaseI, ESR1-, FoxA1-, GATA3- and H3K27ac peaks. The right schematic shows the screening results indicating that both the known enhancer (the right arrow) and potential novel enhancers (the left two arrows) were identified.

FIG. 14 is a chart showing the pgRNA selection matrix. Out of a total of 49 possible pairwise gRNA combinations for a given gene pair, each gene has 7 unique CRISPR gRNAs. The indicated 21 combinations are chosen to ensure that each gRNA is used three times.

FIG. 15 is a chart showing quality control of the 15K pgRNA library. Quality control was assessed for both plasmid and cell libraries by paired-end pgRNA sequencing to ensure the coverage and evenness of all designed pgRNAs and to check for swapping/recombination events.

FIG. 16 is a chart showing the MAGeCK/RRA analysis result of the functional positive control SLGI pairs in the CRISPR screen.

FIG. 17A-FIG. 17D are a series of dot plots showing the analysis of the 15K pgRNA library screen. FIG. 17A is a dot plot anchored on RB1. FIG. 17B is a dot plot anchored on PEN. FIG. 17C is a dot plot anchored on NF1. FIG. 17D is a dot plot anchored on CSK.

DETAILED DESCRIPTION OF THE DISCLOSURE

The present disclosure is based, at least in part, on the discovery that paired-guide RNA (pgRNA)/Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) libraries having

reduced or eliminated rates of internal pgRNA swapping/recombination that may be constructed by using vectors that include two guide RNA (gRNA) cassettes, each having a general structure of promoter-gRNA-scaffold that are constructed from a synthesized oligonucleotide having a general structure of gRNA-1 cassette—unique linker—gRNA-2 cassette such that the unique linker is removed from the final vector containing the two gRNA cassettes.

The promoter used in each gRNA cassette may be different, for example, a gRNA-1 cassette may use a human U6 promoter while a paired gRNA-2 cassette may use a mouse U6 promoter. Additionally, the scaffold sequence in each gRNA cassette will typically include a trans-activating crRNA (tracrRNA), which may include sequences in addition to the tracrRNA. Exemplary human and mouse U6 promoter sequences and RNA scaffolds sequences are listed as below:

1. human U6 promoter (SEQ ID NO: 11):

GAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAG
TAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATA
TGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGG
ACGAAACACCG

2. mouse U6 promoter (SEQ ID NO: 12):

GATCCGACGCGCCATCTCTAGGCCCGCGCCGGCCCCCTCGCACGGACTTGTGGGAG
AAGCTCGGCTACTCCCCTGCCCCGGTTAATTTGCATATAATATTTCCTAGTAACTATA
GAGGCTTAATGTGCGATAAAAGACAGATAATCTGTTCTTTTTAATACTAGCTACATT
TTACATGATAGGCTTGGATTTCTATAACTTCGTATAGCATACATTATACGAAGTTATA
AACAGCACAAAAGGAAACTCACCCTAACTGTAAAGTAATTGTGTGTTTTGAGACTAT
AAGTATCCCTTGGAGAACCACCTTGTTG

3. 1st gRNA scaffold in the hU6 cassette (SEQ ID NO: 13):

GTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

4. 2nd gRNA scaffold in the mU6 cassette (SEQ ID NO: 14):

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTT

5. An exemplary vector may include (SEQ ID NO: 15):
TATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAAT
TAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTGACGTAGAA
AGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCA
TATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAA
GGACGAAACACCG**CCTCCCGCTCCTGGAGCGG**
[gRNA1 in bold]
GTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTCGAGTACTAGGATCCATTA
GGCGGCCGCGTCGACAAGCTTTCTAGAGAATTCGATCCGACGCGCCATCTCTAGGCC
CGCGCCGGCCCCCTCGCACGGACTTGTGGGAGAAGCTCGGCTACTCCCCTGCCCCGG
TTAATTTGCATATAATATTTCCTAGTAACTATAGAGGCTTAATGTGCGATAAAAGAC
AGATAATCTGTTCTTTTTAATACTAGCTACATTTTACATGATAGGCTTGGATTTCTAT
AACTTCGTATAGCATACATTATACGAAGTTATAAACAGCACAAAAGGAAACTCACC
CTAACTGTAAAGTAATTGTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCT
TGTTG**GATATTCACCATTATAGGT**
[gRNA2 in bold]
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGCTTTTTTGAATTCTAGACTTGATGCTAACTAGGTCTTGA
AAGGAGTGGGAATTGGCTCCGGTGCCCGTCAGT.

In an exemplary embodiment, the vectors described herein may include portions of the lentiCRISPRv2 vector (e.g., the World Wide Web at (www) addgene.org/52961/).

The present disclosure provides compositions and methods for making and decoding pgRNA libraries using the CRISPR system. Advantageously, the pgRNA/CRISPR libraries disclosed herein may be used to identify synthetic lethal genetic interactions (SLGI) and non-coding functional elements or cis-elements. The techniques provided herein are important because identifying and characterizing SLGI that occur in combination with cancer causing genes (e.g., tumor suppressor genes) may provide novel therapies with which to treat cancer. In

this regard, the techniques herein provide experimental and computational methods for the large-scale identification of novel therapies to treat cancers with tumor suppressor loss.

Overview

Cancer may be driven by the activation of oncogenes or the deactivation of tumor suppressor genes (TSGs). For example, cancer may be cause by gain-of-function mutations in oncogenes and loss-of-function mutations in TSGs. While activating oncogenic mutations may often be targeted directly by therapeutic intervention, successfully restoring the function of a TSG has thus far not been possible in the clinic. While activating oncogenic mutations may often be directly targeted by therapeutic intervention, successful treatment for tumor suppressor loss has thus far been challenging in the clinic.

Genetic interaction is a phenomenon in which the phenotype of mutations in two genes differs significantly from each mutation's individual effects. In extreme cases, genetic interaction may give rise to synthetic lethality when inactivation of two nonessential genes results in a lethal phenotype. Such synthetic lethal genetic interactions (SLGI) may provide insights on novel cancer therapeutic targets or target combinations that may enhance the efficacy and specificity of targeted drugs. Over the past few years, there have been tremendous efforts to identify SLGI genes in the cancer genome with the primary aim of identifying novel therapeutic targets among the synthetic lethal partners of dysfunctional TSGs. Unfortunately, the accuracy and cost effectiveness of prior art techniques for identifying and validating SLGI pairs in mammalian systems is not sufficient to allow identification of SLGIs at scale.

Historically, many genomic technologies have been developed to map SLGIs in model organisms and humans. For example, two projects of genome-wide quantitative mapping of synthetic lethal interactions have been conducted in yeast based on gene deletion strains (see e.g., references 1 and 2). Based on the same technology, another study screened potential interactions among orthologs of human TSGs and genes encoding drug targets in yeast (see e.g., reference 3). SLGI mapping by directed gene disruptions in human cell lines is very important, as SLGIs involving TSGs or oncogenes may provide insights to precision cancer medicine. Such disruptions generally use RNA interference (e.g., siRNA or shRNA) knockdown or CRISPR/Cas9 knockout and can be roughly categorized as either a "1 x n" design or an "a x b" design.

In a "1 x n" design, genome-wide (n genes) RNAi or CRISPR screens may be used to identify genes showing differential essentiality between cell lines where an anchor gene (1 gene) is active vs inactive. Here, the anchor gene may be inactivated by RNAi or CRISPR (see e.g., references 4–6), drug inhibition (see e.g., reference 7), or inherently lost in the cell line (see e.g., reference 8).

In an "a x b" design, all pairwise combinations of shRNAs or CRISPR guide RNAs (gRNAs) within a starting pool may be randomly combined together to test possible interactions among sets of genes. Although "a" and "b" could theoretically be different, so far published screens are mostly in an "a x a" design, such as 190 x 190 shRNA pairs (see e.g., reference 9) or 153 x 153 CRISPR gRNA pairs (see e.g., reference 10). SLGI screens on specific pairs using simultaneous delivery of two shRNAs have been proposed by The DECIPHER project, although to date no studies have been published using this technique and shRNA is unfortunately known to have significant off-target effect. The "a x b" design may also be carried out in arrayed format with automated technologies (see e.g., reference 11) instead of pooled screens. However, such combinatorial design falls short of the required throughput to interrogate the potential interaction space of all the possible SLGIs involving TSGs.

Many computational approaches have also been developed to systematically study genetic interactions for yeast where genome-wide experimental maps of SLGIs are available (see e.g., references 12–15). In cancer, SLGI has been computationally predicted through mapping yeast genetic interactions to their human orthologs (see e.g., reference 16) and utilizing metabolic models and evolutionary characteristics of metabolic genes (see e.g., references 17–19). With the rapidly accumulating cancer genomic data, a data-driven method, named DAISY, was used to integrate somatic copy number alterations, shRNA-based essentiality screens, and co-expression patterns on hundreds of cancer cell lines to detect SLGI pairs in human (see e.g., reference 20). However, the filtering criterion integrating each data type is determined in an ad hoc manner, and the experimental validation was only conducted on a handful of interactions (see e.g., references 20). Despite these efforts, SLGI identification and validation has been limited by the scale and accuracy of the prior art experimental technology; therefore, it is very difficult to systematically evaluate the performance of SLGI computational predictions with prior art methodologies.

Despite the success of targeted therapies treating cancers with activating mutations, prior art attempts to therapeutically target cancers with TSGs loss (e.g., Tumor Protein P53 (TP53), Phosphatase and Tensin Homolog (PTEN), and the like) have not been effective. SLGI could provide novel insights on therapeutic targets to treat cancers with TSG loss. There have been tremendous efforts to identify SLGI among genes in the genome; unfortunately, the accuracy and cost effectiveness of these efforts in mammalian systems was not sufficient to allow SLGI screening at scale. The CRISPR/Cas9 genome editing technology and CRISPR/Cas9 knockout (KO) screens offers exciting new opportunities to investigate SLGI in mammalian genomes.

CRISPR

The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system is a revolutionary approach for genome editing and functional genomics research in mammalian systems. Cas9 nucleases are directed to specific genomic loci by single-guide RNAs (sgRNAs) containing 19-20 nucleotides that are complementary to the target DNA sequences, thereby creating frameshift insertion/deletion (indel) mutations that result in a loss-of-function allele. The development of lentiviral delivery of a genome-scale CRISPR/Cas9 knockout (KO) library targeting all genes enables both negative and positive selection screening on mammalian cell lines in a cost-effective manner. In genome-wide sgRNA/CRISPR screens, each gene may be targeted by several sgRNAs for KO, and the mutant pool carrying different gene KOs can then be resolved by high throughput sequencing. Those sgRNA targeting genes that inhibit growth under the screening conditions will be enriched while those targeting essential genes will be under-represented. Thus, CRISPR screening is a powerful technology for systematic genetic analysis, and is especially relevant in cancer where growth under various conditions or under drug selection is a critical phenotype.

The delivery of the two sgRNAs into a single cell could create mutations at both targeting locus simultaneously or fragment deletions if two cutting sites are close to each other. Therefore, to build the CRISPR library in which each vector express two gRNAs provides a new approach to investigate gene interactions and functional non-coding elements in a systematic way. Although sgRNA/CRISPR libraries have been constructed and used for screening, prior art libraries suffer from the significant disadvantage that they are prone to recombination that creates undesirable sgRNA pairs and are therefore not amenable to scaling. Accordingly, there remains an urgent unmet need for the construction of high-quality, recombination-free

pgRNA/CRISPR libraries that allow for reliable, scalable functional genomics studies to identify synthetic lethal gene interactions and non-coding elements.

It is contemplated within the scope of the disclosure, that the CRISPR/Cas system may be used to modify any of the nucleotides described herein, either for in vitro or in vivo manipulation of the nucleotides, or for identification of genetic interactions (e.g., SLGIs). For example, the techniques herein provide that the CRISPR/Cas system may be used therapeutically to down regulate expression of, or knockout, pairs of genes in a cancer cell(s). The CRISPR/Cas system is abundantly described in US Patent No. 8,795,965, US Patent No. 8,889,356, US Patent No. 8,771,945, US Patent No. 8,889,418, and US Patent No. 8,895,308, which are hereby incorporated by reference in their entirety.

Briefly, the term "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a trans-activating CRISPR (tracr) sequence (e.g. tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (e.g., guide RNA), or other sequences and transcripts from a CRISPR locus. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III CRISPR system. In some embodiments, one or more elements of a CRISPR system may be derived from a particular organism comprising an endogenous CRISPR system, such as Streptococcus pyogenes. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (e.g., a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence (e.g., gRNA) is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast.

The clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system is a revolutionary approach for genome editing of mammalian systems. Cas9 nucleases are directed to specific genomic loci by single-guide RNAs (sgRNAs) containing 17-27 nucleotides that are complementary to the target DNA sequences and have the ability to create frameshift insertion/deletion (indel) mutations that result in a loss-of-function allele. In an exemplary embodiment, the sgRNAs may be 19-20 nucleotides in length. In an exemplary embodiment, the sgRNAs may be 19 nucleotides in length. Recently, the development of lentiviral delivery of a genome-scale CRISPR/Cas9 knockout (KO) library targeting all genes enables both negative and positive selection screening on mammalian cell lines in a cost-effective manner (see e.g., references 7, 21, and 22). In genome-wide CRISPR KO screens, each gene is targeted by several sgRNAs for KO, and the mutant pool carrying different gene KOs can then be resolved by high throughput sequencing. Those sgRNA targeting genes that inhibit growth under the screening conditions will be enriched while those targeting essential genes will be under-represented. Thus, CRISPR screening is a powerful technology for systematic genetic analysis, and is especially relevant in cancer where growth under various conditions or under drug selection is a critical phenotype.

Over the last few years, techniques have been developed to provide CRISPR screens using paired guide RNAs (pgRNAs) to create deletions in, or to silence, two different genes simultaneously. For example, studies have found that pgRNAs driven by separate U6 promoters work better than consecutive gRNAs transcribed from the same U6 (see e.g., references 23-25). To prevent the two gRNA expression cassettes (U6-gRNA-tracrRNA) from swapping (e.g., recombining) during lentiviral replication, U6 promoters from different species and different tracr RNA sequences for the two gRNAs may be used (see e.g., references 25). This approach also enables the pgRNAs to be read from paired-end sequencing. Unfortunately, pilot studies have indicated that the pgRNAs may still swap or recombine at two different stages during the pooled screen. First, when the synthesized long oligonucleotides carrying the two gRNAs are PCR amplified to construct the custom CRISPR library, the two gRNAs may swap or recombine during PCR due to the common restriction enzyme recognition sites and linker sequence that are shared between the two gRNAs (see e.g., FIG. 1). Second, during the final PCR step to prepare the sequencing library, the two gRNAs may swap or recombine again during PCR due to the first tracrRNA and second U6 sequences that are shared in common between the two gRNAs.

Additionally, the polymerase used in current PCR reactions has a 3' to 5' exonuclease activity that exacerbates the frequency of swapping or recombining during the PCR process (see e.g., FIG. 1). For example, long non-coding RNA (lncRNA) deletion CRISPR screens used 25 pgRNAs to delete the promoter of each lncRNA; however, this deletion screen still suffered from a high false negative rate due to recombination between pgRNAs during PCR (see e.g., reference 23). The techniques herein provide the ability to finally resolve the PCR swapping/recombination issues inherent in the prior art (see e.g., reference 23) and provide a pooled pgRNA CRISPR screening methodology that is robust, effective, accurate, and scalable.

By "Tumor Protein P53 (TP53) nucleic acid molecule" is meant a polynucleotide encoding a TP53 polypeptide. An exemplary TP53 nucleic acid molecule is provided at NCBI Accession No. NM_000546, version NM_000546.5, incorporated herein by reference, and reproduced below (SEQ ID NO: 1):

```
   1 gatgggattg gggttttccc ctcccatgtg ctcaagactg gcgctaaaag ttttgagctt
  61 ctcaaaagtc tagagccacc gtccagggag caggtagctg ctgggctccg gggacacttt
 121 gcgttcgggc tgggagcgtg ctttccacga cggtgacacg cttccctgga ttggcagcca
 181 gactgccttc cgggtcactg ccatggagga gccgcagtca gatcctagcg tcgagccccc
 241 tctgagtcag gaaacatttt cagacctatg gaaactactt cctgaaaaca acgttctgtc
 301 cccttgccg tcccaagcaa tggatgattt gatgctgtcc ccggacgata ttgaacaatg
 361 gttcactgaa gacccaggtc cagatgaagc tcccagaatg ccagaggctg ctccccccgt
 421 ggcccctgca ccagcagctc ctacaccggc ggcccctgca ccagcccct cctggcccct
 481 gtcatcttct gtcccttccc agaaaaccta ccagggcagc tacggtttcc gtctgggctt
 541 cttgcattct gggacagcca agtctgtgac ttgcacgtac tcccctgccc tcaacaagat
 601 gttttgccaa ctggccaaga cctgccctgt gcagctgtgg gttgattcca cacccccgcc
 661 cggcacccgc gtccgcgcca tggccatcta caagcagtca cagcacatga cggaggttgt
 721 gaggcgctgc ccccaccatg agcgctgctc agatagcgat ggtctggccc ctcctcagca
 781 tcttatccga gtggaaggaa atttgcgtgt ggagtatttg gatgacagaa acacttttcg
 841 acatagtgtg gtggtgccct atgagccgcc tgaggttggc tctgactgta ccaccatcca
 901 ctacaactac atgtgtaaca gttcctgcat gggcggcatg aaccggaggc ccatcctcac
 961 catcatcaca ctggaagact ccagtggtaa tctactggga cggaacagct ttgaggtgcg
1021 tgtttgtgcc tgtcctggga gagaccggcg cacagaggaa gagaatctcc gcaagaaagg
1081 ggagcctcac cacgagctgc ccccagggag cactaagcga gcactgccca acaacaccag
1141 ctcctctccc cagccaaaga agaaaccact ggatggagaa tatttcaccc ttcagatccg
1201 tgggcgtgag cgcttcgaga tgttccgaga gctgaatgag gccttggaac tcaaggatgc
1261 ccaggctggg aaggagccag gggggagcag ggctcactcc agccacctga agtccaaaaa
1321 gggtcagtct acctcccgcc ataaaaaact catgttcaag acagaagggc ctgactcaga
1381 ctgacattct ccacttcttg ttccccactg acagcctccc acccccatct ctccctcccc
1441 tgccattttg ggttttgggt ctttgaaccc ttgcttgcaa taggtgtgcg tcagaagcac
1501 ccaggacttc catttgcttt gtcccggggc tccactgaac aagttggcct gcactggtgt
1561 tttgttgtgg ggaggaggat ggggagtagg acataccagc ttagatttta aggtttttac
1621 tgtgagggat gtttgggaga tgtaagaaat gttcttgcag ttaagggtta gtttacaatc
1681 agccacattc taggtagggg cccacttcac cgtactaacc agggaagctg tccctcactg
1741 ttgaattttc tctaacttca aggcccatat ctgtgaaatg ctggcatttg cacctacctc
1801 acagagtgca ttgtgagggt taatgaaata atgtacatct ggccttgaaa ccacctttta
1861 ttacatgggg tctagaactt gacccccttg agggtgcttg ttccctctcc ctgttggtcg
1921 gtgggttggt agtttctaca gttgggcagc tggttaggta gagggagttg tcaagtctct
1981 gctggcccag ccaaaccctg tctgacaacc tcttggtgaa ccttagtacc taaaaggaaa
```

```
2041 tctcacccca tcccacaccc tggaggattt catctcttgt atatgatgat ctggatccac
2101 caagacttgt tttatgctca gggtcaattt ctttttttctt ttttttttttt ttttttcttt
2161 ttctttgaga ctgggtctcg ctttgttgcc caggctggag tggagtggcg tgatcttggc
2221 ttactgcagc ctttgcctcc ccggctcgag cagtcctgcc tcagcctccg gagtagctgg
2281 gaccacaggt tcatgccacc atggccagcc aacttttgca tgttttgtag agatggggtc
2341 tcacagtgtt gcccaggctg gtctcaaact cctgggctca ggcgatccac ctgtctcagc
2401 ctcccagagt gctgggatta caattgtgag ccaccacgtc cagctggaag ggtcaacatc
2461 ttttacattc tgcaagcaca tctgcatttt caccccaccc ttccccctcct tctccctttt
2521 tatatcccat ttttatatcg atctcttatt ttacaataaa actttgctgc cacctgtgtg
2581 tctgaggggt g
```

By "Tumor Protein P53 (TP53) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_000537, version NP_000537.3, incorporated herein by reference, as reproduced below (SEQ ID NO: 2):

```
  1 meepqsdpsv epplsqetfs dlwkllpenn vlsplpsqam ddlmlspddi eqwftedpgp
 61 deaprmpeaa ppvapapaap tpaapapaps wplsssvpsq ktyqgsygfr lgflhsgtak
121 svtctyspal nkmfcqlakt cpvqlwvdst pppgtrvram aiykqsqhmt evvrrcphhe
181 rcsdsdglap pqhlirvegn lrveylddrn tfrhsvvvpy eppevgsdct tihynymcns
241 scmggmnrrp iltiitleds sgnllgrnsf evrvcacpgr drrteeenlr kkgephhelp
301 pgstkralpn ntssspqpkk kpldgeyftl qirgrerfem frelnealel kdaqagkepg
361 gsrahsshlk skkgqstsrh kklmfktegp dsd
```

By "Phosphatase and Tensin Homolog (PTEN) nucleic acid molecule" is meant a polynucleotide encoding a PTEN polypeptide. An exemplary PTEN nucleic acid molecule is provided at NCBI Accession No. NM_000314, version NM_000314.6, incorporated herein by reference, and reproduced below (SEQ ID NO: 3):

```
  1 cctcccctcg cccggcgcgg tcccgtccgc ctctcgctcg cctcccgcct cccctcggtc
 61 ttccgaggcg cccgggctcc cggcgcggcg gcggaggggg cgggcaggcc ggcgggcggt
121 gatgtggcgg gactctttat gcgctgcggc aggatacgcg ctcggcgctg ggacgcgact
181 gcgctcagtt ctctcctctc ggaagctgca gccatgatgg aagtttgaga gttgagccgc
241 tgtgaggcga ggccgggctc aggcgaggga gatgagagac ggcggcggcc gcggcccgga
301 gcccctctca gcgcctgtga gcagccgcgg gggcagcgcc ctcggggagc cggccggcct
361 gcggcggcgg cagcggcggc gtttctcgcc tcctcttcgt cttttctaac cgtgcagcct
421 cttcctcggc ttctcctgaa agggaaggtg gaagccgtgg gctcgggcgg gagccggctg
481 aggcgcggcg gcggcggcgg cacctcccgc tcctggagcg gggggggagaa gcggcggcgg
541 cggcggccgc ggcggctgca gctccaggga gggggtctga gtcgcctgtc accatttcca
601 gggctgggaa cgccggagag ttggtctctc cccttctact gcctccaaca cggcggcggc
661 ggcggcggca catccaggga cccgggccgg ttttaaacct cccgtccgcc gccgccgcac
721 cccccgtggc ccgggctccg gaggccgccg gcggaggcag ccgttcggag gattattcgt
781 cttctcccca ttccgctgcc gccgctgcca ggcctctggc tgctgaggag aagcaggccc
841 agtcgctgca accatccagc agccgccgca gcagccatta cccggctgcg gtccagagcc
901 aagcggcggc agagcgaggg gcatcagcta ccgccaagtc cagagccatt tccatcctgc
961 agaagaagcc ccgccaccag cagcttctgc catctctctc ctccttttc ttcagccaca
1021 ggctcccaga catgacagcc atcatcaaag agatcgttag cagaaacaaa aggagatatc
1081 aagaggatgg attcgactta gacttgacct atatttatcc aaacattatt gctatgggat
1141 ttcctgcaga aagacttgaa ggcgtataca ggaacaatat tgatgatgta gtaaggtttt
```

1201 tggattcaaa gcataaaaac cattacaaga tatacaatct ttgtgctgaa agacattatg
1261 acaccgccaa atttaattgc agagttgcac aatatccttt tgaagaccat aacccaccac
1321 agctagaact tatcaaaccc ttttgtgaag atcttgacca atggctaagt gaagatgaca
1381 atcatgttgc agcaattcac tgtaaagctg gaaagggacg aactggtgta atgatatgtg
1441 catatttatt acatcggggc aaatttttaa aggcacaaga ggccctagat ttctatgggg
1501 aagtaaggac cagagacaaa aagggagtaa ctattcccag tcagaggcgc tatgtgtatt
1561 attatagcta cctgttaaag aatcatctgg attatagacc agtggcactg ttgtttcaca
1621 agatgatgtt tgaaactatt ccaatgttca gtggcggaac ttgcaatcct cagtttgtgg
1681 tctgccagct aaaggtgaag atatattcct ccaattcagg acccacacga cgggaagaca
1741 agttcatgta ctttgagttc cctcagccgt tacctgtgtg tggtgatatc aaagtagagt
1801 tcttccacaa acagaacaag atgctaaaaa aggacaaaat gtttcacttt tgggtaaata
1861 cattcttcat accaggacca gaggaaacct cagaaaaagt agaaaatgga agtctatgtg
1921 atcaagaaat cgatagcatt tgcagtatag agcgtgcaga taatgacaag gaatatctag
1981 tacttacttt aacaaaaaat gatcttgaca aagcaaataa agacaaagcc aaccgatact
2041 tttctccaaa ttttaaggtg aagctgtact tcacaaaaac agtagaggag ccgtcaaatc
2101 cagaggctag cagttcaact tctgtaacac cagatgttag tgacaatgaa cctgatcatt
2161 atagatattc tgacaccact gactctgatc cagagaatga acctttttgat gaagatcagc
2221 atacacaaat tacaaaagtc tgaatttttt tttatcaaga gggataaaac accatgaaaa
2281 taaacttgaa taaactgaaa atggaccttt ttttttttaa tggcaatagg acattgtgtc
2341 agattaccag ttataggaac aattctcttt tcctgaccaa tcttgtttta ccctatacat
2401 ccacagggtt ttgacacttg ttgtccagtt gaaaaaaggt tgtgtagctg tgtcatgtat
2461 atacctttttt gtgtcaaaag gacatttaaa attcaattag gattaataaa gatggcactt
2521 tcccgtttta ttccagtttt ataaaaagtg gagacagact gatgtgtata cgtaggaatt
2581 ttttcctttt gtgttctgtc accaactgaa gtggctaaag agctttgtga tatactggtt
2641 cacatcctac ccctttgcac ttgtggcaac agataagttt gcagttggct aagagaggtt
2701 tccgaagggt tttgctacat tctaatgcat gtattcgggt taggggaatg gagggaatgc
2761 tcagaaagga aataatttta tgctggactc tggaccatat accatctcca gctatttaca
2821 cacacctttc tttagcatgc tacagttatt aatctggaca ttcgaggaat tggccgctgt
2881 cactgcttgt tgtttgcgca tttttttttа aagcatattg gtgctagaaa aggcagctaa
2941 aggaagtgaa tctgtattgg ggtacaggaa tgaaccttct gcaacatctt aagatccaca
3001 aatgaaggga tataaaaata atgtcatagg taagaaacac agcaacaatg acttaaccat
3061 ataaatgtgg aggctatcaa caaagaatgg gcttgaaaca ttataaaaat tgacaatgat
3121 ttattaaata tgttttctca attgtaacga cttctccatc tcctgtgtaa tcaaggccag
3181 tgctaaaatt cagatgctgt tagtacctac atcagtcaac aacttacact tattttacta
3241 gttttcaatc ataatacctg ctgtggatgc ttcatgtgct gcctgcaagc ttctttttтc
3301 tcattaaata taaaatattt tgtaatgctg cacagaaatt ttcaatttga gattctacag
3361 taagcgtttt ttttctttga agatttatga tgcacttatt caatagctgt cagccgttcc
3421 acccttttga ccttacacat tctattacaa tgaattttgc agttttgcac atttttтaaa
3481 tgtcattaac tgttagggaa ttttacttga atactgaata catataatgt ttatattaaa
3541 aaggacattt gtgttaaaaa ggaaattaga gttgcagtaa actttcaatg ctgcacacaa
3601 aaaaaagaca tttgattttt cagtagaaat tgtcctacat gtgctttatt gatttgctat
3661 tgaaagaata gggtttttтt ttttttttтt ttttttttтt ttaaatgtgc agtgttgaat
3721 catttcttca tagtgctccc ccgagttggg actagggctt caatttcact tcttaaaaaa
3781 aatcatcata tatttgatat gcccagactg catacgattt taagcggagt acaactacta
3841 ttgtaaagct aatgtgaaga tattattaaa aaggtttttt tttccagaaa tttggtgtct
3901 tcaaattata ccttcacctt gacatttgaa tatccagcca ttttgtttct taatggtata
3961 aaattccatt ttcaataact tattggtgct gaaattgttc actagctgtg gtctgaccta
4021 gttaatttac aaatacagat tgaataggac ctactagagc agcatttata gagtttgatg
4081 gcaaatagat taggcagaac ttcatctaaa atattcttag taaataatgt tgacacgttt
4141 tccatacctt gtcagtttca ttcaacaatt tttaaatttt taacaaagct cttaggattt
4201 acacatttat atttaaacat tgatatatag agtattgatt gattgctcat aagttaaatt
4261 ggtaaagtta gagacaacta ttctaacacc tcaccattga aatttatatg ccaccttgtc
4321 tttcataaaa gctgaaaatt gttacctaaa atgaaaatca acttcatgtt ttgaagatag
4381 ttataaatat tgttctttgt tacaatttcg ggcaccgcat attaaaacgt aactttattg
4441 ttccaatatg taacatggag ggccaggtca taaataatga cattataatg ggcttttgca
4501 ctgttattat ttttcctttg gaatgtgaag gtctgaatga gggttttgat tttgaatgtt

4561 tcaatgtttt tgagaagcct tgcttacatt ttatggtgta gtcattggaa atggaaaaat
4621 ggcattatat atattatata tataaatata tattatacat actctcctta ctttatttca
4681 gttaccatcc ccatagaatt tgacaagaat tgctatgact gaaaggtttt cgagtcctaa
4741 ttaaaacttt atttatggca gtattcataa ttagcctgaa atgcattctg taggtaatct
4801 ctgagtttct ggaatatttt cttagacttt ttggatgtgc agcagcttac atgtctgaag
4861 ttacttgaag gcatcacttt taagaaagct tacagttggg ccctgtacca tcccaagtcc
4921 tttgtagctc ctcttgaaca tgtttgccat acttttaaaa gggtagttga ataaatagca
4981 tcaccattct ttgctgtggc acaggttata aacttaagtg gagtttaccg gcagcatcaa
5041 atgtttcagc tttaaaaaat aaaagtaggg tacaagttta atgtttagtt ctagaaattt
5101 tgtgcaatat gttcataacg atggctgtgg ttgccacaaa gtgcctcgtt tacctttaaa
5161 tactgttaat gtgtcatgca tgcagatgga aggggtggaa ctgtgcacta aagtgggggc
5221 tttaactgta gtatttggca gagttgcctt ctacctgcca gttcaaaagt tcaacctgtt
5281 ttcatataga atatatatac taaaaaattt cagtctgtta aacagcctta ctctgattca
5341 gcctcttcag atactcttgt gctgtgcagc agtggctctg tgtgtaaatg ctatgcactg
5401 aggatacaca aaaataccaa tatgatgtgt acaggataat gcctcatccc aatcagatgt
5461 ccatttgtta ttgtgtttgt taacaaccct ttatctctta gtgttataaa ctccacttaa
5521 aactgattaa agtctcattc ttgtcattgt gtgggtgttt tattaaatga gagtttataa
5581 ttcaaattgc ttaagtccat tgaagtttta attaatgggc agccaaatgt gaatacaaag
5641 ttttcagttt ttttttttcc tgctgtcctt caaagcctac tgtttaaaaa aaaaaaaaaa
5701 aaaaaacatg gcctgagagt agagtatctg tctactcatg tttaattaag gaaaaacact
5761 tatttttagg gctttagtca tcacttcata aattgtataa gcacattaaa tagcgttcta
5821 gtcctgaaaa agtccaagat tcttagaaaa ttgtgcatat ttttattatg acagatgttt
5881 gaagataatt ccccagaatg gatttgatac tttagatttc aattttgtgg cttttgtcta
5941 ttattctgta ctctgccatc agcatatgga aagcttcatt tactcatcat gacttgtgcc
6001 atataaaaat tgatatttcg gaatagtcta aaggactttt tgtacttgaa tttaatcatg
6061 ttgtttctaa tattcttaaa agcttgaaga ctaaagcata tcctttcaac aaagcatagt
6121 aaggtaataa gaaagtgtag tttgtacaag tgttaaaaaa ataaagtaga caatgttaca
6181 gtgggactta ttatttcaag tttacatttt ctccatgtaa ttttttaaaa agtaaatgaa
6241 aaaatgtgca ataatgtaaa atatgaagtg tatgtgtaca cacattttat ttttcggtat
6301 cttgggtata cgtatggttg aaaactatac tggagtctaa aagtattcta atttataaga
6361 agacattttg gtgatgtttg aaaaatagaa atgtgctagt tttgttttta tatcatgtcc
6421 tttgtacgtt gtaatatgag ctggcttggt tcagtaaatg ccatcaccat ttccattgag
6481 aatttaaaac tcaccagtgt ttaatatgca ggcttccaaa ggcttatgaa aaaaatcaag
6541 acccttaaat ctagttaatt tgctgctaac atgaaactct ttggttcttt tatttttgcc
6601 agataattag acacacatct aaagcttagt cttaaatggc ttaagtgtag ctattgatta
6661 gtgctgttgc tagttcagaa agaaatgttt gtgaatggaa acaagaatat tcagtccaaa
6721 ctgttgtaag gacagtacct gaaaaccagg aaacaggata atggaaaaag tcttttaaag
6781 atgaaatgtt ggagccaact ttcttataga attaattgta tgtggctata gaaagcctaa
6841 tgattgttgc ttatttttga gagcatatta ttcttttatg accataatct tgctgttttt
6901 ccatcttcca aaagatcttc cttctaatat gtatatcaga atgtgggtag ccagtcagac
6961 aaattcatat tggttggtag ctttaaaaag tttgtaatgt gaagacagga aaggacaaaa
7021 tagtttgctt tggtggtagt actctggttg ttaagctagg tattttgaga ctacttcccc
7081 atcacaacaa caataaaata atcactcata atcctatcac ctggagacat agccatcgtt
7141 aatatgttag tgactataca atcatgtttt cttctgtata tccatgtata ttctttaaaa
7201 atgaaattta tactgtacct gatctcaaag ctttttagct tagtatatct gtcatgaatt
7261 tgtaggatgt tccattgcat cagaaaacgg acagtgattt gattactttc taatgccaca
7321 gatgcagatt acatgtagtt attgagaatc ctttcgaatt cagtggctta atcatgaatg
7381 tctaaatatt gttgacatta ggatgataca tgtaaattaa agttacattt gtttagcata
7441 gacaagctta acattgtaga tgtttctctt caaaaatcat cttaaacatt tgcatttgga
7501 attgtgttaa atagaatgtg tgaaacactg tattagtaaa cttcatcacc tttctacttc
7561 cttatagttt gaacttttca gtttttgtag ttcccaaaca gttgctcaat ttagagcaaa
7621 ttaatttaac acctgccaaa aaaaggctgc tgttggctta tcagttgtct ttaaattcaa
7681 atgctcatgt gacttttatc acatcaaaaa atatttcatt aatgattcac ctttagctct
7741 gaaaattacc gcgtttagta attatagtgg gcttataaaa acatgcaact ctttttgata
7801 gttatttgag aattttggtg aaaaatattt agctgagggc agtatagaac ttataaacca
7861 atatattgat atttttaaaa catttttaca tataagtaaa ctgccatctt tgagcataac

```
7921 tacatttaaa aataaagctg catattttta aatcaagtgt ttaacaagaa tttatatttt
7981 ttatttttta aaattaaaaa taatttatat ttcctctgtt gcatgaggat tctcatctgt
8041 gcttataatg gttagagatt ttatttgtgt ggaatgaagt gaggcttgta gtcatggttc
8101 tagtgtttca gtttgccaag tctgtttact gcagtgaaat tcatcaaatg tttcagtgtg
8161 gttttctgta gcctatcatt tactggctat ttttttatgt acacctttag gattttctgc
8221 ctactctatc cagttgtcca aatgatatcc tacattttac aaatgccctt tcagtttcta
8281 ttttcttttt ccattaaatt gccctcatgt cctaatgtgc agtttgtaag tgtgtgtgtg
8341 tgtgtctgtg tgtgtgtgaa tttgattttc aagagtgcta gacttccaat ttgagagatt
8401 aaataattta attcaggcaa acatttttca ttggaatttc acagttcatt gtaatgaaaa
8461 tgttaatcct ggatgacctt tgacatacag taatgaatct tggatattaa tgaatttgtt
8521 agtagcatct tgatgtgtgt tttaatgagt tattttcaaa gttgtgcatt aaaccaaagt
8581 tggcatactg gaagtgttta tatcaagttc catttggcta ctgatggaca aaaaatagaa
8641 atgccttcct atggagagta tttttccttt aaaaaattaa aaaggttaat tattttgact
8701 aaaaaaaaaa aaaaaaaa
```

By "Phosphatase and Tensin Homolog (PTEN) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_000305, version NP_000305.3, incorporated herein by reference, as reproduced below (SEQ ID NO: 4):

```
  1 mtaiikeivs rnkrryqedg fdldltyiyp niiamgfpae rlegvyrnni ddvvrfldsk
 61 hknhykiynl caerhydtak fncrvaqypf edhnppqlel ikpfcedldq wlseddnhva
121 aihckagkgr tgvmicayll hrgkflkaqe aldfygevrt rdkkgvtips qrryvyyysy
181 llknhldyrp vallfhkmmf etipmfsggt cnpqfvvcql kvkiyssnsg ptrredkfmy
241 fefpqplpvc gdikveffhk qnkmlkkdkm fhfwvntffi pgpeetsekv engslcdqei
301 dsicsierad ndkeylvltl tkndldkank dkanryfspn fkvklyftkt veepsnpeas
361 sstsvtpdvs dnepdhyrys dttdsdpene pfdedqhtqi tkv
```

By "Tuberous Sclerosis 1 (TSC1) nucleic acid molecule" is meant a polynucleotide encoding a TSC1 polypeptide. An exemplary TSC1 nucleic acid molecule is provided at NCBI Accession No. NM_000368, version NM_000368.4, incorporated herein by reference, and reproduced below (SEQ ID NO: 5):

```
  1 acgacggggg aggtgctgta cgtccaagat ggcggcgccc tgtaggctgg agggactgtg
 61 aggtaaacag ctgaggggga ggagacggtg gtgaccatga aagacaccag gttgacagca
121 ctggaaactg aagtaccagt tgtcgctaga acagtttggt agtggcccca atgaagaacc
181 ttcagaacct gtagcacacg tcctggagcc agcacagcgc cttcgagcga gagaatggcc
241 caacaagcaa atgtcgggga gcttcttgcc atgctggact ccccccatgct gggtgtgcgg
301 gacgacgtga cagctgtctt taaagagaac ctcaattctg accgtggccc tatgcttgta
361 aacaccttgg tggattatta cctggaaacc agctctcagc cggcattgca catcctgacc
421 accttgcaag agccacatga caagcacctc ttggacagga ttaacgaata tgtgggcaaa
481 gccgccactc gtttatccat cctctcgtta ctgggtcatg tcataagact gcagccatct
541 tggaagcata agctctctca agcacctctt ttgccttctt tactaaaatg tctcaagatg
601 gacactgacg tcgttgtcct cacaacaggc gtcttggtgt tgataaccat gctaccaatg
661 attccacagt ctgggaaaca gcatcttctt gatttctttg acatttttgg ccgtctgtca
721 tcatggtgcc tgaagaaacc aggccacgtg gcggaagtct atctcgtcca tctccatgcc
781 agtgtgtacg cactctttca tcgcctttat ggaatgtacc cttgcaactt cgtctccttt
841 ttgcgttctc attacagtat gaaagaaaac ctggagactt ttgaagaagt ggtcaagcca
901 atgatggagc atgtgcgaat tcatccggaa ttagtgactg gatccaagga ccatgaactg
```

```
 961 gaccctcgaa ggtggaagag attagaaact catgatgttg tgatcgagtg tgccaaaatc
1021 tctctggatc ccacagaagc ctcatatgaa gatggctatt ctgtgtctca ccaaatctca
1081 gcccgctttc ctcatcgttc agccgatgtc accaccagcc cttatgctga cacacagaat
1141 agctatgggt gtgctacttc tacccccttac tccacgtctc ggctgatgtt gttaaatatg
1201 ccagggcagc tacctcagac tctgagttcc ccatcgacac ggctgataac tgaaccacca
1261 caagctactc tttggagccc atctatggtt tgtggtatga ccactcctcc aacttctcct
1321 ggaaatgtcc cacctgatct gtcacaccct tacagtaaag tctttggtac aactgcaggt
1381 ggaaaaggaa ctcctctggg aaccccagca acctctcctc ctccagcccc actctgtcat
1441 tcggatgact acgtgcacat ttcactcccc caggccacag tcacaccccc caggaaggaa
1501 gagagaatgg attctgcaag accatgtcta cacagacaac accatcttct gaatgacaga
1561 ggatcagaag agccacctgg cagcaaaggt tctgtcactc taagtgatct tccagggttt
1621 ttaggtgatc tggcctctga agaagatagt attgaaaaag ataaagaaga agctgcaata
1681 tctagagaac tttctgagat caccacagca gaggcagagc ctgtggttcc tcgaggaggc
1741 tttgactctc ccttttaccg agacagtctc ccaggttctc agcggaagac ccactcggca
1801 gcctccagtt ctcagggcgc cagcgtgaac cctgagcctt tacactcctc cctggacaag
1861 cttgggcctg acacaccaaa gcaagccttt actcccatag acctgccctg cggcagtgct
1921 gatgaaagcc ctgcgggaga cagggaatgc cagacttctt tggagaccag tatcttcact
1981 cccagtcctt gtaaaattcc acctccgacg agagtgggct ttggaagcgg gcagcctccc
2041 ccgtatgatc atctttttga ggtggcattg ccaaagacag cccatcattt tgtcatcagg
2101 aagactgagg agctgttaaa gaaagcaaaa ggaaacacag aggaagatgg tgtgccctct
2161 acctccccaa tggaagtgct ggacagactg atacagcagg gagcagacgc gcacagcaag
2221 gagctgaaca agttgccttt acccagcaag tctgtcgact ggacccactt tggaggctct
2281 cctccttcag atgagatccg caccctccga gaccagttgc ttttactgca caaccagtta
2341 ctctatgagc gttttaagag gcagcagcat gccctccgga acaggcggct cctccgcaag
2401 gtgatcaaag cagcagctct ggaggaacat aatgctgcca tgaaagatca gttgaagtta
2461 caagagaagg acatccagat gtggaaggtt agtctgcaga aagaacaagc tagatacaat
2521 cagctccagg agcagcgtga cactatggta accaagctcc acagccagat cagacagctg
2581 cagcatgacc gagaggaatt ctacaaccag agccaggaat tacagacgaa gctggaggac
2641 tgcaggaaca tgattgcgga gctgcggata gaactgaaga aggccaacaa caaggtgtgt
2701 cacactgagc tgctgctcag tcaggtttcc caaaagctct caaacagtga gtcggtccag
2761 cagcagatgg agttcttgaa caggcagctg ttggttcttg gggaggtcaa cgagctctat
2821 ttggaacaac tgcagaacaa gcactcagat accacaaagg aagtagaaat gatgaaagcc
2881 gcctatcgga aagagctaga aaaaaacaga agccatgttc tccagcagac tcagaggctt
2941 gatacctccc aaaaacggat tttggaactg gaatctcacc tggccaagaa agaccacctt
3001 cttttggaac agaagaaata tctagaggat gtcaaactcc aggcaagagg acagctgcag
3061 gccgcagaga gcaggtatga ggctcagaaa aggataaccc aggtgtttga attggagatc
3121 ttagatttat atggcaggtt ggagaaagat ggcctcctga aaaaacttga agaagaaaaa
3181 gcagaagcag ctgaagcagc agaagaaagg cttgactgtt gtaatgacgg gtgctcagat
3241 tccatggtag ggcacaatga agaggcatct ggccacaacg gtgagaccaa gacccccagg
3301 cccagcagcg cccggggcag tagtggaagc agaggtggtg gaggcagcag cagcagcagc
3361 agcgagcttt ctaccccaga gaaaccccca caccagaggg caggcccatt cagcagtcgg
3421 tgggagacga ctatgggaga agcgtctgcc agcatcccca ccactgtggg ctcacttccc
3481 agttcaaaaa gcttcctggg tatgaaggct cgagagttat ttcgtaataa gagcgagagc
3541 cagtgtgatg aggacggcat gaccagtagc ctttctgaga gcctaaagac agaactgggc
3601 aaagacttgg gtgtggaagc caagattccc ctgaacctag atggccctca cccgtctccc
3661 ccgaccccgg acagtgttgg acagctacat atcatggact acaatgagac tcatcatgaa
3721 cacagctaag gaatgatggt caatcagtgt taacttgcat attgttggca cagaacagga
3781 ggtgtgaatg cacgtttcaa agctttcctg tttccagggt ctgagtgcaa gttcatgtgt
3841 ggaaatggga cggaggtcct ttggacagct gactgaatgc agaacggttt ttggatctgg
3901 cattgaaatg cctcttgacc ttcccctcca cccgccctaa ccccctctca tttacctcgc
3961 agtgtgttct aatccaaggg ccagttggtg ttcctcagta gctttacttt cttcctttcc
4021 cccccaaatg gttgcgtcct ttgaacctgt gcaatatgag gccaaattta atctttgagt
4081 ctaacacacc actttctgct ttcccgaagt tcagataact gggttggctc tcaattagac
4141 caggtagttt gttgcattgc aggtaagtct ggttttgtcc cttccaggag gacatagcct
4201 gcaaagctgg ttgtctttac atgaaagcgt ttacatgaga ctttccgact gctttttga
4261 ttctgaagtt cagcatctaa agcagcaggt ctagaagaac aacggtttat tcatacttgc
```

4321 attcttttgg cagttctgat aagcttccta gaaagttctg tgtaaacaga agcctgtttc
4381 agaaatctgg agctggcact gtggagacca cacacccttt gggaaagctc ttgtctcttc
4441 ttcccccact acctcttatt tatttggtgt ttgcttgaat gctggtacta ttgtgaccac
4501 aggctggtgt gtaggtggta aaacctgttc tccataggag ggaaggagca gtcactggga
4561 gaggttaccc gagaagcact tgagcatgag gaactgcacc tttaggccat ctcagcttgc
4621 tgggcctttt gttaaaccct tctgtctact ggcctccctt tgtgtgcata cgcctcttgt
4681 tcatgtcagc ttatatgtga cactgcagca gaaaggctct gaaggtccaa agagtttctg
4741 caaagtgtat gtgaccatca tttcccaggc cattagggtt gcctcactgt agcaggttct
4801 aggctaccag aagaggggca gctttttcat accaattcca actttcaggg gctgactctc
4861 cagggagctg atgtcatcac actctccatg ttagtaatgg cagagcagtc taaacagagt
4921 ccgggagaat gctggcaaag gctggctgtg tatacccact aggctgcccc acgtgctccc
4981 gagagatgac actagtcaga aaattggcag tggcagagaa tccaaactca acaagtgctc
5041 ctgaaagaaa cgctagaagc ctaagaactg tggtctggtg ttccagctga ggcaggggga
5101 tttggtagga aggagccagt gaacttggct ttcctgtttc tatctttcat taaaaagaat
5161 agaaggattc agtcataaag aggtaaaaaa ctgtcacggt acgaaatctt agtgcccacg
5221 gaggcctcga gcagagagaa tgaaagtctt ttttttttt ttttttttt agcatggcaa
5281 taaatattct agcatcccta actaaagggg actagacagt tagagactct gtcaccctag
5341 ctataccagc agaaaacctg ttcaggcagg ctttctgggt gtgactgatt cccagcctgt
5401 ggcagggcgt ggtcccaact actcagccta gcacaggctg gcagttggta ctgaattgtc
5461 agatgtggag tattagtgac accacacatt taattcagct ttgtccaaag gaaagcttaa
5521 aacccaatac agtctagttt cctggttccg ttttagaaaa ggaaaacgtg aacaaactta
5581 gaaagggaag gaaatcccat cagtgaatcc tgaaactggt tttaagtgct ttccttctcc
5641 tcatgcccaa gagatctgtg ccatagaaca agataccagg cacttaaagc cttttcctga
5701 attggaaagg aaaagaggcc caagtgcaaa agaaaaaaca ttttagaaac ggacagctta
5761 taaaaataaa gggaagaaag gaggcagcat ggagagaggc ctgtgctaga agctccatgg
5821 acgtgtctgc acagggtcct cagctcatcc atgcggcctg ggtgtccttt tactcagctt
5881 tataacaaat gtggctccaa gctcaggtgc ctttgagttc taggaggctg tgggttttat
5941 tcaactacgg ttgggagaat gagacctgga gtcatgttga aggtgcccaa cctaaaaatg
6001 taggctttca tgttgcaaag aactccagag tcagtagtta ggtttggttt ggttttggac
6061 atgataaacc tgccaagagt caacaggtca cttgatcatg ctgcagtggg tagttctaag
6121 gatggaaagg tgacagtatt actctcgaga ggcaattcag tcctgggcaa aggtattagt
6181 acaataagcg ttaagggcag agtctacctt gaaaccaatt aagcagcttg gtattcataa
6241 atattgggat tggatggcct ccatccagaa atcactatgg gtgagcatac ctgtctcagc
6301 tgtttggcca atgtgcataa cctactcgga tccccacctg acactaacca gagtcagcac
6361 aggccccgag gagcccgaag tctgctgctg tgcagcatgg aattccttta aaaaggtgca
6421 ctacagtttt agcggggagg gggataggaa gacgcagagc aaatgagctc cggagtccct
6481 gcaggtgaat aaacacacag atctgcatct gatagaactt tgatggattt tcaaaaagcc
6541 gttgacaagg ctctgctata cagtctataa aaattgttat tatgggattg gaagaaacac
6601 gtggtcatga atagaaaaaa aacaaaccca aaggtaggaa ggtcaaggtc atttcttaga
6661 tggagaagtt gtgaaagatg tccttggaga tgagttttag gaccagcatt actaaggcag
6721 gtgggcagac agtgacctct ctaggtgtgt ccacagagtt tttcaggaga gaaaactgcc
6781 tgacctttgg gactaagctg cggaatcttc ttactaagct tgaagagtgg agaggcgaga
6841 ggtgagctac tttgtgagcc aaagcttatg tgacatggtt ggggaaacag tccaaactgt
6901 tctgagaagg tgaactgtta cgacccagga caattagaaa aattcaccca ccatgccgca
6961 cattactggg taaaagcagg gcagcaggga acaaaactcc agactcttgg gccgtcccca
7021 tttgcaacag cacacatagt ttctggtata tttgttggga aagataaaac tctagcagtt
7081 gttgagggga ggatgtataa aatggtcatg gggatgaaag gatctctgag accacagagg
7141 ctcagactca ctgttaagaa tagaaaactg ggtatgcgtt tcatgtagcc agcagaactg
7201 aagtgtgctg tgacaagcca atgtgaattt ctaccaaata gtagagcata ccacttgaag
7261 aaggaaagaa ccgaagagca aacaaaagtt ctgcgtaatg agactcacct tttctcgctg
7321 aaagcactaa gaggtgggag gaggcctgca caggctggag gagggtttgg gcagagcgaa
7381 gacccggcca ggaccttggt gagatggggt gccgcccacc tcctgcggat actcttggag
7441 agttgttccc ccaggggggct ctgccccacc tggagaagga agctgcctgg tgtggagtga
7501 ctcaaatcag tatacctatc tgctgcacct tcactctcca gggtacatgc tttaaaaccg
7561 acccgcaaca agtattggaa aaatgtatcc agtctgaaga tgtttgtgta tctgtttaca
7621 tccagagttc tgtgacacat gccccccaga ttgctgcaaa gatcccaagg cattgattgc

```
7681 acttgattaa gcttttgtct gtaggtgaaa gaacaagttt aggtcgagga ctggccccta
7741 ggctgctgct gtgacccttg tcccatgtgg cttgtttgcc tgtccgggac tcttcgatgt
7801 gcccagggga gcgtgttcct gtctcttcca tgccgtcctg cagtccttat ctgctcgcct
7861 gagggaagag tagctgtagc tacaagggaa gcctgcctgg aagagccgag cacctgtgcc
7921 catggcttct ggtcatgaaa cgagttaatg atggcagagg agcttcctcc ccacttcgca
7981 gcgccacatt atccatcctc tgagataagt aggctggttt aaccattgga atggaccttt
8041 cagtggaaac cctgagagtc tgagaacccc cagaccaacc cttccctccc tttccccacc
8101 tcttacagtg tttggacagg agggtatggt gctgctctgt gtagcaagta ctttggctta
8161 tgaaagaggc agccacgcat tttgcactag gaagaatcag taatcacttt tcagaagact
8221 tctatggacc acaaatatat tacggaggaa cagattttgc taagacataa tctagtttta
8281 taactcaatc atgaatgaac catgtgtggc aaacttgcag tttaaagggg tcccatcagt
8341 gaaagaaact gatttttttt aacggactgc ttttagttaa attgaagaaa gtcagctctt
8401 gtcaaaaggt ctaaactttc ccgcctcaat cctaaaagca tgtcaacaat ccacatcaga
8461 tgccataaat atgaactgca ggataaaatg gtacaatctt agtgaatggg aattggaatc
8521 aaaagagttt gctgtccttc ttagaatgtt ctaaaatgtc aaggcagttg cttgtgttta
8581 actgtgaaca aataaaaatt tattgttttg cactacaaaa aaaaaa
```

By "Tuberous Sclerosis 1 (TSC1) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_000359, version NP_000359.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 6):

```
   1 maqqanvgel lamldspmlg vrddvtavfk enlnsdrgpm lvntlvdyyl etssqpalhi
  61 lttlqephdk hlldrineyv gkaatrlsil sllghvirlq pswkhklsqa pllpsllkcl
 121 kmdtdvvvlt tgvlvlitml pmipqsgkqh lldffdifgr lsswclkkpg hvaevylvhl
 181 hasvyalfhr lygmypcnfv sflrshysmk enletfeevv kpmmehvrih pelvtgskdh
 241 eldprrwkrl ethdvvieca kisldpteas yedgysvshq isarfphrsa dvttspyadt
 301 qnsygcatst pystsrlmll nmpgqlpqtl sspstrlite ppqatlwsps mvcgmttppt
 361 spgnvppdls hpyskvfgtt aggkgtplgt patspppapl chsddyvhis lpqatvtppr
 421 keermdsarp clhrqhhlln drgseeppgs kgsvtlsdlp gflgdlasee dsiekdkeea
 481 aisrelseit taeaepvvpr ggfdspfyrd slpgsqrkth saasssqgas vnpeplhssl
 541 dklgpdtpkq aftpidlpcg sadespagdr ecqtsletsi ftpspckipp ptrvgfgsgq
 601 pppydhlfev alpktahhfv irkteellkk akgnteedgv pstspmevld rliqqgadah
 661 skelnklplp sksvdwthfg gsppsdeirt lrdqllllhn qllyerfkrq qhalrnrrll
 721 rkvikaaale ehnaamkdql klqekdiqmw kvslqkeqar ynqlqeqrdt mvtklhsqir
 781 qlqhdreefy nqsqelqtkl edcrnmiael rielkkannk vchtelllsq vsqklsnses
 841 vqqqmeflnr qllvlgevne lyleqlqnkh sdttkevemm kaayrkelek nrshvlqqtq
 901 rldtsqkril eleshlakkd hllleqkkyl edvklqargq lqaaesryea qkritqvfel
 961 eildlygrle kdgllkklee ekaeaaeaae erldccndgc sdsmvghnee asghngetkt
1021 prpssargss gsrggggsss ssselstpek pphqragpfs srwettmgea sasipttvgs
1081 lpssksflgm karelfrnks esqcdedgmt sslseslkte lgkdlgveak iplnldgphp
1141 spptpdsvgq lhimdyneth hehs
```

By "Neurofibromin 1 (NF1) nucleic acid molecule" is meant a polynucleotide encoding a NF1 polypeptide. An exemplary NF1 nucleic acid molecule is provided at NCBI Accession No. NM_001042492, version NM_001042492.2, incorporated herein by reference, and reproduced below (SEQ ID NO: 7):

```
   1 aatctctagc tcgctcgcgc tccctctccc cgggccgtgg aaaggatccc acttccggtg
  61 gggtgtcatg gcggcgtctc ggactgtgat ggctgtgggg agacggcgct agtggggaga
 121 gcgaccaaga ggccccctcc cctccccggg tccccttccc ctatcccccc ccccccagcc
```

```
 181 tccttgccaa cgcccccttt ccctctcccc ctcccgctcg gcgctgaccc cccatcccca
 241 cccccgtggg aacactggga gcctgcactc cacagaccct ctccttgcct cttccctcac
 301 ctcagcctcc gctccccgcc ctcttcccgg cccagggcgc cggcccaccc ttccctccgc
 361 cgcccccccgg ccgcggggag gacatggccg cgcacaggcc ggtggaatgg gtccaggccg
 421 tggtcagccg cttcgacgag cagcttccaa taaaaacagg acagcagaac acacatacca
 481 aagtcagtac tgagcacaac aaggaatgtc taatcaatat ttccaaatac aagttttctt
 541 tggttataag cggcctcact actattttaa agaatgttaa caatatgaga atatttggag
 601 aagctgctga aaaaaattta tatctctctc agttgattat attggataca ctggaaaaat
 661 gtcttgctgg gcaaccaaag gacacaatga gattagatga aacgatgctg gtcaaacagt
 721 tgctgccaga aatctgccat tttcttcaca cctgtcgtga aggaaaccag catgcagctg
 781 aacttcggaa ttctgcctct ggggttttat tttctctcag ctgcaacaac ttcaatgcag
 841 tctttagtcg catttctacc aggttacagg aattaactgt ttgttcagaa gacaatgttg
 901 atgttcatga tatagaattg ttacagtata tcaatgtgga ttgtgcaaaa ttaaaacgac
 961 tcctgaagga aacagcattt aaatttaaag ccctaaagaa ggttgcgcag ttagcagtta
1021 taaatagcct ggaaaaggca ttttggaact gggtagaaaa ttatccagat gaatttacaa
1081 aactgtacca gatcccacag actgatatgg ctgaatgtgc agaaaagcta tttgacttgg
1141 tggatggttt tgctgaaagc accaaacgta aagcagcagt ttggccacta caaatcattc
1201 tccttatctt gtgtccagaa ataatccagg atatatccaa agacgtggtt gatgaaaaca
1261 acatgaataa gaagttattt ctggacagtc tacgaaaagc tcttgctggc catggaggaa
1321 gtaggcagct gacagaaagt gctgcaattg cctgtgtcaa actgtgtaaa gcaagtactt
1381 acatcaattg ggaagataac tctgtcattt tcctacttgt tcagtccatg gtggttgatc
1441 ttaagaacct gctttttaat ccaagtaagc cattctcaag aggcagtcag cctgcagatg
1501 tggatctaat gattgactgc cttgtttctt gctttcgtat aagccctcac aacaaccaac
1561 actttaagat ctgcctggct cagaattcac cttctacatt tcactatgtg ctggtaaatt
1621 cactccatcg aatcatcacc aattccgcat tggattggtg gcctaagatt gatgctgtgt
1681 attgtcactc ggttgaactt cgaaatatgt ttggtgaaac acttcataaa gcagtgcaag
1741 gttgtggagc acacccagca atacgaatgg caccgagtct tacatttaaa gaaaaagtaa
1801 caagccttaa atttaaagaa aaacctacag acctggagac aagaagctat aagtatcttc
1861 tcttgtccat ggtgaaacta attcatgcag atccaaagct cttgctttgt aatccaagaa
1921 aacaggggcc cgaaacccaa ggcagtacag cagaattaat tacagggctc gtccaactgg
1981 tccctcagtc acacatgcca gagattgctc aggaagcaat ggaggctctg ctggttcttc
2041 atcagttaga tagcattgat ttgtggaatc ctgatgctcc tgtagaaaca ttttgggaga
2101 ttagctcaca aatgcttttt tacatctgca agaaattaac tagtcatcaa atgcttagta
2161 gcacagaaat tctcaagtgg ttgcgggaaa tattgatctg caggaataaa tttcttctta
2221 aaaataagca ggcagataga agttcctgtc actttctcct tttttacggg gtaggatgtg
2281 atattccttc tagtggaaat accagtcaaa tgtccatgga tcatgaagaa ttactacgta
2341 ctcctggagc ctctctccgg aagggaaaag ggaactcctc tatggatagt gcagcaggat
2401 gcagcggaac ccccccgatt tgccgacaag cccagaccaa actagaagtg gccctgtaca
2461 tgtttctgtg gaaccctgac actgaagctg ttctggttgc catgtcctgt ttccgccacc
2521 tctgtgagga agcagatatc cggtgtgggg tggatgaagt gtcagtgcat aacctcttgc
2581 ccaactataa cacattcatg gagtttgcct ctgtcagcaa tatgatgtca acaggaagag
2641 cagcacttca gaaaagagtg atggcactgc tgaggcgcat tgagcatccc actgcaggaa
2701 acactgaggc ttgggaagat acacatgcaa aatgggaaca agcaacaaag ctaatcctta
2761 actatccaaa agccaaaatg gaagatggcc aggctgctga aagccttcac aagaccattg
2821 ttaagaggcg aatgtcccat gtgagtggag gaggatccat agatttgtct gacacagact
2881 ccctacagga atggatcaac atgactggct tcctttgtgc ccttggggga gtgtgcctcc
2941 agcagagaag caattctggc ctggcaacct atagcccacc catgggtcca gtcagtgaac
3001 gtaagggttc tatgatttca gtgatgtctt cagagggaaa cgcagataca cctgtcagca
3061 aatttatgga tcggctgttg tccttaatgg tgtgtaacca tgagaaagtg ggacttcaaa
3121 tacggaccaa tgttaaggat ctggtgggtc tagaattgag tcctgctctg tatccaatgc
3181 tatttaacaa attgaagaat accatcagca agtttttga ctcccaagga caggttttat
3241 tgactgatac caatactcaa tttgtagaac aaaccatagc tataatgaag aacttgctag
3301 ataatcatac tgaaggcagc tctgaacatc tagggcaagc tagcattgaa acaatgatgt
3361 taaatctggt caggtatgtt cgtgtgcttg ggaatatggt ccatgcaatt caaataaaaa
3421 cgaaactgtg tcaattagtt gaagtaatga tggcaaggag agatgacctc tcattttgcc
3481 aagagatgaa atttaggaat aagatggtag aatacctgac agactgggtt atgggaacat
```

3541 caaaccaagc agcagatgat gatgtaaaat gtcttacaag agatttggac caggcaagca
3601 tggaagcagt agtttcactt ctagctggtc tccctctgca gcctgaagaa ggagatggtg
3661 tggaattgat ggaagccaaa tcacagttat ttcttaaata cttcacatta tttatgaacc
3721 ttttgaatga ctgcagtgaa gttgaagatg aaagtgcgca aacaggtggc aggaaacgtg
3781 gcatgtctcg gaggctggca tcactgaggc actgtacggt ccttgcaatg tcaaacttac
3841 tcaatgccaa cgtagacagt ggtctcatgc actccatagg cttaggttac cacaaggatc
3901 tccagacaag agctacattt atggaagttc tgacaaaaat ccttcaacaa ggcacagaat
3961 ttgacacact tgcagaaaca gtattggctg atcggtttga gagattggtg gaactggtca
4021 caatgatggg tgatcaagga gaactcccta tagcgatggc tctggccaat gtggttcctt
4081 gttctcagtg ggatgaacta gctcgagttc tggttactct gtttgattct cggcatttac
4141 tctaccaact gctctggaac atgttttcta aagaagtaga attggcagac tccatgcaga
4201 ctctcttccg aggcaacagc ttggccagta aaataatgac attctgtttc aaggtatatg
4261 gtgctaccta tctacaaaaa ctcctggatc ctttattacg aattgtgatc acatcctctg
4321 attggcaaca tgttagcttt gaagtggatc ctaccaggtt agaaccatca gagagccttg
4381 aggaaaacca gcggaacctc cttcagatga ctgaaaagtt cttccatgcc atcatcagtt
4441 cctcctcaga attccccccc caacttcgaa gtgtgtgcca ctgtttatac caggcaactt
4501 gccactccct actgaataaa gctacagtaa aagaaaaaaa ggaaaacaaa aaatcagtgg
4561 ttagccagcg tttccctcag aacagcatcg gtgcagtagg aagtgccatg ttcctcagat
4621 ttatcaatcc tgccattgtc tcaccgtatg aagcagggat tttagataaa aagccaccac
4681 ctagaatcga aaggggcttg aagttaatgt caaagatact tcagagtatt gccaatcatg
4741 ttctcttcac aaaagaagaa catatgcggc ctttcaatga ttttgtgaaa agcaactttg
4801 atgcagcacg caggtttttc cttgatatag catctgattg tcctacaagt gatgcagtaa
4861 atcatagtct ttccttcata agtgacggca atgtgcttgc tttacatcgt ctactctgga
4921 acaatcagga gaaaattggg cagtatcttt ccagcaacag ggatcataaa gctgttggaa
4981 gacgaccttt tgataagatg gcaacacttc ttgcatacct gggtcctcca gagcacaaac
5041 ctgtggcaga tacacactgg tccagcctta accttaccag ttcaaagttt gaggaattta
5101 tgactaggca tcaggtacat gaaaaagaag aattcaaggc tttgaaaacg ttaagtattt
5161 tctaccaagc tgggacttcc aaagctggga atcctatttt ttattatgtt gcacggaggt
5221 tcaaaactgg tcaaatcaat ggtgatttgc tgatatacca tgtcttactg actttaaagc
5281 catattatgc aaagccatat gaaattgtag tggaccttac ccataccggg cctagcaatc
5341 gctttaaaac agactttctc tctaagtggt ttgttgtttt tcctggcttt gcttacgaca
5401 acgtctccgc agtctatatc tataactgta actcctgggt cagggagtac accaagtatc
5461 atgagcggct gctgactggc ctcaaaggta gcaaaaggct tgttttcata gactgtcctg
5521 ggaaactggc tgagcacata gagcatgaac aacagaaact acctgctgcc accttggctt
5581 tagaagagga cctgaaggta ttccacaatg ctctcaagct agctcacaaa gacaccaaag
5641 tttctattaa agttggttct actgctgtcc aagtaacttc agcagagcga acaaaagtcc
5701 tagggcaatc agtctttcta aatgacattt attatgcttc ggaaattgaa gaaatctgcc
5761 tagtagatga gaaccagttc accttaacca ttgcaaacca gggcacgccg ctcaccttca
5821 tgcaccagga gtgtgaagcc attgtccagt ctatcattca tatccggacc cgctgggaac
5881 tgtcacagcc cgactctatc ccccaacaca ccaagattcg gccaaaagat gtccctggga
5941 cactgctcaa tatcgcatta cttaatttag gcagttctga cccgagttta cggtcagctg
6001 cctataatct tctgtgtgcc ttaacttgta cctttaattt aaaaatcgag ggccagttac
6061 tagagacatc aggtttatgt atccctgcca acaacaccct ctttattgtc tctattagta
6121 agacactggc agccaatgag ccacacctca cgttagaatt tttggaagag tgtatttctg
6181 gatttagcaa atctagtatt gaattgaaac acctttgttt ggaatacatg actccatggc
6241 tgtcaaatct agttcgtttt tgcaagcata atgatgatgc caaacgacaa agagttactg
6301 ctattcttga caagctgata acaatgacca tcaatgaaaa acagatgtac ccatctattc
6361 aagcaaaaat atggggaagc cttgggcaga ttacagatct gcttgatgtt gtactagaca
6421 gtttcatcaa aaccagtgca acaggtggct tgggatcaat aaaagctgag gtgatggcag
6481 atactgctgt agctttggct tctggaaatg tgaaattggt ttcaagcaag gttattggaa
6541 ggatgtgcaa aataattgac aagacatgct tatctccaac tcctacttta gaacaacatc
6601 ttatgtggga tgatattgct attttagcac gctacatgct gatgctgtcc ttcaacaatt
6661 cccttgatgt ggcagctcat cttccctacc tcttccacgt tgttactttc ttagtagcca
6721 caggtccgct ctcccttaga gcttccacac atggactggt cattaatatc attcactctc
6781 tgtgtacttg ttcacagctt cattttagtg aagagaccaa gcaagttttg agactcagtc
6841 tgacagagtt ctcattaccc aaattttact tgctgtttgg cattagcaaa gtcaagtcag

```
 6901 ctgctgtcat tgccttccgt tccagttacc gggacaggtc attctctcct ggctcctatg
 6961 agagagagac ttttgctttg acatccttgg aaacagtcac agaagctttg ttggagatca
 7021 tggaggcatg catgagagat attccaacgt gcaagtggct ggaccagtgg acagaactag
 7081 ctcaaagatt tgcattccaa tataatccat ccctgcaacc aagagctctt gttgtctttg
 7141 ggtgtattag caaacgagtg tctcatgggc agataaagca gataatccgt attcttagca
 7201 aggcacttga gagttgctta aaaggacctg acacttacaa cagtcaagtt ctgatagaag
 7261 ctacagtaat agcactaacc aaattacagc cacttcttaa taaggactcg cctctgcaca
 7321 aagccctctt ttgggtagct gtggctgtgc tgcagcttga tgaggtcaac ttgtattcag
 7381 caggtaccgc acttcttgaa caaaacctgc atactttaga tagtctccgt atattcaatg
 7441 acaagagtcc agaggaagta tttatggcaa tccggaatcc tctggagtgg cactgcaagc
 7501 aaatggatca ttttgttgga ctcaatttca actctaactt taactttgca ttggttggac
 7561 acctttttaaa agggtacagg catccttcac ctgctattgt tgcaagaaca gtcagaattt
 7621 tacatacact actaactctg gttaacaaac acagaaattg tgacaaattt gaagtgaata
 7681 cacagagcgt ggcctactta gcagctttac ttacagtgtc tgaagaagtt cgaagtcgct
 7741 gcagcctaaa acatagaaag tcacttcttc ttactgatat ttcaatggaa aatgttccta
 7801 tggatacata tcccattcat catggtgacc cttcctatag gacactaaag gagactcagc
 7861 catggtcctc tcccaaaggt tctgaaggat accttgcagc cacctatcca actgtcggcc
 7921 agaccagtcc ccgagccagg aaatccatga gcctggacat ggggcaacct tctcaggcca
 7981 acactaagaa gttgcttgga acaaggaaaa gttttgatca cttgatatca gacacaaagg
 8041 ctcctaaaag gcaagaaatg gaatcaggga tcacaacacc ccccaaaatg aggagagtag
 8101 cagaaactga ttatgaaatg gaaactcaga ggatttcctc atcacaacag cacccacatt
 8161 tacgtaaagt ttcagtgtct gaatcaaatg ttctcttgga tgaagaagta cttactgatc
 8221 cgaagatcca ggcgctgctt cttactgttc tagctacact ggtaaaatat accacagatg
 8281 agtttgatca acgaattctt tatgaatact tagcagaggc cagtgttgtg tttcccaaag
 8341 tctttcctgt tgtgcataat ttgttggact ctaagatcaa caccctgtta tcattgtgcc
 8401 aagatccaaa tttgttaaat ccaatccatg gaattgtgca gagtgtggtg taccatgaag
 8461 aatccccacc acaataccaa acatcttacc tgcaaagttt tggttttaat ggcttgtggc
 8521 ggtttgcagg accgttttca aagcaaacac aaattccaga ctatgctgag cttattgtta
 8581 agtttcttga tgccttgatt gacacgtacc tgcctggaat tgatgaagaa accagtgaag
 8641 aatccctcct gactcccaca tctccttacc ctcctgcact gcagagccag cttagtatca
 8701 ctgccaacct taacctttct aattccatga cctcacttgc aacttcccag cattccccag
 8761 gaatcgacaa ggagaacgtt gaactctccc ctaccactgg ccactgtaac agtggacgaa
 8821 ctcgccacgg atccgcaagc caagtgcaga agcaaagaag cgctggcagt ttcaaacgta
 8881 atagcattaa gaagatcgtg tgaagcttgc ttgctttctt ttttaaaatc aacttaacat
 8941 gggctcttca ctagtgaccc cttccctgtc cttgcccttt ccccccatgt tgtaatgctg
 9001 cacttcctgt tttataatga acccatccgg tttgccatgt tgccagatga tcaactcttc
 9061 gaagccttgc ctaaatttaa tgctgccttt tctttaactt tttttcttct acttttggcg
 9121 tgtatctggt atatgtaagt gttcagaaca actgcaaaga aagtgggagg tcaggaaact
 9181 tttaactgag aaatctcaat tgtaagagag gatgaattct tgaatactgc tactactggc
 9241 cagtgatgaa agccatttgc acagagctct gccttctgtg gttttccctt cttcatccta
 9301 cagagtaaag tgttagtcct atttatacat ttttcaagat acaagtttat gagagaaata
 9361 gtattataac cccagtatgt ttaatctttt agctgtggac tttttttta accgtacaaa
 9421 actgaaagaa ccatagaggt caagcctcag tgacttgaca ccataaagcc acagacaagg
 9481 tacttggggg ggagggcagg gaaatttcat attttatagt ggattcttaa gaaatactaa
 9541 cacttgagta ttagcaataa ttacaggaaa ataagtgcga ccacatatat cttaacatta
 9601 ctgaattaaa actatggctt ctaagtcctt atccaaactc agtcatccaa actagtttat
 9661 ttttttctcc agttgattat cttttaattt ttaattttgc taaaggtggt ttttttgtgt
 9721 tttgtttttt gtaaaccaaa actatactaa gtatagtaat tatatatata tatatatttt
 9781 ttcccctccc cctcttcttt cctaactaat tctgagcagg gtaatcagtg aacaaagtgt
 9841 tgaaaattgt tcccagaagg taattttcat agatgtttgc attagctcca tagcaaaatg
 9901 gaatggtacg tgacatttag ggtagctgat atttttattt tgttaaataa tttccaagaa
 9961 tagagtatgg tgtatattat aaatttcttt gataagatgt attttgaatg tcttttaatc
10021 ttcctcctcc tctccaaaaa aatcagaaac ctctttaaga aaacatgtag gttatatatg
10081 ctagaattgc atttaatcac tgtgaaaaga ctggtcagcc tgcattagta tgacagtagg
10141 ggggctgtta gaattgctgc tatactggtg gtatggatta tcatggcatt ggaattttca
10201 tagtaatgca gatccaattt ctttgtggta cctgcagttt acaaaataat ttgacttcag
```

```
10261 tgagcatatt ggtatctgga tgttccaatt tagaactaaa ccatatttat tacaaaaaga
10321 tattaatccc tctactccca ggttcccttt atatgttaag atataatggc tttgaggggg
10381 gaaaaaataa acctagggga gaggggagtt tcctgtagtg ctgtttcatt agaggatttc
10441 agtaaattaa attccacagc taattcaata aataatggta catttaagtg ttctgatttt
10501 aataatatat ttcacattta tccacacagt aacaatgtaa tatgttaatg taaataaaat
10561 tggttttgat actcagaaat aacaagaatt taatttttta aatttgttta cagtcctggg
10621 aaaagtaaga attatttgcc aaaataagag gaaagaaaac cttagtatta ttaatgagtt
10681 taccatagaa ttgttggaaa tactgaagac aggtgcaatt tactaaactt ttgtttttaa
10741 actattgtag aggctgcatt agaagaaaat gtttataatg acagagcaac tatgactata
10801 taaaaaagct gaaattagaa ctgtgtttag aaatagatca gtaacccagt gccaaggatg
10861 ccaagctgcc accatggtct tggctctccc acaacccagt gtttctgggg taagtttcac
10921 agtttctagg ccctggaata gcaggcagtg taagcctttg ataactttag ttcgatgttt
10981 ttcttgtttt tgtttgttgg tttggtgcat atgatagtgg gtgttatgct attttgctct
11041 tcccatcaaa ataaagaaac ttccagaggt ttactgttaa aaatactgat atttccataa
11101 acgggtttac caagggtgta gtatttcata ccgcctgaaa tgatcagcat tggcacaaat
11161 caaaattcag ccgcctttga aatgcaaaaa taccttttgac tagtaagtac atcctaggag
11221 tttgaaaact taactaaggt ttaaaattta ccttgtttaa agaacttctg acttttgagg
11281 aaaatctagc tttccaagta actaaaatgt acatgagata aacctctcac cactatgtgt
11341 cccttgagaa atgcaacact tttttagtct tcatacttgt aatctataaa agaaattctg
11401 aagtttagac caagttgccc atttctgcgt aattgacata agttctgtta aaaatattat
11461 aagtaattcg tttcggtttg tagatgtttc ccctgacttg ttaaagagga aaccaggaac
11521 tcagtcatgt ttttgtcctg gataatctac ctgttatgcc agtactccca tccgaggggc
11581 atgcccttag ttgcccagat ggagatgcag ttcagtagat ttggggcaaa gtggctacag
11641 ctctgtcttc cattcactca acacctgttc atgactgagc caggtgccca ggacacatcc
11701 taaacagtca gcttctatcc tgtgtcctag ttggggagac agagtgccag ccagcaaccc
11761 tcccaggttt gtaggtttta ggggttttca gttttgtttg ggtttttttgt tttttgtttt
11821 tgtttctaca tccttccccg actcccaggc ataatgaggc atgtcttact caatgttatg
11881 caatggattt aggcaaaaat tcattcttag tgtcagccac acaatttttt ttaatgcagt
11941 atattcacct gtaaatagtt tgtgtaaaat ttgacaaaaa aagtatattt actatactgt
12001 aaatatatgt gatgatatat tgtattattt tgctttttttg taaagcagtt agttgctgca
12061 catggataac aacaaaaatt tgattattct cgtgttagta ttgttaactt ctttttgcga
12121 ctgcgttaca tcatttaaag aaaatgctgt gtattgtaaa cttaaattgt atatgataac
12181 ttactgtcct ttccatccgg gcctaaactt tggcagttcc tttgtctaca accttgttaa
12241 tactgtaaac agttgtacgc cagcaggaaa aatactgccc aacagacaaa atcgatcatt
12301 gtaggggaaa atcatagaaa tccatttcag atctttattg ttcctcaccc cattttcctc
12361 cttgtgtatg tacttccccc acccccctttt ttttaagtaa aatgtaaatt caatctgctc
12421 taagaaaaaa aaaaaaaaaa aaaa
```

By "Neurofibromin 1 (NF1) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_001035957, version NP_001035957.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 8):

```
  1 maahrpvewv qavvsrfdeq lpiktgqqnt htkvstehnk ecliniskyk fslvisgltt
 61 ilknvnnmri fgeaaeknly lsqliildtl ekclagqpkd tmrldetmlv kqllpeichf
121 lhtcregnqh aaelrnsasg vlfslscnnf navfsristr lqeltvcsed nvdvhdiell
181 qyinvdcakl krllketafk fkalkkvaql avinslekaf wnwvenypde ftklyqipqt
241 dmaecaeklf dlvdgfaest krkaavwplq iillilcpei iqdiskdvvd ennmnkklfl
301 dslrkalagh ggsrqltesa aiacvklcka styinwedns vifllvqsmv vdlknllfnp
361 skpfsrgsqp advdlmidcl vscfrisphn nqhfkiclaq nspstfhyvl vnslhriitn
421 saldwwpkid avychsvelr nmfgetlhka vqgcgahpai rmapsltfke kvtslkfkek
481 ptdletrsyk ylllsmvkli hadpkllicn prkqgpetqg staelitglv qlvpqshmpe
541 iaqeameall vlhqldsidl wnpdapvetf weissqmlfy ickkltshqm lssteilkwl
601 reilicrnkf llknkqadrs schfllfygv gcdipssgnt sqmsmdheel lrtpgaslrk
```

```
 661 gkgnssmdsa agcsgtppic rqaqtkleva lymflwnpdt eavlvamscf rhlceeadir
 721 cgvdevsvhn llpnyntfme fasvsnmmst graalqkrvm allrriehpt agnteawedt
 781 hakweqatkl ilnypkakme dgqaaeslhk tivkrrmshv sgggsidlsd tdslqewinm
 841 tgflcalggv clqqrsnsgl atysppmgpv serkgsmisv mssegnadtp vskfmdrlls
 901 lmvcnhekvg lqirtnvkdl vglelspaly pmlfnklknt iskffdsqgq vlltdtntqf
 961 veqtiaimkn lldnhtegss ehlgqasiet mmlnlvryvr vlgnmvhaiq iktklcqlve
1021 vmmarrddls fcqemkfrnk mveyltdwvm gtsnqaaddd vkcltrdldq asmeavvsll
1081 aglplqpeeg dgvelmeaks qlflkyftlf mnllndcsev edesaqtggr krgmsrrlas
1141 lrhctvlams nllnanvdsg lmhsiglgyh kdlqtratfm evltkilqqg tefdtlaetv
1201 ladrferlve lvtmmgdqge lpiamalanv vpcsqwdela rvlvtlfdsr hllyqllwnm
1261 fskevelads mqtlfrgnsl askimtfcfk vygatylqkl ldpllrivit ssdwqhvsfe
1321 vdptrlepse sleenqrnll qmtekffhai issssefppq lrsvchclyq atchsllnka
1381 tvkekkenkk svvsqrfpqn sigavgsamf lrfinpaivs pyeagildkk ppprierglk
1441 lmskilqsia nhvlftkeeh mrpfndfvks nfdaarrffl diasdcptsd avnhslsfis
1501 dgnvlalhrl lwnnqekigq ylssnrdhka vgrrpfdkma tllaylgppe hkpvadthws
1561 slnltsskfe efmtrhqvhe keefkalktl sifyqagtsk agnpifyyva rrfktgqing
1621 dlliyhvllt lkpyyakpye ivvdlthtgp snrfktdfls kwfvvfpgfa ydnvsavyiy
1681 ncnswvreyt kyherlltgl kgskrlvfid cpgklaehie heqqklpaat laleedlkvf
1741 hnalklahkd tkvsikvgst avqvtsaert kvlgqsvfln diyyaseiee iclvdenqft
1801 ltianqgtpl tfmhqeceai vqsiihirtr welsqpdsip qhtkirpkdv pgtllniall
1861 nlgssdpslr saaynllcal tctfnlkieg qlletsglci panntlfivs isktlaanep
1921 hltlefleec isgfskssie lkhlcleymt pwlsnlvrfc khnddakrqr vtaildklit
1981 mtinekqmyp siqakiwgsl gqitdlldvv ldsfiktsat gglgsikaev madtavalas
2041 gnvklvsskv igrmckiidk tclsptptle qhlmwddiai larymlmlsf nnsldvaahl
2101 pylfhvvtfl vatgplslra sthglvinii hslctcsqlh fseetkqvlr lsltefslpk
2161 fyllfgiskv ksaaviafrs syrdrsfspg syeretfalt sletvteall eimeacmrdi
2221 ptckwldqwt elaqrfafqy npslqpralv vfgciskrvs hgqikqiiri lskalesclk
2281 gpdtynsqvl ieatvialtk lqpllnkdsp lhkalfwvav avlqldevnl ysagtalleq
2341 nlhtldslri fndkspeevf mairnplewh ckqmdhfvgl nfnsnfnfal vghllkgyrh
2401 pspaivartv rilhtlltlv nkhrncdkfe vntqsvayla alltvseevr srcslkhrks
2461 llltdismen vpmdtypihh gdpsyrtlke tqpwsspkgs egylaatypt vgqtsprark
2521 smsldmgqps qantkkllgt rksfdhlisd tkapkrqeme sgittppkmr rvaetdyeme
2581 tqrisssqqh phlrkvsvse snvlldeevl tdpkiqalll tvlatlvkyt tdefdqrily
2641 eylaeasvvf pkvfpvvhnl ldskintlls lcqdpnllnp ihgivqsvvy heesppqyqt
2701 sylqsfgfng lwrfagpfsk qtqipdyael ivkfldalid tylpgideet seeslltpts
2761 pyppalqsql sitanlnlsn smtslatsqh spgidkenve lspttghcns grtrhgsasq
2821 vqkqrsagsf krnsikkiv
```

By "RB Transcriptional Corepressor 1 (RB1) nucleic acid molecule" is meant a polynucleotide encoding a RB1 polypeptide. An exemplary RB1 nucleic acid molecule is provided at NCBI Accession No. NM_000321, version NM_000321.2, incorporated herein by reference, and reproduced below (SEQ ID NO: 9):

```
   1 gctcagttgc cgggcggggg agggcgcgtc cggtttttct caggggacgt tgaaattatt
  61 tttgtaacgg gagtcgggag aggacggggc gtgccccgac gtgcgcgcgc gtcgtcctcc
 121 ccggcgctcc tccacagctc gctggctccc gccgcggaaa ggcgtcatgc cgcccaaaac
 181 cccccgaaaa acggccgcca ccgccgccgc tgccgccgcg gaacccccgg caccgccgcc
 241 gccgcccccc cctgaggagg acccagagca ggacagcggc ccggaggacc tgcctctcgt
 301 caggcttgag tttgaagaaa cagaagaacc tgattttact gcattatgtc agaaattaaa
 361 gataccagat catgtcagag agagagcttg gttaacttgg gagaaagttt catctgtgga
 421 tggagtattg ggaggttata ttcaaaagaa aaaggaactg tggggaatct gtatctttat
 481 tgcagcagtt gacctagatg agatgtcgtt cacttttact gagctacaga aaaacataga
 541 aatcagtgtc cataaattct ttaacttact aaaagaaatt gataccagta ccaaagttga
```

601 taatgctatg tcaagactgt tgaagaagta tgatgtattg tttgcactct tcagcaaatt
661 ggaaaggaca tgtgaactta tatatttgac acaacccagc agttcgatat ctactgaaat
721 aaattctgca ttggtgctaa aagtttcttg gatcacattt ttattagcta aaggggaagt
781 attacaaatg gaagatgatc tggtgatttc atttcagtta atgctatgtg tccttgacta
841 ttttattaaa ctctcacctc ccatgttgct caaagaacca tataaaacag ctgttatacc
901 cattaatggt tcacctcgaa cacccaggcg aggtcagaac aggagtgcac ggatagcaaa
961 acaactagaa aatgatacaa gaattattga agttctctgt aaagaacatg aatgtaatat
1021 agatgaggtg aaaaatgttt atttcaaaaa ttttatacct tttatgaatt ctcttggact
1081 tgtaacatct aatggacttc cagaggttga aaatctttct aaacgatacg aagaaattta
1141 tcttaaaaat aaagatctag atgcaagatt atttttggat catgataaaa ctcttcagac
1201 tgattctata gacagttttg aaacacagag aacaccacga aaaagtaacc ttgatgaaga
1261 ggtgaatgta attcctccac acactccagt taggactgtt atgaacacta tccaacaatt
1321 aatgatgatt ttaaattcag caagtgatca accttcagaa aatctgattt cctattttaa
1381 caactgcaca gtgaatccaa aagaaagtat actgaaaaga gtgaaggata taggatacat
1441 ctttaaagag aaatttgcta aagctgtggg acagggttgt gtcgaaattg gatcacagcg
1501 atacaaactt ggagttcgct tgtattaccg agtaatggaa tccatgctta aatcagaaga
1561 agaacgatta tccattcaaa attttagcaa acttctgaat gacaacattt ttcatatgtc
1621 tttattggcg tgcgctcttg aggttgtaat ggccacatat agcagaagta catctcagaa
1681 tcttgattct ggaacagatt tgtctttccc atggattctg aatgtgctta atttaaaagc
1741 ctttgatttt tacaaagtga tcgaaagttt tatcaaagca gaaggcaact tgacaagaga
1801 aatgataaaa catttagaac gatgtgaaca tcgaatcatg gaatcccttg catggctctc
1861 agattcacct ttatttgatc ttattaaaca atcaaaggac cgagaaggac caactgatca
1921 ccttgaatct gcttgtcctc ttaatcttcc tctccagaat aatcacactg cagcagatat
1981 gtatctttct cctgtaagat ctccaaagaa aaaaggttca actacgcgtg taaattctac
2041 tgcaaatgca gagacacaag caacctcagc cttccagacc cagaagccat tgaaatctac
2101 ctctctttca ctgttttata aaaaagtgta tcggctagcc tatctccggc taaatacact
2161 ttgtgaacgc cttctgtctg agcacccaga attagaacat atcatctgga ccctttccca
2221 gcacaccctg cagaatgagt atgaactcat gagagacagg catttggacc aaattatgat
2281 gtgttccatg tatggcatat gcaaagtgaa gaatatagac cttaaattca aaatcattgt
2341 aacagcatac aaggatcttc ctcatgctgt tcaggagaca ttcaaacgtg ttttgatcaa
2401 agaagaggag tatgattcta ttatagtatt ctataactcg gtcttcatgc agagactgaa
2461 aacaaatatt ttgcagtatg cttccaccag gccccctacc ttgtcaccaa tacctcacat
2521 tcctcgaagc ccttacaagt ttcctagttc acccttacgg attcctggag ggaacatcta
2581 tatttcaccc ctgaagagtc catataaaat ttcagaaggt ctgccaacac caacaaaaat
2641 gactccaaga tcaagaatct tagtatcaat tggtgaatca ttcgggactt ctgagaagtt
2701 ccagaaaata aatcagatgg tatgtaacag cgaccgtgtg ctcaaaagaa gtgctgaagg
2761 aagcaaccct cctaaaccac tgaaaaaact acgctttgat attgaaggat cagatgaagc
2821 agatggaagt aaacatctcc caggagagtc caaatttcag cagaaactgg cagaaatgac
2881 ttctactcga acacgaatgc aaaagcagaa aatgaatgat agcatggata cctcaaacaa
2941 ggaagagaaa tgaggatctc aggaccttgg tggacactgt gtacacctct ggattcattg
3001 tctctcacag atgtgactgt ataactttcc caggttctgt ttatggccac atttaatatc
3061 ttcagctctt tttgtggata taaaatgtgc agatgcaatt gtttgggtga ttcctaagcc
3121 acttgaaatg ttagtcattg ttatttatac aagattgaaa atcttgtgta aatcctgcca
3181 tttaaaaagt tgtagcagat tgtttcctct tccaaagtaa aattgctgtg ctttatggat
3241 agtaagaatg gccctagagt gggagtcctg ataacccagg cctgtctgac tactttgcct
3301 tcttttgtag catataggtg atgtttgctc ttgtttttat taatttatat gtatattttt
3361 ttaatttaac atgaacaccc ttagaaaatg tgtcctatct atcttccaaa tgcaatttga
3421 ttgactgccc attcaccaaa attatcctga actcttctgc aaaaatggat attattagaa
3481 attagaaaaa aattactaat tttacacatt agattttatt ttactattgg aatctgatat
3541 actgtgtgct tgttttataa aattttgctt ttaattaaat aaaagctgga agcaaagtat
3601 aaccatatga tactatcata ctactgaaac agatttcata cctcagaatg taaaagaact
3661 tactgattat tttcttcatc caacttatgt ttttaaatga ggattattga tagtactctt
3721 ggtttttata ccattcagat cactgaattt ataaagtacc catctagtac ttgaaaaagt
3781 aaagtgttct gccagatctt aggtatagag gaccctaaca cagtatatcc caagtgcact
3841 ttctaatgtt tctgggtcct gaagaattaa gatacaaatt aattttactc cataaacaga
3901 ctgttaatta taggagcctt aatttttttt tcatagagat ttgtctaatt gcatctcaaa

```
3961 attattctgc cctccttaat ttgggaaggt ttgtgttttc tctggaatgg tacatgtctt
4021 ccatgtatct tttgaactgg caattgtcta tttatctttt atttttttaa gtcagtatgg
4081 tctaacactg gcatgttcaa agccacatta tttctagtcc aaaattacaa gtaatcaagg
4141 gtcattatgg gttaggcatt aatgtttcta tctgattttg tgcaaaagct tcaaattaaa
4201 acagctgcat tagaaaaaga ggcgcttctc ccctccccta cacctaaagg tgtatttaaa
4261 ctatcttgtg tgattaactt atttagagat gctgtaactt aaaatagggg atatttaagg
4321 tagcttcagc tagcttttag gaaaatcact ttgtctaact cagaattatt tttaaaaaga
4381 aatctggtct tgttagaaaa caaaatttta ttttgtgctc atttaagttt caaacttact
4441 attttgacag ttattttgat aacaatgaca ctagaaaact tgactccatt tcatcattgt
4501 ttctgcatga atatcataca aatcagttag tttttaggtc aagggcttac tatttctggg
4561 tcttttgcta ctaagttcac attagaatta gtgccagaat tttaggaact tcagagatcg
4621 tgtattgaga tttcttaaat aatgcttcag atattattgc tttattgctt ttttgtattg
4681 gttaaaactg tacatttaaa attgctatgt tactattttc tacaattaat agtttgtcta
4741 ttttaaaata aattagttgt taagagtctt aa
```

By "RB Transcriptional Corepressor 1 (RB1) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_000312, version NP_000312.2, incorporated herein by reference, as reproduced below (SEQ ID NO: 10):

```
  1 mppktprkta ataaaaaaep papppppppe edpeqdsgpe dlplvrlefe eteepdftal
 61 cqklkipdhv rerawltwek vssvdgvlgg yiqkkkelwg icifiaavdl demsftftel
121 qknieisvhk ffnllkeidt stkvdnamsr llkkydvlfa lfsklertce liyltqpsss
181 isteinsalv lkvswitfll akgevlqmed dlvisfqlml cvldyfikls ppmllkepyk
241 tavipingsp rtprrgqnrs ariakqlend triievlcke hecnidevkn vyfknfipfm
301 nslglvtsng lpevenlskr yeeiylknkd ldarlfldhd ktlqtdsids fetqrtprks
361 nldeevnvip phtpvrtvmn tiqqlmmiln sasdqpsenl isyfnnctvn pkesilkrvk
421 digyifkekf akavgqgcve igsqryklgv rlyyrvmesm lkseeerlsi qnfskllndn
481 ifhmsllaca levvmatysr stsqnldsgt dlsfpwilnv lnlkafdfyk viesfikaeg
541 nltremikhl ercehrimes lawlsdsplf dlikqskdre gptdhlesac plnlplqnnh
601 taadmylspv rspkkkgstt rvnstanaet qatsafqtqk plkstslslf ykkvyrlayl
661 rlntlcerll sehpelehii wtlfqhtlqn eyelmrdrhl dqimmcsmyg ickvknidlk
721 fkiivtaykd lphavqetfk rvlikeeeyd siivfynsvf mqrlktnilq yastrpptls
781 piphiprspy kfpssplrip ggniyisplk spykiseglp tptkmtprsr ilvsigesfg
841 tsekfqkinq mvcnsdrvlk rsaegsnppk plkklrfdie gsdeadgskh lpgeskfqqk
901 laemtstrtr mqkqkmndsm dtsnkeek
```

By "C-Src Tyrosine Kinase (CSK) nucleic acid molecule" is meant a polynucleotide encoding a CSK polypeptide. An exemplary CSK nucleic acid molecule is provided at NCBI Accession No. NM_004383, version NM_004383.2, incorporated herein by reference, and reproduced below (SEQ ID NO: 30):

```
  1 ccgggccgcg cttcctctcg ccaggcctgc gagcttcctc ccagcggagc cctgggcgag
 61 ccgaggttgg ccgccgccgc cgccgagccc gctgccgccc tcccgctcct gccccacccg
121 cgccttgccc ggggggcttct gccggggtgg ggtccgagcc gggcgaccgc ccggctgcgc
181 cgccgtcggg gccgtaaccc ggcccgccgt ccctcccgcc ccagccagcc tctggccgcc
241 ggagcccgcg gggcgtggag cgcgaggagc cccgcggccc cgatcgagcg tccggggcgg
301 cccccggcag ccagcgcgac gttccaaaat cgaacctcag tggcggcgct cggaagcgga
```

```
 361 actctgccgg ggccgcgccg gctacattgt ttcctccccc cgactccctc ccgccccctt
 421 cccccgcctt tcttccctcc gcgacccggg ccgtgcgtcc gtccccctgc ctctgcctgg
 481 cggtccctcc tcccctctcc ttgcacccat acctctttgt accgcacccc ctggggaccc
 541 ctgcgccccc ccccctcccc ctgaccgcat ggaccgtccc gcaggccgct gatgccgccc
 601 gcggcgaggt ggcccggacc gcagtgcccc aagagagctc taatggtacc aagtgacagg
 661 ttggctttac tgtgactcgg ggacgccaga gctcctgaga agatgtcagc aatacaggcc
 721 gcctggccat ccggtacaga atgtattgcc aagtacaact tccacggcac tgccgagcag
 781 gacctgccct tctgcaaagg agacgtgctc accattgtgg ccgtcaccaa ggaccccaac
 841 tggtacaaag ccaaaaacaa ggtgggccgt gagggcatca tcccagccaa ctacgtccag
 901 aagcgggagg gcgtgaaggc gggtaccaaa ctcagcctca tgccttggtt ccacggcaag
 961 atcacacggg agcaggctga gcggcttctg tacccgccgg agacaggcct gttcctggtg
1021 cgggagagca ccaactaccc cggagactac acgctgtgcg tgagctgcga cggcaaggtg
1081 gagcactacc gcatcatgta ccatgccagc aagctcagca tcgacgagga ggtgtacttt
1141 gagaacctca tgcagctggt ggagcactac acctcagacg cagatggact ctgtacgcgc
1201 ctcattaaac caaaggtcat ggagggcaca gtggcggccc aggatgagtt ctaccgcagc
1261 ggctgggccc tgaacatgaa ggagctgaag ctgctgcaga ccatcgggaa gggggagttc
1321 ggagacgtga tgctgggcga ttaccgaggg aacaaagtcg ccgtcaagtg cattaagaac
1381 gacgccactg cccaggcctt cctggctgaa gcctcagtca tgacgcaact gcggcatagc
1441 aacctggtgc agctcctggg cgtgatcgtg gaggagaagg gcgggctcta catcgtcact
1501 gagtacatgg ccaaggggag ccttgtggac tacctgcggt ctaggggtcg gtcagtgctg
1561 ggcggagact gtctcctcaa gttctcgcta gatgtctgcg aggccatgga atacctggag
1621 ggcaacaatt tcgtgcatcg agacctggct gcccgcaatg tgctggtgtc tgaggacaac
1681 gtggccaagg tcagcgactt tggtctcacc aaggaggcgt ccagcaccca ggacacgggc
1741 aagctgccag tcaagtggac agcccctgag gccctgagag agaagaaatt ctccactaag
1801 tctgacgtgt ggagtttcgg aatccttctc tgggaaatct actcctttgg gcgagtgcct
1861 tatccaagaa ttcccctgaa ggacgtcgtc cctcgggtgg agaagggcta caagatggat
1921 gcccccgacg gctgcccgcc cgcagtctat gaagtcatga agaactgctg gcacctggac
1981 gccgccatgc ggccctcctt cctacagctc cgagagcagc ttgagcacat caaaacccac
2041 gagctgcacc tgtgacggct ggcctccgcc tgggtcatgg gcctgtgggg actgaacctg
2101 gaagatcatg gacctggtgc ccctgctcac tgggcccgag cctgaactga gccccagcgg
2161 gctggcgggc ctttttcctg cgtcccagcc tgcacccctc cggccccgtc tctcttggac
2221 ccacctgtgg ggcctgggga gcccactgag gggccaggga ggaaggaggc cacggagcgg
2281 gaggcagcgc cccaccacgt cgggcttccc tggcctcccg ccactcgcct tcttagagtt
2341 ttattccttt ccttttttga gatttttttt ccgtgtgttt atttttatt atttttcaag
2401 ataaggagaa agaaagtacc cagcaaatgg gcattttaca agaagtacga atcttatttt
2461 tcctgtcctg cccgtgaggg tgggggggac cgggcccctc tctagggacc cctcgcccca
2521 gcctcattcc ccattctgtg tcccatgtcc cgtgtctcct cggtcgcccc gtgtttgcgc
2581 ttgaccatgt tgcactgttt gcatgcgccc gaggcagacg tctgtcaggg gcttggattt
2641 cgtgtgccgc tgccacccgc ccacccgcct tgtgagatgg aattgtaata aaccacgcca
2701 tgaggacacc gccgcccgcc tcggcgcttc ctccaccgag aaaaaaaaaa aaaaa
```

By "C-Src Tyrosine Kinase (CSK) polypeptide" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. BAG70102, version BAG70102.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 31):

```
  1 msaiqaawps gteciakynf hgtaeqdlpf ckgdvltiva vtkdpnwyka knkvgregii
 61 panyvqkreg vkagtklslm pwfhgkitre qaerllyppe tglflvrest nypgdytlcv
121 gcdgkvehyr imyhasklsi deevyfenlm qlvehytsda dglctrlikp kvmegtvaaq
181 defyrsgwal nmkelkllqt igkgefgdvm lgdyrgnkva vkcikndata qaflaeasvm
241 tqlrhsnlvq llgviveekg glyivteyma kgslvdylrs rgrsvlggdc llkfsldvce
301 ameylegnnf vhrdlaarnv lvsednvakv sdfgltkeas stqdtgklpv kwtapealre
361 kkfstksdvw sfgillweiy sfgrvpypri plkdvvprve kgykmdapdg cppavyevmk
421 ncwhldaamr psflqlreql ehikthelhl
```

By "Mitogen-Activated Protein Kinase 8 (MAPK8) nucleic acid molecule" is meant a polynucleotide encoding a MAPK8 polypeptide. An exemplary MAPK8 nucleic acid molecule is provided at NCBI Accession No. NM_001323320, version NM_001323320.1, incorporated herein by reference, and reproduced below (SEQ ID NO: 32):

```
   1 gacgtgcgcg ggcgtgcgcg gtgacggccc gcgtctctgt tactcagccg agcggccgag
  61 gccggacgac gcggcttgga ttgcggagcc gcgagcagcg ctgggtaacg gccgcggcga
 121 ccaccccgga cggcccctgt ccccgctggc gggcttccct gtcgccgttc gctgcgctgc
 181 cggcttcttg gtgaattttt ggatgaagcc attaaattaa ttgcttgcca tcatgagcag
 241 aagcaagcgt gacaacaatt tttatagtgt agagattgga gattctacat tcacagtcct
 301 gaaacgatat cagaatttaa aacctatagg ctcaggagct caaggaatag tatgcgcagc
 361 ttatgatgcc attcttgaaa gaaatgttgc aatcaagaag ctaagccgac catttcagaa
 421 tcagactcat gccaagcggg cctacagaga gctagttctt atgaaatgtg ttaatcacaa
 481 aaatataatt ggccttttga atgttttcac accacagaaa tccctagaag aatttcaaga
 541 tgtttacata gtcatggagc tcatggatgc aaatctttgc caagtgattc agatggagct
 601 agatcatgaa agaatgtcct accttctcta tcagatgctg tgtggaatca agcaccttca
 661 ttctgctgga attattcatc gggacttaaa gcccagtaat atagtagtaa aatctgattg
 721 cactttgaag attcttgact tcggtctggc caggactgca ggaacgagtt ttatgatgac
 781 gccttatgta gtgactcgct actacagagc acccgaggtc atccttggca tgggctacaa
 841 ggaaaacgct gactcagaac acaacaaact taaagccagt caggcaaggg atttgttatc
 901 caaaatgctg gtaatagatg catctaaaag gatctctgta gatgaagctc tccaacaccc
 961 gtacatcaat gtctggtatg atccttctga agcagaagct ccaccaccaa agatccctga
1021 caagcagtta gatgaaaggg aacacacaat agaagagtgg aaagaattga tatataagga
1081 agttatggac ttggaggaga gaaccaagaa tggagttata cgggggcagc cctctccttt
1141 aggtgcagca gtgatcaatg gctctcagca tccatcatca tcgtcgtctg tcaatgatgt
1201 gtcttcaatg tcaacagatc cgactttggc ctctgataca gacagcagtc tagaagcagc
1261 agctgggcct ctgggctgct gtagatgact acttgggcca tcgggggtg ggagggatgg
1321 ggagtcggtt agtcattgat agaactactt tgaaaacaat tcagtggtct tatttttggg
1381 tgatttttca aaaaatgtag aattcatttt gtagtaaagt agtttatttt ttttaatttc
1441 aagtgatgta atttaaaacc taagttgtgt ttcaaaacag caacaaaact gtattgtatt
1501 ttttttgctg taattaactg tataatgtaa acctaattat tttatcatgg tttaaatttt
1561 ttgcatattt gctttatctt atgctgctga tttttttaac tgaatttgta agattttgtt
1621 tatcaaagca actattatgt ggtgacttgc ctatatcatg aattatttaa gatttttata
1681 gtttttttta attagaattt atttcagatg ttttgttcat gatactatcc ttcagggtta
1741 tgtgcttatc aatgaaataa ccccagagga gtgagggaaa ataacttgta gccagttata
1801 ttcaggaata actactgtaa atgatgaacg tgttaggaga cctccaatat ttgctacttg
1861 ccaatcctaa tttagttaca agaattggta ggcaatccta cttaattttg gcaaaagccc
1921 cgtcatctaa atggcagaat aactcagagc atgtctttga agatgctggg cgtctaccac
1981 caccttatgt ccccacccta cccaacaaaa ataagtaaaa agaatatggt gtattctaca
2041 aatttgtggc atgctcaaag tttatgatca cataaaggca agaggatact tcatgaataa
2101 tacatttcaa tgcaaataaa cagatggttc acttctacta gctatgagcc tgtttttgta
2161 tacactgagt taatctactc aggctgtagg tcccagcaat gttctagagt ctggtctttc
2221 cctttcctgc agcttcgggt ccttggacct ttcctgtttc ctattacttg gagtgtctgt
2281 cagttgagca ccagttgttc tggtgtttca tttgattcta cttgtagcat aatcatttat
2341 acgagctatt gggaggttcc aaaccctacc tagatttgtg taggtgatgt atcaaatgag
2401 caatataccg ttcatctgaa aatagtagca cacagccata tataggatat cattttctaa
2461 ggactgtttc ttcacattga gcagagcagg cataaatggt ggttatttag tctaagtctt
2521 ttatttttt atacctgatt ttcaacataa cacgcaatgt ggatgtcgag tagtgttaag
2581 aatggtgctg ctcctgacaa gtgtatgtta actgtttaca ttttctatct gtagaattat
2641 ttctctatta ctgaactttt cctaagtaaa atgtctttga agtctcgtta tttctgaaat
```

```
2701 acgttgtctg taatagaccc aggcaccttt taaattatct ctggaacaag agggatttca
2761 tgtaatgaac taggaaatgc atactcacat aagcaacaag gttctaggca gaaagcccct
2821 tggaatttgt gaccaacagg agcaagaaca ggtgcggctc aacatgcaat gtctgaaaat
2881 ttgcttggca ttttattcat atatttagtg caaaattatt tttgagtgag atattttaca
2941 tcactgttaa tgtgcaatat ttaagattaa aatacattag ctttttttata tactttgaag
3001 tagcaagttt gttttcgatg gcttagagtc atgatttcca gcttcccagc ctttttatca
3061 gtccctttc taatacaaca aggtgcatta atttgattag gcaaattaga gttctaagac
3121 acttcttgaa ttgtagacag aaaatattgg attcacaatt tcagcagaaa tttgagaatg
3181 agtgtgttta tattaatttc acaattagct gtattttctg tagcatagat tatgtcactg
3241 ttgcactttc acagcagaca tgctttcaga aggttctcat attttatgtt tgattgctga
3301 taagccatct ctattgatac agattttggt taagtaagga aaaccaggtg tgtgtctgta
3361 tcatttattg taaatgccag ctgccacttg ccaaccatca tgttcagttc aattcaaaga
3421 aaacaaactc tcattactta gtgtaaacta aaatacttaa caaattatat cctaaaaaca
3481 aggtctcttt gttaaatgtt gcatgcccta ggttttaaat tactacatcc aaatacagtt
3541 ttcgtcttaa atttgttaag ctaaatatat gttggttctt tttattttgg aatcctttaa
3601 gcatcttaaa catttttttt ttgaagagaa gttacaaata acatttctat caggtagtac
3661 ttgtatgaaa ccacctttct tattctataa ttttgatttt tcaattttat atacttaata
3721 tactcactgt cttactatca gaaagttatt ttgaccaaga tttttattat cttcatagat
3781 tcagaaagag atgctaattc tgtaccaatg tcttcctggt tactattctc ttccctctaa
3841 tatatactgg ccatttgtaa aaccattgtg ttgttgggat cacttagtta tactatacgc
3901 agatagagca tctcaactct gtcatagtgt ttgctgaaca gttttcagtg tcatgcacct
3961 ttggctgcta attgttcctg acgtgcactc ttccgagttg gtaaaggcac agtgtgttca
4021 tgccagactt ctaagagaaa caccagcctc ttaaatcaga agcctacaca caacccccctt
4081 aacaatccaa agaagcttga tggtgtgcaa agaagcatcc tgccagcctt gtcattgttc
4141 tgttctatgc taatcctgct gtgttgtcta aaagatggag ggaagaggac atcagtgtct
4201 gatagtgaaa tcatcagcag gaaagtgaag ctctttcctt ggttacagat aagacttggt
4261 ttacactatt ggccagtatc tgctaaacat atgaagactt aactattcag tgttgcctag
4321 gcattcgcct gcacaacatt ttgaggttag aacatagaat attttcagaa atactgttgt
4381 agtttgtgag tgttgttcat tagttacaca ttagctatag agtggatgca tgaagcccca
4441 tgacaccagt aaacttctct taccagtagg taaaccaaac accattctgt cattagcagc
4501 cctcttaaat gttgcctctc cgtatcctgt tgcatttttg tgtgcattgt gtttctactg
4561 atctctctta ggtttttacg gaatcaaagg aaactaattt ttccttaata gcaagaaaga
4621 tgaagaggta aagggcattg aagcagaaat gtatagtttg gggtacgatt agaaaactcg
4681 taaggaaaac agaagtccta atttcaaact gactgctctt cgttaagtgc tcttaaggag
4741 agtctagtaa cagtaacact ttctggccat ttctagttta gattctcttc gttactgaaa
4801 cttttgagaa atattacctg tggattaatt ttgcacaatg ttctattctc ataatgactt
4861 acaaattaaa ctaggttttt attgaactac ctcacactaa ttttctatgc tttcccaagt
4921 aagctgttgc cctgttagat ctttactgag tgaattataa atgtgtgtta aatactttct
4981 agccaatgtt gacacaatac cagtaagtat gtaaagtata taccttacat cagtaagaga
5041 cacgtgtaaa atctttgact gtatgtcttg caaaattgtg ctcgttgaca ttattactgt
5101 ttttgtaagt agaaacctgc tcgtgatatc ggtccattta cattttacaa aaggagtaaa
5161 tcttagtaaa aattttacga agaaataaat tacttttgta ggcccaatat ttggtatatt
5221 tttgagaagc tgttaatctt ttagctgaat aatgaagtta gactgaatta cgtgtctccc
5281 tggactgtga catctatttt ctcattacag tttatcctgg tcagcagggt gtcacacctg
5341 gaaacctgag tatgatagct gacatttgct tttctccctc tgcgatgtca ttcctcctcc
5401 attcctctcc ttccctgtgt tccgttccct ctcctttcct ctagacaaaa caaaatgggg
5461 cacttttag ggaatgctga gatcattatt gtggtttttc atcattcatg ccctagtcat
5521 taaacatgca ccactggaat gtaaacaatg ttatctagta tgtcaattgg ttataatatt
5581 ttaaataaaa aagaaaaaag tggtatgaaa attatgaaa
```

By "Mitogen-Activated Protein Kinase 8 (MAPK8)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAI30573, version AAI30573.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 33):

```
  1 msrskrdnnf ysveigdstf tvlkryqnlk pigsgaqgiv caaydailer nvaikklsrp
 61 fqnqthakra yrelvlmkcv nhkniiglln vftpqkslee fqdvyivmel mdanlcqviq
121 meldhermsy llyqmlcgik hlhsagiihr dlkpsnivvk sdctlkildf glartagtsf
181 mmtpyvvtry yrapevilgm gykenvdiws vgcimgemik ggvlfpgtdh idqwnkvieq
241 lgtpcpefmk klqptvrtyv enrpkyagys feklfpdvlf padsehnklk asqardllsk
301 mlvidaskri svdealqhpy invwydpsea eapppkipdk qlderehtie ewkeliykev
361 mdleertkng virgqpsplg aavingsqhp sssssvndvs smstdptlas dtdssleaaa
421 gplgccr
```

By "Janus Kinase 3 (JAK3) nucleic acid molecule" is meant a polynucleotide encoding a JAK3 polypeptide. An exemplary JAK3 nucleic acid molecule is provided at NCBI Accession No. NM_000215, version NM_000215.3, incorporated herein by reference, and reproduced below (SEQ ID NO: 34):

```
   1 cacacaggaa ggagccgagt gggactttcc tctcgctgcc tcccggctct gcccgccctt
  61 cgaaagtcca gggtccctgc ccgctaggca agttgcactc atggcacctc caagtgaaga
 121 gacgcccctg atccctcagc gttcatgcag cctcttgtcc acggaggctg gtgccctgca
 181 tgtgctgctg cccgctcggg gccccgggcc cccccagcgc ctatctttct cctttgggga
 241 ccacttggct gaggacctgt gcgtgcaggc tgccaaggcc agcggcatcc tgcctgtgta
 301 ccactccctc tttgctctgg ccacggagga cctgtcctgc tggttccccc cgagccacat
 361 cttctccgtg gaggatgcca gcacccaagt cctgctgtac aggattcgct tttacttccc
 421 caattggttt gggctggaga agtgccaccg cttcgggcta cgcaaggatt tggccagtgc
 481 tatccttgac ctgccagtcc tggagcacct ctttgcccag caccgcagtg acctggtgag
 541 tgggcgcctc cccgtgggcc tcagtctcaa ggagcagggt gagtgtctca gcctggccgt
 601 gttggacctg gcccggatgg cgcgagagca ggcccagcgg ccgggagagc tgctgaagac
 661 tgtcagctac aaggcctgcc tacccccaag cctgcgcgac ctgatccagg gcctgagctt
 721 cgtgacgcgg aggcgtattc ggaggacggt gcgcagagcc ctgcgccgcg tggccgcctg
 781 ccaggcagac cggcactcgc tcatggccaa gtacatcatg gacctggagc ggctggatcc
 841 agccggggcc gccgagacct tccacgtggg cctccctggg gcccttggtg gccacgacgg
 901 gctggggctg ctccgcgtgg ctggtgacgg cggcatcgcc tggacccagg gagaacagga
 961 ggtcctccag cccttctgcg actttccaga aatcgtagac attagcatca agcaggcccc
1021 gcgcgttggc ccggccggag agcaccgcct ggtcactgtt accaggacag acaaccagat
1081 tttagaggcc gagttcccag ggctgcccga ggctctgtcg ttcgtggcgc tcgtggacgg
1141 ctacttccgg ctgaccacgg actcccagca cttcttctgc aaggaggtgg caccgccgag
1201 gctgctggag gaagtggccg agcagtgcca cggccccatc actctggact ttgccatcaa
1261 caagctcaag actgggggct cacgtcctgg ctcctatgtt ctccgccgca gcccccagga
1321 ctttgacagc ttcctcctca ctgtctgtgt ccagaacccc cttggtcctg attataaggg
1381 ctgcctcatc cggcgcagcc ccacaggaac cttccttctg gttggcctca gccgacccca
1441 cagcagtctt cgagagctcc tggcaacctg ctgggatggg gggctgcacg tagatggggt
1501 ggcagtgacc ctcacttcct gctgtatccc cagacccaaa gaaaagtcca acctgatcgt
1561 ggtccagaga ggtcacagcc cacccacatc atccttggtt cagccccaat cccaatacca
1621 gctgagtcag atgacatttc acaagatccc tgctgacagc ctggagtggc atgagaacct
1681 gggccatggg tccttcacca agatttaccg gggctgtcgc catgaggtgg tggatgggga
1741 ggcccgaaag acagaggtgc tgctgaaggt catggatgcc aagcacaaga actgcatgga
1801 gtcattcctg gaagcagcga gcttgatgag ccaagtgtcg taccggcatc tcgtgctgct
1861 ccacggcgtg tgcatggctg gagacagcac catggtgcag gaatttgtac acctgggggc
1921 catagacatg tatctgcgaa aacgtggcca cctggtgcca gccagctgga agctgcaggt
1981 ggtcaaacag ctggcctacg ccctcaacta tctggaggac aaaggcctgc cccatggcaa
2041 tgtctctgcc cggaaggtgc tcctggctcg ggaggggggct gatgggagcc cgcccttcat
```

2101 caagctgagt gaccctgggg tcagccccgc tgtgttaagc ctggagatgc tcaccgacag
2161 gatcccctgg gtggcccccg agtgtctccg ggaggcgcag acacttagct tggaagctga
2221 caagtggggc ttcggcgcca cggtctggga agtgtttagt ggcgtcacca tgcccatcag
2281 tgccctggat cctgctaaga aactccaatt ttatgaggac cggcagcagc tgccggcccc
2341 caagtggaca gagctggccc tgctgattca acagtgcatg gcctatgagc cggtccagag
2401 gccctccttc cgagccgtca ttcgtgacct caatagcctc atctcttcag actatgagct
2461 cctctcagac cccacacctg gtgccctggc acctcgtgat gggctgtgga atggtgccca
2521 gctctatgcc tgccaagacc ccacgatctt cgaggagaga cacctcaagt acatctcaca
2581 gctgggcaag ggcaactttg gcagcgtgga gctgtgccgc tatgacccgc taggcgacaa
2641 tacaggtgcc ctggtggccg tgaaacagct gcagcacagc gggccagacc agcagaggga
2701 ctttcagcgg gagattcaga tcctcaaagc actgcacagt gatttcattg tcaagtatcg
2761 tggtgtcagc tatggcccgg gccgccagag cctgcggctg gtcatggagt acctgcccag
2821 cggctgcttg cgcgacttcc tgcagcggca ccgcgcgcgc ctcgatgcca gccgcctcct
2881 tctctattcc tcgcagatct gcaagggcat ggagtacctg ggctcccgcc gctgcgtgca
2941 ccgcgacctg gccgcccgaa acatcctcgt ggagagcgag gcacacgtca agatcgctga
3001 cttcggccta gctaagctgc tgccgcttga caaagactac tacgtggtcc gcgagccagg
3061 ccagagcccc atttttctggt atgcccccga atccctctcg gacaacatct tctctcgcca
3121 gtcagacgtc tggagcttcg gggtcgtcct gtacgagctc ttcacctact gcgacaaaag
3181 ctgcagcccc tcggccgagt tcctgcggat gatgggatgt gagcgggatg tccccgccct
3241 ctgccgcctc ttggaactgc tggaggaggg ccagaggctg ccggcgcctc ctgcctgccc
3301 tgctgaggtt cacgagctca tgaagctgtg ctgggccccc agcccacagg accggccatc
3361 attcagcgcc ctgggccccc agctggacat gctgtggagc ggaagccggg ggtgtgagac
3421 tcatgccttc actgctcacc cagagggcaa acaccactcc ctgtcctttt catagctcct
3481 gcccgcagac ctctggatta ggtctctgtt gactggctgt gtgaccttag gcccggagct
3541 gcccctctct gggcctcaga ggccttatga gggtcctcta cttcaggaac acccccatga
3601 cattgcattt gggggggctc ccgtggcctg tagaatagcc tgtggccttt gcaatttgtt
3661 aaggttcaag acagatgggc atatgtgtca gtggggctct ctgagtcctg gcccaaagaa
3721 gcaaggaacc aaatttaaga ctctcgcatc ttcccaaccc cttaagccct ggcccccctga
3781 gtttccttt ctgtctctct ctttttattt tttttatttt tatttttatt tttgagacag
3841 agcctcgctc tgttacccag ggtggagtgc agtggtgcga tctcggctca gtgcaacctc
3901 tgcttcccag gttcaagcga ttctcctgcc tcagcctccc gagtagctgg gattacaggt
3961 gtgcaccacc acacccggct aatttttttt atttttaata gagatgaggt ttcaccatga
4021 tggccaggct gatctcgaac tcctaacctc aagtgatcct cccacctcag cctcccaaag
4081 tgttggaata ataggcatga gccactgcac ccaggctttt tttttttaa atttattatt
4141 attattttta agagacagga tcttgctacg ttgcccaggc tggtcttgaa ctcctgggct
4201 acagtgatcc tcctgcctta tcctcctaaa tagctgggac tacagcacct agttttgagt
4261 ttcctgtctt atttccaatg gggacattca tgtagctttt tttttttttt ttttttgag
4321 acggagtctc gctctgtcgc ccaggctgga gtacagtggc gcaatctagg ctcactgcaa
4381 gctccgcctc ctgggttcac accattctct cgcctcagcc tcccaagtag ctgggactac
4441 aggcgcccgc caccacaccc ggctaatttt ttgtattttt agtagagacg gggtttcacc
4501 ttgttagcca ggatggtttc catctcctga cctcgtgatc tgcccgtctc ggcctcccaa
4561 agtgctggga ttacaggcat gagccactgc gcccggccct catgtagctt taaatgtatg
4621 atctgacttc tgctccccga tctctgtttc tctggaggaa gccaaggaca agagcagttg
4681 ctgtggctgg gactctgcct tttaggggag cccgtgtatc tctttgggat cctgaaaggg
4741 ggcaggaaag gctggggtcc cagtccaccc taatggtatc tgagtgtcct agggcttcag
4801 ttttcccacc tgtccaatgg gacccttttct gtcctcaccc tacaaggggc acaaagggat
4861 gacaccaaac ctggcaggaa cttttcacgc aatcaaggga aggaaaggca ttcctggcag
4921 agggaacagc atgccaagcg tgagaaggct cagagtaagg aggttaagag cccaagtatt
4981 ggagcctaca gttttgcccc ttccatgcag tgtgacagtg ggcaagttcc tttccctctc
5041 tgggtctcag ttctgtcccc tgcaaaatgg tcagagctta ccccttggct gtgcagggtc
5101 aactttctga ctggtgagag ggattctcat gcaggttaag cttctgctgc tcctcctcac
5161 ctgcaaagct tttctgccac ttttgcctcc ttggaaaact cttatccatc tctcaaaact
5221 ccagctacca catccttgca gccttccctc atataccccc actactactg tagccctgtc
5281 cttccctcca gccccactct ggccctgggg ctggggaagt gtctgtgtcc agctgtctcc
5341 cctgacctca gggttccttg ggggctgggc tgaggcctca gtacagaggg ggctctggaa
5401 atgtttgttg actgaataaa ggaattcagt ggaaaaaaaa aaaaaaaaa

By "Janus Kinase 3 (JAK3)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAC50950, version AAC50950.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 35):

```
   1 mappseetpl ipqrscslls teagalhvll pargpgppqr lsfssgdhla edlcvqaaka
  61 sgilpvyhsl falatedlsc wfppshifsv edastqvlly rirsfyfpnw fglekchrfg
 121 lrkdlasail dlpvlehlfa qhrsdlvsgr lpvglslkeq geclslavld larmareqaq
 181 rpgellktvs ykaclppslr dliqglsfvt rrairrtvrr alprvaacqa drhslmakyi
 241 mdlerldpag aaetfhvglp galgghdglg lfrvagdggi awtqgeqvlq pfcdfpeivd
 301 isikqaprvg pagehrlvtv trtdnqilea efpglpeals fvalvdgyfr lttdsqhffc
 361 kevapprlle evaeqchgpi tldfainklk tggsrpgsyv lrrspqdfds flltvcvqnp
 421 lgpdykgcli rrsptgtfll vglsrphssl rellatcwdg glhvdgvavt ltscciprpk
 481 eksnlivvqr ghspptsslv qpqsqyqlsq mtfhkipads lewhenlghg sftkiyrgcr
 541 hevvdgeark tevllkvmda khkncmesfl eaaslmsqvs yrhlvllhgv cmagdstmvq
 601 efvhlgaidm ylrkrghlvp aswklqvvkq layalnyled kglphgnvsa rkvllarega
 661 dgsppfikls dpgvspavls lemltdripw vapeclreaq tlsleadkwg fgatvwevfs
 721 gvtmpisald pakklqfyed rqqlpapkwt elalliqqcm ayepvqrpsf ravirdlnsl
 781 issdyellsd ptpgalaprd glwngaqlya cqdptifeer hlkyisqlgk gnfgsvelcr
 841 ydplgdntga lvavkqlqhs gpdqqrdfqr eiqilkalhs dfivkyrgvs ygpgepelrl
 901 vmeylpsgcl rdflqrhrar ldasrlllys sqickgmeyl gsrrcvhrdl aarnilvese
 961 ahvkiadfgl akllpldkdy yvvrepgqsp ifwyapesls dnifsrqsdv wsfgvvlyel
1021 ftycdkscsp saeflrmmgc erdvpalcrl lelleegqrl pappacpaev helmklcwap
1081 spqdrpsfsa lgpqldmlws gsrgcethaf tahpegkhhs lsfs
```

By "Cyclin Dependent Kinase 12 (CDK12) nucleic acid molecule" is meant a polynucleotide encoding a CDK12 polypeptide. An exemplary CDK12 nucleic acid molecule is provided at NCBI Accession No. NM_015083, version NM_015083.2, incorporated herein by reference, and reproduced below (SEQ ID NO: 36):

```
   1 gtgtgactgg gtctgtgtga gggagagagt gtgtgtggtg tggaggtgaa acggaggcaa
  61 gaaagggggc tacctcagga gcgagggaca aaggggggcgt gaggcaccta ggccgcggca
 121 ccccggcgac aggaagccgt cctgaaccgg gctaccgggt aggggaaggg cccgcgtagt
 181 cctcgcaggg ccccagagct ggagtcggct ccacagcccc gggccgtcgg cttctcactt
 241 cctggacctc cccggcgccc gggcctgagg actggctcgg cggagggaga agaggaaaca
 301 gacttgagca gctccccgtt gtctcgcaac tccactgccg aggaactctc atttcttccc
 361 tcgctccttc acccccacc tcatgtagaa gggtgctgag gcgtcgggag ggaggaggag
 421 cctgggctac cgtccctgcc ctccccaccc ccttcccggg gcgctttggt gggcgtggag
 481 ttggggttgg gggggtgggt gggggttgct ttttggagtg ctggggaact tttttccctt
 541 cttcaggtca ggggaaaggg aatgcccaat tcagagagac atgggggcaa gaaggacggg
 601 agtggaggag cttctggaac tttgcagccg tcatcgggag gcggcagctc taacagcaga
 661 gagcgtcacc gcttggtatc gaagcacaag cggcataagt ccaaacactc caaagacatg
 721 gggttggtga cccccgaagc agcatccctg ggcacagtta tcaaaccttt ggtggagtat
 781 gatgatatca gctctgattc cgacaccttc tccgatgaca tggccttcaa actagaccga
 841 agggagaacg acgaacgtcg tggatcagat cggagcgacc gcctgcacaa acatcgtcac
 901 caccagcaca ggcgttcccg ggacttacta aaagctaaac agaccgaaaa agaaaaaagc
```

```
 961 caagaagtct ccagcaagtc gggatcgatg aaggaccgga tatcgggaag ttcaaagcgt
1021 tcgaatgagg agactgatga ctatgggaag gcgcaggtag ccaaaagcag cagcaaggaa
1081 tccaggtcat ccaagctcca caaggagaag accaggaaag aacgggagct gaagtctggg
1141 cacaaagacc ggagtaaaag tcatcgaaaa agggaaacac ccaaaagtta caaaacagtg
1201 gacagcccaa aacggagatc caggagcccc cacaggaagt ggtctgacag ctccaaacaa
1261 gatgatagcc cctcgggagc ttcttatggc caagattatg accttagtcc ctcacgatct
1321 catacctcga gcaattatga ctcctacaag aaaagtcctg gaagtacctc gagaaggcag
1381 tcggtcagtc ccccttacaa ggagccttcg gcctaccagt ccagcacccg gtcaccgagc
1441 ccctacagta ggcgacagag atctgtcagt ccctatagca ggagacggtc gtccagctac
1501 gaaagaagtg gctcttacag cgggcgatcg cccagtccct atggtcgaag gcggtccagc
1561 agcccctttcc tgagcaagcg gtctctgagt cggagtccac tccccagtag gaaatccatg
1621 aagtccagaa gtagaagtcc tgcatattca agacattcat cttctcatag taaaaagaag
1681 agatccagtt cacgcagtcg tcattccagt atctcacctg tcaggcttcc acttaattcc
1741 agtctgggag ctgaactcag taggaaaaag aaggaaagag cagctgctgc tgctgcagca
1801 aagatggatg gaaaggagtc caagggttca cctgtatttt tgcctagaaa agagaacagt
1861 tcagtagagg ctaaggattc aggtttggag tctaaaaagt tacccagaag tgtaaaattg
1921 gaaaaatctg ccccagatac tgaactggtg aatgtaacac atctaaacac agaggtaaaa
1981 aattcttcag atacagggaa agtaaagttg gatgagaact ccgagaagca tcttgttaaa
2041 gatttgaaag cacagggaac aagagactct aaacccatag cactgaaaga ggagattgtt
2101 actccaaagg agacagaaac atcagaaaag gagacccctc cacctcttcc cacaattgct
2161 tctcccccac cccctctacc aactactacc cctccacctc agacaccccc tttgccacct
2221 ttgcctccaa taccagctct tccacagcaa ccacctctgc ctccttctca gccagcattt
2281 agtcaggttc ctgcttccag tacttcaact ttgccccctt ctactcactc aaagacatct
2341 gctgtgtcct ctcaggcaaa ttctcagccc cctgtacagg tttctgtgaa gactcaagta
2401 tctgtaacag ctgctattcc acacctgaaa acttcaacgt tgcctccttt gcccctccca
2461 cccttattac ctggagatga tgacatggat agtccaaaag aaactcttcc ttcaaaacct
2521 gtgaagaaag agaaggaaca gaggacacgt cacttactca cagaccttcc tctcctctca
2581 gagctccctg gtggagatct gtctccccca gactctccag aaccaaaggc aatcacacca
2641 cctcagcaac catataaaaa gagaccaaaa atttgttgtc ctcgttatgg agaaagaaga
2701 caaacagaaa gcgactgggg gaaacgctgt gtggacaagt ttgacattat tgggattatt
2761 ggagaaggaa cctatggcca agtatataaa gccaaggaca aagacacagg agaactagtg
2821 gctctgaaga aggtgagact agacaatgag aaagagggct tcccaatcac agccattcgt
2881 gaaatcaaaa tccttcgtca gttaatccac cgaagtgttg ttaacatgaa ggaaattgtc
2941 acagataaac aagatgcact ggatttcaag aaggacaaag gtgcctttta ccttgtattt
3001 gagtatatgg accatgactt aatgggactg ctagaatctg gtttggtgca cttttctgag
3061 gaccatatca agtcgttcat gaaacagcta atggaaggat tggaatactg tcacaaaaag
3121 aatttcctgc atcgggatat taagtgttct aacattttgc tgaataacag tgggcaaatc
3181 aaactagcag attttggact tgctcggctc tataactctg aagagagtcg cccttacaca
3241 aacaaagtca ttactttgtg gtaccgacct ccagaactac tgctaggaga ggaacgttac
3301 acaccagcca tagatgtttg gagctgtgga tgtattcttg gggaactatt cacaaagaag
3361 cctatttttc aagccaatct ggaactggct cagctagaac tgatcagccg actttgtggt
3421 agcccttgtc cagctgtgtg gcctgatgtt atcaaactgc cctacttcaa caccatgaaa
3481 ccgaagaagc aatatcgaag gcgtctacga gaagaattct ctttcattcc ttctgcagca
3541 cttgatttat tggaccacat gctgacacta gatcctagta agcggtgcac agctgaacag
3601 accctacaga gcgacttcct taaagatgtc gaactcagca aaatggctcc tccagacctc
3661 ccccactggc aggattgcca tgagttgtgg agtaagaaac ggcgacgtca gcgacaaagt
3721 ggtgttgtag tcgaagagcc acctccatcc aaaacttctc gaaaagaaac tacctcaggg
3781 acaagtactg agcctgtgaa gaacagcagc ccagcaccac ctcagcctgc tcctggcaag
3841 gtggagtctg gggctgggga tgcaataggc cttgctgaca tcacacaaca gctgaatcaa
3901 agtgaattgg cagtgttatt aaacctgctg cagagccaaa ccgacctgag catccctcaa
3961 atggcacagc tgcttaacat ccactccaac ccagagatgc agcagcagct ggaagccctg
4021 aaccaatcca tcagtgccct gacggaagct acttcccagc agcaggactc agagaccatg
4081 gccccagagg agtctttgaa ggaagcaccc tctgccccag tgatcctgcc ttcagcagaa
4141 cagacgaccc ttgaagcttc aagcacacca gctgacatgc agaatatatt ggcagttctc
4201 ttgagtcagc tgatgaaaac ccaagagcca gcaggcagtc tggaggaaaa caacagtgac
4261 aagaacagtg ggccacaggg gccccgaaga actcccacaa tgccacagga ggaggcagca
```

4321 gagaagaggc cccctgagcc ccccggacct ccaccgccgc cacctccacc ccctctggtt
4381 gaaggcgatc tttccagcgc cccccaggag ttgaacccag ccgtgacagc cgccttgctg
4441 caacttttat cccagcctga agcagagcct cctggccacc tgccacatga gcaccaggcc
4501 ttgagaccaa tggagtactc cacccgaccc cgtccaaaca ggacttatgg aaacactgat
4561 gggcctgaaa cagggttcag tgccattgac actgatgaac gaaactctgg tccagccttg
4621 acagaatcct tggtccagac cctggtgaag aacaggacct tctcaggctc tctgagccac
4681 cttggggagt ccagcagtta ccagggcaca gggtcagtgc agtttccagg ggaccaggac
4741 ctccgttttg ccagggtccc cttagcgtta cacccggtgg tcgggcaacc attcctgaag
4801 gctgagggaa gcagcaattc tgtggtacat gcagagacca aattgcaaaa ctatggggag
4861 ctggggccag gaaccactgg ggccagcagc tcaggagcag gccttcactg gggggccca
4921 actcagtctt ctgcttatgg aaaactctat cgggggccta caagagtccc accaagaggg
4981 ggaagaggga gaggagttcc ttactaaccc agagacttca gtgtcctgaa agattccttt
5041 cctatccatc cttccatcca gttctctgaa tctttaatga aatcatttgc cagagcgagg
5101 taatcatctg catttggcta ctgcaaagct gtccgttgta ttccttgctc acttgctact
5161 agcaggcgac ttacgaaata atgatgttgg caccagttcc ccctggatgg gctatagcca
5221 gaacatttac ttcaactcta ccttagtaga tacaagtaga gaatatggag aggatcatta
5281 cattgaaaag taaatgtttt attagttcat tgcctgcact tactgatcgg aagagagaaa
5341 gaacagtttc agtattgaga tggctcagga gaggctcttt gatttttaaa gttttggggt
5401 gggggattgt gtgtggtttc tttcttttga attttaattt aggtgttttg ggtttttttc
5461 ctttaaagag aatagtgttc acaaaatttg agctgctctt tggcttttgc tataagggaa
5521 acagagtggc ctggctgatt tgaataaatg tttctttcct ctccaccatc tcacattttg
5581 cttttaagtg aacacttttt ccccattgag catcttgaac atacttttttt tccaaataaa
5641 ttactcatcc ttaaagttta ctccactttg acaaaagata cgcccttctc cctgcacata
5701 aagcaggttg tagaacgtgg cattcttggg caagtaggta gactttaccc agtctctttc
5761 ctttttgct gatgtgtgct ctctctctct ctttctctct ctctctctct ctctctctct
5821 ctctctctct ctctgtctcg cttgctcgct ctcgctgttt ctctctcttt gaggcatttg
5881 tttggaaaaa atcgttgaga tgcccaagaa cctgggataa ttctttactt tttttgaaat
5941 aaaggaaagg aaattcagac tcttacattg ttctctgtaa ctcttcaatt ctaaaatgtt
6001 ttgtttttta aaccatgttc tgatggggaa gttgatttgt aagtgtggac agcttggaca
6061 ttgctgctga gctgtggtta gagatgatgc ctccattcct agagggctaa taacagcatt
6121 tagcatattg tttacacata tatttttatg tcaaaaaaaa aacaaaaacc tttcaaacag
6181 agcattgtga tattgtcaaa gagaaaaaca aatcctgaag atacatggaa atgtaaccta
6241 gtttagggtg ggtatttttc tgaagataca tcaatacctg acctttttta aaaaaataat
6301 tttaaaacag catactgtga ggaagaacag tattgacata cccacatccc agcatgtgta
6361 ccctgccagt tcttttaggg atttttcctc caaagagatt tggatttggt tttggtaaaa
6421 ggggttaaat tgtgcttcca ggcaagaact ttgccttatc ataaacagga aatgaaaaag
6481 ggaagggctg tcaggatggg ataatttggg aggcttctca ttctggcttc tatttctatg
6541 tgagtaccag catatagagt gttttaaaaa cagatacatg tcatataatt tatctgcaca
6601 gacttagacc ttcaggaaac ataggttaag ccccctttta caaagaaaaa gtaaacatac
6661 ttcagcatct tggagggtag ttttcaaaac tcaagtttca tgtttcaatg ccaagttctt
6721 attttaaaaa ataaaatcta cttataagag aaaggtgcat tacttaaaaa aaaaaaactt
6781 taaagaaatg aaagaagaac cctcttcaga tacttacttg aagactgttt tcccctgtta
6841 atgagatata gctagatatc ggtgtgtgta tttctttatt attctctggt ttttgatctg
6901 gccttgcctc cagggccaaa cactgattta gaaagagagc cttctagcta ttttggcatt
6961 gatggctttt tataccagtg tgtccagtta gatttactag gcttactgac atgctattgg
7021 taaatcgcat taaagttcat ctgaaccttc tgtctgttga cttcttagtc ctcagacatg
7081 ggcctttgtg ttttagaata tttgaatttg agttattggg ccccactccc tgtttttttat
7141 taaagaacgt gagcctggga tactttcaga agtatctgtt caatgaaaaa aagttggttt
7201 cccatcaaat atgaataaaa ttctctatat atttcattgt attttggtta tcagcagtca
7261 tcaataatgt ttttccctcc cctctcccac ctcttatttt taattatgcc aaatatccta
7321 aataatatac ttaagcctcc attccctcat ccctactagg gaaggggtg agtgtatgtg
7381 tgagtgtatg tgtatgtatg atcccatctc acccccaccc ccattttggg agtcttttaa
7441 aatgaaaaca aagtttggta gttttgacta tttctaaaag cagaggagaa aaaaaaactt
7501 atttaaatat cctggaatct gtatggagga agaaaaggta tttgttaatt tttcagttac
7561 gttatctata aacatgatgg aagtaaaggt ttggcagaat ttcaccttga ctatttgaaa
7621 attacagacc caattaattc cattcaaaag tggttttcgt tttgttttaa ttattgtaca

```
7681 atgagagata ttgtctatta aatacattat tttgaacaga tgagaaatct gattctgttc
7741 atgagtggga ggcaaaactg gtttgaccgt gatcattttt gtggttttga aaacaaatat
7801 acttgaccca gtttccttag ttttttcttc aactgtccat aggaacgata agtatttgaa
7861 agcaacatca aatctatacg tttaaagcag ggcagttagc acaaatttgc aagtagaact
7921 tctattagct tatgccatag acatcaccca accacttgta tgtgtgtgtg tatatataat
7981 atgcatatat agttaccgtg ctaaaatggt taccagcagg ttttgagaga gaatgctgca
8041 tcagaaaagt gtcagttgcc acctcattct ccctgattta ggttcctgac actgattcct
8101 ttctctctcg tttttgaccc ccattgggtg tatcttgtct atgtacagat attttgtaat
8161 atattaaatt tttttctttc agtttataaa aatggaaagt ggagattgga aaattaaata
8221 tttcctgtta ctataccact tttgctccat tgcatt
```

By "Cyclin Dependent Kinase 12 (CDK12)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_057591, version NP_057591.2, incorporated herein by reference, as reproduced below (SEQ ID NO: 37):

```
   1 mpnserhggk kdgsggasgt lqpssgggss nsrerhrlvs khkrhkskhs kdmglvtpea
  61 aslgtvikpl veyddissds dtfsddmafk ldrrenderr gsdrsdrlhk hrhhqhrrsr
 121 dllkakqtek eksqevssks gsmkdrisgs skrsneetdd ygkaqvakss skesrssklh
 181 kektrkerel ksghkdrsks hrkretpksy ktvdspkrrs rsphrkwsds skqddspsga
 241 sygqdydlsp srshtssnyd sykkspgsts rrqsvsppyk epsayqsstr spspysrrqr
 301 svspysrrrs ssyersgsys grspspygrr rssspflskr slsrsplpsr ksmksrsrsp
 361 aysrhssshs kkkrsssrsr hssispvrlp lnsslgaels rkkkeraaaa aaakmdgkes
 421 kgspvflprk enssveakds gleskklprs vkleksapdt elvnvthlnt evknssdtgk
 481 vkldensekh lvkdlkaqgt rdskpialke eivtpketet seketppplp tiaspppplp
 541 tttpppqtpp lpplppipal pqqpplppsq pafsqvpass tstlppsths ktsavssqan
 601 sqppvqvsvk tqvsvtaaip hlktstlppl plppllpgdd dmdspketlp skpvkkekeq
 661 rtrhlltdlp lppelpggdl sppdspepka itppqqpykk rpkiccpryg errqtesdwg
 721 krcvdkfdii giigegtygq vykakdkdtg elvalkkvrl dnekegfpit aireikilrq
 781 lihrsvvnmk eivtdkqdal dfkkdkgafy lvfeymdhdl mgllesglvh fsedhiksfm
 841 kqlmegleyc hkknflhrdi kcsnillnns gqikladfgl arlynseesr pytnkvitlw
 901 yrppelllge erytpaidvw scgcilgelf tkkpifqanl elaqlelisr lcgspcpavw
 961 pdviklpyfn tmkpkkqyrr rlreefsfip saaldlldhm ltldpskrct aeqtlqsdfl
1021 kdvelskmap pdlphwqdch elwskkrrrq rqsgvvveep ppsktsrket tsgtstepvk
1081 nsspappqpa pgkvesgagd aigladitqq lnqselavll nllqsqtdls ipqmaqllni
1141 hsnpemqqql ealnqsisal teatsqqqds etmapeeslk eapsapvilp saeqttleas
1201 stpadmqnil avllsqlmkt qepagsleen nsdknsgpqg prrtptmpqe eaaacpphil
1261 ppekrppepp gpppppppppp lvegdlssap qelnpavtaa llqllsqpea eppghlpheh
1321 qalrpmeyst rprpnrtygn tdgpetgfsa idtdernsgp alteslvqtl vknrtfsgsl
1381 shlgesssyq gtgsvqfpgd qdlrfarvpl alhpvvgqpf lkaegssnsv vhaetklqny
1441 gelgpgttga sssgaglhwg gptqssaygk lyrgptrvpp rggrgrgvpy
```

By "AKT3 nucleic acid molecule" is meant a polynucleotide encoding a AKT3 polypeptide. An exemplary AKT3 nucleic acid molecule is provided at NCBI Accession No. NM_005465, version NM_005465.4, incorporated herein by reference, and reproduced below (SEQ ID NO: 38):

```
   1 gctgagtcat cactagagag tgggaagggc agcagcagca gagaatccaa accctaaagc
  61 tgatatcaca aagtaccatt tctccaagtt gggggctcag aggggagtca tcatgagcga
 121 tgttaccatt gtgaaagaag gttgggttca gaagagggga gaatatataa aaaactggag
```

```
 181 gccaagatac ttccttttga agacagatgg ctcattcata ggatataaag agaaacctca
 241 agatgtggat ttaccttatc ccctcaacaa cttttcagtg gcaaaatgcc agttaatgaa
 301 aacagaacga ccaaagccaa acacatttat aatcagatgt ctccagtgga ctactgttat
 361 agagagaaca tttcatgtag atactccaga ggaaagggaa gaatggacag aagctatcca
 421 ggctgtagca gacagactgc agaggcaaga agaggagaga atgaattgta gtccaacttc
 481 acaaattgat aatataggag aggaagagat ggatgcctct acaacccatc ataaaagaaa
 541 gacaatgaat gattttgact atttgaaact actaggtaaa ggcacttttg ggaaagttat
 601 tttggttcga gagaaggcaa gtggaaaata ctatgctatg aagattctga agaaagaagt
 661 cattattgca aaggatgaag tggcacacac tctaactgaa agcagagtat taaagaacac
 721 tagacatccc tttttaacat ccttgaaata ttccttccag acaaaagacc gtttgtgttt
 781 tgtgatggaa tatgttaatg ggggcgagct gtttttccat ttgtcgagag agcgggtgtt
 841 ctctgaggac cgcacacgtt tctatggtgc agaaattgtc tctgccttgg actatctaca
 901 ttccggaaag attgtgtacc gtgatctcaa gttggagaat ctaatgctgg acaaagatgg
 961 ccacataaaa attacagatt ttggactttg caaagaaggg atcacagatg cagccaccat
1021 gaagacattc tgtggcactc cagaatatct ggcaccagag gtgttagaag ataatgacta
1081 tggccgagca gtagactggt ggggcctagg ggttgtcatg tatgaaatga tgtgtgggag
1141 gttacctttc tacaaccagg accatgagaa acttttttgaa ttaatattaa tggaagacat
1201 taaatttcct cgaacactct cttcagatgc aaaatcattg ctttcagggc tcttgataaa
1261 ggatccaaat aaacgccttg gtggaggacc agatgatgca aaagaaatta tgagacacag
1321 tttcttctct ggagtaaact ggcaagatgt atatgataaa aagcttgtac ctccttttaa
1381 acctcaagta acatctgaga cagatactag atattttgat gaagaattta cagctcagac
1441 tattacaata acaccacctg aaaaatatga tgaggatggt atggactgca tggacaatga
1501 gaggcggccg catttccctc aattttccta ctctgcaagt ggacgagaat aagtctcttt
1561 cattctgcta cttcactgtc atcttcaatt tattactgaa aatgattcct ggacatcacc
1621 agtcctagct cttacacata gcaggggcac cttccgacat cccagaccag ccaagggtcc
1681 tcacccctcg ccacctttca ccctcatgaa aacacacata cacgcaaata cactccagtt
1741 tttgtttttg catgaaattg tatctcagtc taaggtctca tgctgttgct gctactgtct
1801 tactattata gcaactttaa gaagtaattt tccaaccttt ggaagtcatg agcccaccat
1861 tgttcatttg tgcaccaatt atcatctttt gatcttttag tttttccctc agtgaaggct
1921 aaatgagata cactgattct aggtacattt tttaactttc tagaagagaa aaactaacta
1981 gactaagaag atttagttta taaattcaga acaagcaatt gtggaagggt ggtggcgtgc
2041 atatgtaaag cacatcagat ccgtgcgtga agtaggcata tatcactaag ctgtggctgg
2101 aattgattag gaagcatttg gtagaaggac tgaacaactg ttgggatata tatatatata
2161 tataattttt ttttttaaa ttcctggtgg atactgtaga agaagcccat atcacatgtg
2221 gatgtcgaga cttcacgggc aatcatgagc aagtgaacac tgttctacca agaactgaag
2281 gcatatgcac agtcaaggtc acttaaaggg tcttatgaaa caatttgagc cagagagcat
2341 ctttcccctg tgcttggaaa cctttttttcc ttcttgacat ttatcacctc tgatggctga
2401 agaatgtaga caggtataat gatactgctt ttcaccaaaa tttctacacc aaggtaaaca
2461 ggtgtttgcc ttatttaatt ttttactttc agttctacgt gaattagctt tttctcagat
2521 gttgaaactt tgaatgtcct tttatgattt tgtttatatt gcagtagtat ttatttttta
2581 gtgatgagaa ttgtatgtca tgttagcaaa cgcagctcca acttatataa aatagactta
2641 ctgcagttac ttttgaccca tgtgcaagga ttgtacacgc tgatgagaat catgcacttt
2701 ttctcctctg ttaaaaaaaa tgataaggct ctgaaatgga atatattggt tagaatttgg
2761 ctttgggaga agagatgctg ccatttaacc ccttggtact gaaaatgaga aaatccccaa
2821 ctatgcatgc caaggggtta atgaaacaaa tagctgttga cgtttgctca tttaagaatt
2881 tgaaacgtta tgatgacctg gcaacaaaaa gtaatgaaga aaattgagac ctgagtgaag
2941 ataagaaatg atctttacgt ggcaaaatga acacatcttg agtatttagg aaatgggcag
3001 tgaaggctaa gaacctggtg tgtttcttgg gatcatggta catttatcac tgaattaagc
3061 catcagggaa aaaacaacaa aaaaagagaa cacctccagc ttttcttttt ctgtatatac
3121 tcatgtcccc cagattccaa catttctcac tgaaaggggg catgtatgca aacctcatct
3181 ttctccttca ttaatgatga tcttcagatt aaaccctttg gtgctaggag ctgacaattt
3241 ccaaagcagc ctgtgaagtc ctaggggctg ggggccactc ttgcggcaag cagaaggcca
3301 tcctactccg cggagtgatc atggaaatgt attttagtta aactctgaca gctcccaaac
3361 ggaagactac agcatgacgt agtattatga ttgcattgta tgaaagagca agtgactttc
3421 taagtaggat gaatcatatt catatgcaga tgtcttagcc tcttgacgct ggaagtgtgg
3481 atttatagct atgaaaccac tgctggcagt gggtgggcca ctgggactga cgggggttaa
```

3541 agggcatttt actaaggcag ctaagacata ttcagacatc aacgttatcc ttctttttca
3601 tatttctacc tgagtgaagt tcatccttag tattgagtag gaagttacag taaatggtag
3661 ttcattctta cttacacaca tagctaatct ttttttttc acttggaatt atgttgaatg
3721 tttcattttg acaaaaaagt agactagaag gtatgttctt taagttgtct tgcatccatt
3781 atataagaaa gaaacaggtg agaggaagag cagaaagctg agactggctg atgttcagag
3841 cacttactcc tctagaggga aagcatgaca ccgaacacta agcacacagc tttttgttgt
3901 tttggttttt tctcccgcaa atcttaaagt gattcccatg accttggcca aggacacttc
3961 ttaaagatta atgactggca ctgacattgc cccaggcggg ccactcctca cactggctct
4021 cagttcccag ccatgcctgg ggctcagtca cttctattcc accctctgag actccattgg
4081 tgtcacacaa ggtgtcttct tggctttgat tttgagaatc ccctattttc acttccagat
4141 ctgtcagctg ccatggagga ataatagaaa accagaaatg cgtgtagagg gagatttcta
4201 aaacttccct tgtgtcgccc atagttgtag ttttgggttc tggcaggtgg aacaccctga
4261 aacctggaat cattctatga gaatacagtt cagactttgc agactccagc ccatactaac
4321 tgtcatgaag cttgacttct tgtcataatg cagccatctt ggaggaaatt ggccatttct
4381 gcttagatgg ttggcagggt cgcgctcagc tttgctttct acactaatta catagcatta
4441 ttcaagtatt gttttccatt tcccatccct gatttccagc ttcttaaagc tgactgttct
4501 tgcaggggcc acttgcttct cctagagtac aaaagtaagg gccttcctta ctaactgcag
4561 ggtctctcta ttacacctca acatacacac tttgctgcta ctgtttgtac tgtctacagt
4621 agaatttcct tatcttgctc ctggtagtgc attacaggca agcatgaaat gtaaagtatt
4681 tatttaaata aaaagaaaac ctctaaattg gtaattgaat tacctccctg tagctttata
4741 gtttgtgaca tttcttgacc ttgctagttc tttcattaga tctgcgcaag atctagtcat
4801 ctggttaagg attttaagca gatgcaacta taaacccaag aaactgtatt actattactg
4861 ttggtcatac taaacctgtc tatttcctga agtatatgac ccacaaggat gtggaataac
4921 taggagaaac tgtttttgta cactgtacat ccttagtatt tttacacgta tatgataggg
4981 atgaacatga ttttccttcg tacagacagc ttaaataaag cactatgtca atctgctact
5041 tctctgttta ttgttgttgg atgtggttct ataatccccc caaattaaat cttctttaat
5101 gaaaacatga tttttaatag ccccagctgg tattaaccta ccttgtataa aatgtgacag
5161 gaaaatatag aaataattcc ttgtagctca cacacacaca cataggggat catttttact
5221 tcagtgaaat ggcagtagtg cggttgtgca aactttgatg aacggctgct tctgagggga
5281 aacgctgacc tctcagcact ggatttagga tggatgtact gtgaagccag ggatgaagga
5341 ggtctcagac cctggggaca ttcagacccg aatcatctat acaacacacg gtttggaccc
5401 agaatctgaa ggaatgtagc ttttcattaa cgtcttcctg ataatgtact gctctgcata
5461 tttccttttct tagagtgtat ttctaacaac atgtcatggc aaattaacaa acttagacgt
5521 gggtgatgta gatgggtagg atggctggac tgcagtctga cttcacgttg aatcattctg
5581 gatggggcct ttttctgatt ttacctcata aagctactat tgtagaaact tggctttgct
5641 cctgtgacga agccagacag aggaatggct tttgggacca gagtgagtca agcatgtatg
5701 tgtatgtcac acggccaaat ttgagggcat tctcacatgt gctcttctct caaaaccact
5761 ggggttgaca gatccaggag gctaaaaaaa agtgacctct ataattcttt aaaggtgcta
5821 tttttagaat attgtataat ttattcacag tatatctaaa acagaattaa ggacaattaa
5881 aatatcttat gtgacagcct ttatgtctag cacatttgat gaaataaaaa acttctgaat
5941 ctgaatagaa gttctactgt ttcaggcttg aacctttttac atgctcaaga gattcaaatg
6001 gtctctgtgt gtagatcatg ccaccgcctc caaagcctaa tccacatcac ttctgagagg
6061 caaggctgag catatggtga catcagctct gtgttgagat ggtgatgagg atgatggctc
6121 gctggccagg cagggcagcc gaaggtcagg gacctgtcct aactaactgc agccttgcct
6181 ttagtgtttg tcattctcag atacaacacg gtatgtccag tgtccgtttt tattacttta
6241 aagcatttga gggcttaatt gtgtatagta gaaatactat tttagacaaa taattatctg
6301 tgtacagata tttgatatac tctaagtaaa ttttctaatt tcactaagta cgtttttagg
6361 ctcctctcaa atactgcgta ttgaagaaaa aaatctgaca ccaccgagcc aaagatgctt
6421 ttttgtctgt tttcgttgtt taacagaatg gaaagagtaa tgcatagtgc ttcctggtgt
6481 ctcctgattg attgattgtg cacaaagtag gacgataaat aaataaaatg gagtctgatg
6541 ggacattgat taaaggtgaa ggatgattga tatatagatc atgaaaagaa aaatgaatgg
6601 caggaaaaaa agtttggtcc ttaatatact ttggcctagt taaaatatgt gccttttttgg
6661 tgtgttttgt tcatcactac aagataaaaa ggaaacatta caactcaagt ctttaaaaag
6721 ttcatttatt gaaaatcata tgtataacct agcatacgaa tgagcagatt taaacacata
6781 acttcaagcc atttctgaaa acatacacca ggagctctgc tcagctagag tcagactcca
6841 gctccagccc gactgcgtgc ggggacagcg cccgcgttga tgaggaccag ccccactgca

```
6901 ggctgaggcg gtgtcaccct gggaaggtcg tggtgcgttg tggcatatta agtctaaacc
6961 agatgaatgt aaatatctct ttgtaaatca tttatttcac tctgttccat ccaggtcagc
7021 aatcagattg tggcatgctg ggtaactgga aaaaataata aaaagtaagt ttcaatagct
7081 caaaaaaaaa a
```

By "AKT3" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. CAB53537, version CAB53537.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 39):

```
   1 msdvtivkeg wvqkrgeyik nwrpryfllk tdgsfigyke kpqdvdlpyp lnnfsvakcq
  61 lmkterpkpn tfiirclqwt tviertfhvd tpeereewte aiqavadrlq rqeeermncs
 121 ptsqidnige eemdastthh krktmndfdy lkllgkgtfg kvilvrekas gkyyamkilk
 181 keviiakdev ahtltesrvl kntrhpflts lkysfqtkdr lcfvmeyvng gelffhlsre
 241 rvfsedrtrf ygaeivsald ylhsgkivyr dlklenlmld kdghikitdf glckegitda
 301 atmktfcgtp eylapevled ndygravdww glgvvmyemm cgrlpfynqd heklfelilm
 361 edikfprtls sdaksllsgl likdpnkrlg ggpddakeim rhsffsgvnw qdvydkklvp
 421 pfkpqvtset dtryfdeeft aqtititppe kydedgmdcm dnerrphfpq fsysasgre
```

By "Tyrosine-Protein Kinase Receptor 3 (TYRO3) nucleic acid molecule" is meant a polynucleotide encoding a TYRO3 polypeptide. An exemplary TYRO3 nucleic acid molecule is provided at NCBI Accession No. X72886, version X72886.1, incorporated herein by reference, and reproduced below (SEQ ID NO: 40):

```
   1 accctgggcc ggatgttggg caaaggagag tttggttcag tgcgggaggc ccagctgaag
  61 caagaggatg gctcctttgt gaaagtggct gtgaagatgc tgaaagctga catcattgcc
 121 tcaagcgaca ttgaagagtt cctcagggaa gcagcttgca tgaaggagtt tgaccatcca
 181 cacgtggcca aacttgttgg ggtaagcctc cggagcaggg ctaaaggccg tctccccatc
 241 cccatggtca tcttgccctt catgaagcat ggggacctgc atgccttcct gctcgcctcc
 301 cggattgggg agaacccctt taacctaccc ctccagaccc tgatccggtt catggtggac
 361 attgcctgcg gcatggagta cctgagctct cggaacttca tccaccgaga cctggctgct
 421 cggaattgca tgctggcaga ggacatgaca gtgtgtgtgg ctgacttcgg actctcccgg
 481 aagatctaca gtggggacta ctatcgtcaa ggctgtgcct ccaaactgcc tgtcaagtgg
 541 ctggccctgg agagcctggc cgacaacctg tatactgtgc agagtgacgt gtgggcgttc
 601 ggggtgacca tgtgggagat catgacacgt gggcagacgc catatgctgg catcgaaaac
 661 gctgagattt acaactacct cattggcggg aaccgcctga aacagcctcc ggagtgtatg
 721 gaggacgtgt atgatctcat gtaccagtgc tggagtgctg accccaagca gcgcccgagc
 781 tttacttgtc tgcgaatgga actggagaac atcttg
```

By "Tyrosine-Protein Kinase Receptor 3 (TYRO3)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAH51756, version AAH51756.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 41):

```
   1 malrrsmgrp glpplplppp prlglllaal aslllpesaa aglklmgapv kltvsqgqpv
```

```
  61 klncsvegme epdiqwvkdg avvqnldqly ipvseqhwig flslksvers dagrywcqve
 121 dggeteisqp vwltvegvpf ftvepkdlav ppnapfqlsc eavgppepvt ivwwrgttki
 181 ggpapspsvl nvtgvtqstm fsceahnlkg lassrtatvh lqalpaapfn itvtklsssn
 241 asvawmpgad grallqsctv qvtqapggwe vlavvvpvpp ftcllrdlvp atnyslrvrc
 301 analgpspya dwvpfqtkgl apasapqnlh airtdsglil eweevipeap legplgpykl
 361 swvqdngtqd eltvegtran ltgwdpqkdl ivrvcvsnav gcgpwsqplv vsshdragqq
 421 gpphsrtswv pvvlgvltal vtaaalalil lrkrrketrf gqafdsvmar gepavhfraa
 481 rsfnrerper ieatldslgi sdelkekled vlipeqqftl grmlgkgefg svreaqlkqe
 541 dgsfvkvavk mlkadiiass dieeflreaa cmkefdhphv aklvgvslrs rakgrlpipm
 601 vilpfmkhgd lhafllasri genpfnlplq tlirfmvdia cgmeylssrn fihrdlaarn
 661 cmlaedmtvc vadfglsrki ysgdyyrqgc asklpvkwla lesladnlyt vqsdvwafgv
 721 tmweimtrgq tpyagienae iynyliggnr lkqppecmed vydlmyqcws adpkqrpsft
 781 clrmelenil gqlsvlsasq dplyiniera eeptaggsle lpgrdqpysg agdgsgmgav
 841 ggtpsdcryi ltpgglaeqp gqaehqpesp lnetqrllll qqgllphssc
```

By "Ephrin Type-A Receptor 5 (EPHA5) nucleic acid molecule" is meant a polynucleotide encoding a EPHA5 polypeptide. An exemplary EPHA5 nucleic acid molecule is provided at NCBI Accession No. NM_004439, version NM_004439.7, incorporated herein by reference, and reproduced below (SEQ ID NO: 42):

```
   1 ccagacacag caggagcgcg ccttggcggt gcagtttcaa cctcgacttc gcagccgcgc
  61 acacaccgcc tgctccccga gcagcgggaa ccgcagcagc tcctgggccg ccgcagcgca
 121 gctccgcgct cctaccggcc ggcagccgtc agtccctccc ctcttcagca ctcagcccgc
 181 agctatttcc ttctgccagt ctctttgaac tctggatctt tgcttttgct cgctgctctc
 241 ctgtttttca ttctccacat tttctcaagt cctctttctt tatccttagc caccctgctt
 301 ttttcctcct tttttaaaaa atcggagatt tcgtcttaaa atgatttgcc ttccttacct
 361 tcgtccattt caacactgaa ggctgcaaag aaccttcacc tttcccctag tggtatttaa
 421 aaattctcaa tccgtaaaaa gtctttttga aaggcaaagg aacaggaccc aggaccctct
 481 cgacaccctt gatccgagtc cagatctgca ctagcaacca gaactaatat ttcatttaac
 541 cccaccaaag ggggaggcga gaggagccag aagcaaactt cagctgtctc agccggatcc
 601 gtggttccta catttggagg agccgcgtgc cagaaggcgt aggaccccaa ggggggacaa
 661 ggaggactcc cgagtctccc ttcgccgctc tgcgaggccg aagcggtgga ctgagccgct
 721 cgggacagcg gcaccggagg aggctcggag aagatgcggg gctcggggcc ccggggtgcg
 781 ggacgccggc ggcccccaag cggcggcggc gacaccccca tcaccccagc gtccctggcc
 841 ggctgctact ctgcacctcg acgggctccc ctctggacgt gccttctcct gtgcgccgca
 901 ctccggaccc tcctggccag ccccagcaac gaagtgaatt tattggattc acgcactgtc
 961 atgggggacc tgggatggat tgcttttcca aaaaatgggt gggaagagat tggtgaagtg
1021 gatgaaaatt atgcccctat ccacacatac caagtatgca aagtgatgga acagaatcag
1081 aataactggc ttttgaccag ttggatctcc aatgaaggtg cttccagaat cttcatagaa
1141 ctcaaattta ccctgcggga ctgcaacagc cttcctggag gactggggac ctgtaaggaa
1201 acctttaata tgtattactt tgagtcagat gatcagaatg ggagaaacat caaggaaaac
1261 caatacatca aaattgatac cattgctgcc gatgaaagct ttacagaact tgatcttggt
1321 gaccgtgtta tgaaactgaa tacagaggtc agagatgtag gacctctaag caaaaaggga
1381 ttttatcttg cttttcaaga tgttggtgct tgcattgctc tggtttctgt gcgtgtatac
1441 tataaaaaat gcccttctgt ggtacgacac ttggctgtct tccctgacac catcactgga
1501 gctgattctt cccaattgct cgaagtgtca ggctcctgtg tcaaccattc tgtgaccgat
1561 gaacctccca aaatgcactg cagcgccgaa ggggagtggc tggtgcccat cgggaaatgc
1621 atgtgcaagg caggatatga agagaaaaat ggcacctgtc aagtgtgcag acctgggttc
1681 ttcaaagcct cacctcacat ccagagctgc ggcaaatgtc cacctcacag ttatacccat
1741 gaggaagctt caacctcttg tgtctgtgaa aaggattatt tcaggagaga gtctgatcca
1801 cccacaatgg catgcacaag acccccctct gctcctcgga atgccatctc aaatgttaat
1861 gaaactagtg tctttctgga atggattccg cctgctgaca ctggtggaag gaaagacgtg
```

1921 tcatattata ttgcatgcaa gaagtgcaac tcccatgcag gtgtgtgtga ggagtgtggc
1981 ggtcatgtca ggtaccttcc ccggcaaagc ggcctgaaaa acacctctgt catgatggtg
2041 gatctactcg ctcacacaaa ctataccttt gagattgagg cagtgaatgg agtgtccgac
2101 ttgagcccag gagcccggca gtatgtgtct gtaaatgtaa ccacaaatca agcagctcca
2161 tctccagtca ccaatgtgaa aaaagggaaa attgcaaaaa acagcatctc tttgtcttgg
2221 caagaaccag atcgtcccaa tggaatcatc ctagagtatg aaatcaagta ttttgaaaag
2281 gaccaagaga ccagctacac gattatcaaa tctaaagaga caactattac tgcagagggc
2341 ttgaaaccag cttcagttta tgtcttccaa attcgagcac gtacagcagc aggctatggt
2401 gtcttcagtc gaagatttga gtttgaaacc accccagtgt ttgcagcatc cagcgatcaa
2461 agccagattc ctgtaattgc tgtgtctgtg acagtgggag tcattttgtt ggcagtggtt
2521 atcggcgtcc tcctcagtgg aagttgctgc gaatgtggct gtgggagggc ttcttccctg
2581 tgcgctgttg cccatccaag cctaatatgg cggtgtggct acagcaaagc aaaacaagat
2641 ccagaagagg aaaagatgca ttttcataat gggcacatta aactgccagg agtaagaact
2701 tacattgatc cacataccta tgaggatccc aatcaagctg tccacgaatt tgctaaggag
2761 atagaagcat catgtatcac cattgagaga gttattggag caggtgaatt tggtgaagtt
2821 tgtagtggac gtttgaaact accaggaaaa agagaattac ctgtggctat caaaaccctt
2881 aaagtaggct atactgaaaa gcaacgcaga gatttcctag gtgaagcaag tatcatggga
2941 cagtttgatc atcctaacat catccattta gaaggtgtgg tgaccaaaag taaaccagtg
3001 atgatcgtga cagagtatat ggagaatggc tctttagata catttttgaa gaaaaacgat
3061 gggcagttca ctgtgattca gcttgttggc atgctgagag gtatctctgc aggaatgaag
3121 tacctttctg acatgggcta tgtgcataga gatcttgctg ccagaaacat cttaatcaac
3181 agtaaccttg tgtgcaaagt gtctgacttt ggactttccc gggtactgga agatgatccc
3241 gaggcagcct acaccacaag gggaggaaaa attccaatca gatggactgc cccagaagca
3301 atagctttcc gaaagtttac ttctgccagt gatgtctgga gttatggaat agtaatgtgg
3361 gaagttgtgt cttatggaga gagaccctac tgggagatga ccaatcaaga tgtgattaaa
3421 gcggtagagg aaggctatcg tctgccaagc cccatggatt gtcctgctgc tctctatcag
3481 ttaatgctgg attgctggca gaaagagcga aatagcaggc ccaagtttga tgaaatagtc
3541 aacatgttgg acaagctgat acgtaaccca agtagtctga agacgctggt taatgcatcc
3601 tgcagagtat ctaatttatt ggcagaacat agcccactag gatctggggc ctacagatca
3661 gtaggtgaat ggctagaggc aatcaagatg ggccggtata cagagatttt catggaaaat
3721 ggatacagtt caatggacgc tgtggctcag gtgaccttgg aggatttgag acggcttgga
3781 gtgactcttg tcggtcacca gaagaagatc atgaacagcc ttcaagaaat gaaggtgcag
3841 ctggtaaacg gaatggtgcc attgtaactt catgtaaatg tcgcttcttc aagtgaatga
3901 ttctgcactt tgtaaacagc actgagattt attttaacaa aaaaaggggg aaaagggaaa
3961 acagtgattt ctaaacctta gaaaacattt gcctcagcca cagaatttgt aatcatggtt
4021 ttactgaagt atccagttct tagtccttag tcttcatttt tcatgaagca aacatatctt
4081 gcattaaaag ggacatgaag ttagacatca tcttaagtta caacaacaga atccttccca
4141 ctacttctac aaaatttttgt acatgaaata tataattata tagcactttt atagactgaa
4201 ttaaggcaac ccctttcaaa acttccaggg atctacttga aaggaaatgt tttatagcca
4261 tttgtgagct aacaaaagct acagtttact gaagtttact tcaagtctta attgtctaca
4321 aaagtgtatt gaagagcaat atgattagat tatttcttaa tagatatctt cttttgtaat
4381 tttaaaatgc tgttacacag cgttaagtta tagaaactag tgtgtaaaca tgttgcttga
4441 tcaagaaaaa gtacaataca gggtgtatat ttatttttttg tgttataaag tttacttttta
4501 gttgctcttc tagagattat taggtaataa atgtgtatat actgtataat ttgcaatata
4561 cccaggaact gatttaagat ggaattgtgt gtgtgtttgc ttgcacatgt gtgtgttacc
4621 attctgcttg catttctaat agtgtttcaa ttttagcaac atataggagc aagtgttcca
4681 gaatgtaata tgaataggag aaatagggaa gcagtaaaac aaaatttaac acaagcttgt
4741 gtctcttttc ctctcatgtg tccaaaagct aatctcttta ttcactaaaa ggaaatgtgt
4801 ataagactaa atccccctttg gctttttaaa aacattttgt gatatcagtg acaatgcagt
4861 tcttctagcc attaatcttg taccccctgta agaaatttca cctttctgag tcctgaaaag
4921 tatcttgtca agcaaagttg acaccgaagg gcacattttc agcaggatgt agaaggattt
4981 aactgtgcag gcttctgtta atgttgttaa atccaggcac atagcacgaa gcatcaccct
5041 taagtgttaa tccgttgtaa ccattcccat tttgactcag ttctagaaat tttgactcta
5101 aggcagcaat ggaattatga caaatatata tatatacaca tacacacaca caaatatata
5161 tatatgctta tttcccttca gaattttatt tcaataattg ataagtttta tttttaatgg
5221 atgtttattc atttgttaat ttcagtttgt attcagttac atggctttga gtttatttta

5281 ttgaaccaat ccaaggtatt atgtaattag gcttattaag gaatataaca tatcttctat
5341 gtatgctcta tataccacac atgatgaaga cttatatttg tgatacaaat tatgcagtgt
5401 attagaaatt gtccaaaact gagctcctga gataaccaac atttttgtat tattacactt
5461 tataaaaatt atggaatgtc atgtgattag tcaaaatcag agcatttaaa taaaatacag
5521 cctactcatg attatcaatg ttaatagtac agtagatcag agaagcctgg attaatgaaa
5581 tttatcagaa atatagccac gttatattaa tggtattctt ttttccaagc acttgtttta
5641 tttgaatatt agcacatata agaataaact tttgggaaat aactacttcc tattctgccc
5701 acccttttact taagcagtaa agtacctctt ttctaaagaa ccaggaaaca tcaaaataac
5761 cacaacaaaa atctaaacag ttttcaaata atacattcag ttctcataat aataagggct
5821 atataccatg ggtccacaca aactgtgctg caaacatata atctattagt ttagtaaaat
5881 ttccagagat gtgaatttaa tattttcttg ataaaatcat aaaaaagcat caccattatt
5941 aaagatgcat ttgttcattt caataaacag aaattaatga aacaaattac ttttgtacaa
6001 aataaatgat gattatgtgg cactttattg aattttatag aaatattcca atctaagtga
6061 taaagtaaca actgagaaac ttcagaatga aagccttatt gattcaaaag gaaaattacc
6121 atatgcttta gggctgatga acttgccaca attgcttaga cataacaaag atattttcta
6181 cattgttttc caactttata tgctgaaaga taattaggag gtccgggtgg agtgtatcta
6241 aatgactaga actttaagtt gcaatataga ttttctcttt ttaacaaagg aagaatataa
6301 attaatttgg atcaaattta tttgccttct ttgcaatctt ggtgatcatt ttggaaagta
6361 aattgaaagg aaagttaaat agccacatag ttttcttttg catctcaatt tggttgagaa
6421 tttctaagga aggttaatat gactttagaa tgaataaaga gactgtcaaa caacaagtac
6481 tctccctaaa aaagaggaaa gaatgccttc ctaaaatatt tgtttcctat tgtatatgtc
6541 accaattgga aaatgccaat ctcatctgca agtgtgagaa atggaacaag gagaggagcc
6601 aactctagta aaatgtagtg aggccataat ccaaggcaaa gtttaacaaa tcttgattgt
6661 gtaatttctt attctgttga ttcattttgt cacccaccag actaccaaaa ataaagcaca
6721 gacatgacaa ttttagtata tacaacagat ttggcaaata atacataatc ttattgtaat
6781 tgccagtagc aatatcttaa ggggcatgag tctttcacag tattggactt tgaaaaattg
6841 gtacagggag taattttcaa cagactgata cttgaagtca ctcaattact actttgacct
6901 atcataattt gtttctaatc ttgtcatttg ctcatatata ctattaatat aaccagttgc
6961 tcttttcata atgtactatt tctatgtaaa gatgtaaaga ttttgatttt cctatgtttt
7021 aaaattgtaa tgtgcatgat cttttatttc ttgtgtttca attaccaagt cagtattcta
7081 aggttgtaag attactgtaa aaattatatt ccaatatttt atctaaactt aagtgtggcc
7141 agcattattg gtttcagagt tgcctaaaca acacagaata tccttaattc agtttgtaaa
7201 agagttttga ccagtgtaaa aaagtgaaga cacagtttca tgtttcaagt ttaaaatgga
7261 aagttaatat catttgagca cttatgtgtt ttcatgtctt ctacttaatc tgtatgtgaa
7321 agtaaaatat ttttaacacc attatttatg ccatgtaaag tgggttctca gaagcacaag
7381 cacagatatt taccttctga agacattttt ggctataaaa gtgcattgga tggcaaacac
7441 tatttgagtc aggtgttaga aatttattag atgtattata ttacatagta gaataaggtc
7501 cctttctcat tttgtttttc ctaaaaataa gaaaaaaaag aagtgtgata gtaaactgtc
7561 tttgctaatg tcctgccaga tgtttaactt caaattaagc aaaaagtagg tacaatatga
7621 tgatgatcct gatgatgatg atgatgatga tgatgaagaa taaatggaat caaaatgcta
7681 gctttatttg acattaataa gtaaaataag tcatcatttt ttcaactctg tagcacagct
7741 gtttacattg aataattttc tctattgtgc tgttaattat atagtaatgt atcaaaatag
7801 aaataggaac tttcttttca tttcttattc atttatgaag aaatcttttc acacataaat
7861 gattgaatgt tcttgttcag aatgaccaca taatcattgc ttagaagaag ataccaattc
7921 ttttaaaaga aaaaaagtct gtttataata atcagaatca aatgttgttt gtttcttcta
7981 aacgttaatg gagaaaattg aaggtggtaa aatgtcatgt ttattcaggc tgggaactgt
8041 attcacagta gaagtttcag tggtcaacat atctatgact ctttaggctg ctgtagtttt
8101 acagtcaatt atttaaaagt gagtagttac atttataaga gcctgagaat acttagactc
8161 agtcatttgt tagtattttt accaaaatct cttagtttca gacatgtcag aagcagctat
8221 atagcatatc ttattctatg atatacatca ggctatctca agttcctgtc tcacagttaa
8281 ttcaaagaag gattaggatt tctgtatttt ttctcatttg aatctttatg tgcatttggt
8341 ttgtgtacat gctttttgta gtgtaagata tgaaatttta tatttttca gaaaataaaa
8401 acccttttgaa tacagttaaa aaaaaaaaaa aaaaa

By "Ephrin Type-A Receptor 5 (EPHA5)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAI43428, version AAI43428.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 43):

```
   1 mrgsgprgag rrrppsgggd tpitpaslag cysaprrapl wtclllcaal rtllaspsne
  61 vnlldsrtvm gdlgwiafpk ngweeigevd enyapihtyq vckvmeqnqn nwlltswisn
 121 egasrifiel kftlrdcnsl pgglgtcket fnmyyfesdd qngrnikenq yikidtiaad
 181 esfteldlgd rvmklntevr dvgplskkgf ylafqdvgac ialvsvrvyy kkcpsvvrhl
 241 avfpdtitga dssqllevsg scvnhsvtde ppkmhcsaeg ewlvpigkcm ckagyeekng
 301 tcqvcrpgff kasphiqscg kcpphsythe eastscvcek dyfrresdpp tmactrppsa
 361 prnaisnvne tsvflewipp adtggrkdvs yyiackkcns hagvceecgg hvrylprqsg
 421 lkntsvmmvd llahtnytfe ieavngvsdl spgarqyvsv nvttnqaaps pvtnvkkgki
 481 aknsislswq epdrpngiil eyeikyfekd qetsytiiks kettitaegl kpasvyvfqi
 541 rartaagygv fsrrfefett pvsvaassdq sqipviavsv tvgvillavv igvllsgrrc
 601 gyskakqdpe eekmhfhngh iklpgvrtyi dphtyedpnq avhefakeie ascitiervi
 661 gagefgevcs grlklpgkre lpvaiktlkv gytekqrrdf lgeasimgqf dhpniihleg
 721 vvtkskpvmi vteymengsl dtflkkndgq ftviqlvgml rgisagmkyl sdmgyvhrdl
 781 aarnilinsn lvckvsdfgl srvleddpea ayttrggkip irwtapeaia frkftsasdv
 841 wsygivmwev vsygerpywe mtnqdvikav eegyrlpspm dcpaalyqlm ldcwqkerns
 901 rpkfdeivnm ldklirnpss lktlvnascr vsnllaehsp lgsgayrsvg ewleaikmgr
 961 yteifmengy ssmdavaqvt ledlrrlgvt lvghqkkimn slqemkvqlv ngmvpl
```

By "Neurotrophic Receptor Tyrosine Kinase 3 (NTRK3) nucleic acid molecule" is meant a polynucleotide encoding a NTRK3 polypeptide. An exemplary NTRK3 nucleic acid molecule is provided at NCBI Accession No. NM_001012338, version NM_001012338.2, incorporated herein by reference, and reproduced below (SEQ ID NO: 44):

```
   1 acatttctgc agccgcgcgg cgagccattc gcggcggctg ctgcagctcc tactgcatct
  61 tccttctctt cctttcctcg ggctccggtc tcggagtcgg agagcgcgcc tcgcttccag
 121 agcccccgga cccggcgagt cagcgatcgc cgagccggcc accatgcccg gcagaccgcg
 181 ccactaggcg ctcctcgcgg ctcccacccg gcggcggcgg cggcggcggc ggcgtccgcg
 241 atggtttcag acgctgaagg attttgcatc tgatcgctcg gcgtttcaaa gaagcagcga
 301 tcggagatgg atgtctctct ttgcccagcc aagtgtagtt tctggcggat tttcttgctg
 361 ggaagcgtct ggctggacta tgtgggctcc gtgctggctt gccctgcaaa ttgtgtctgc
 421 agcaagactg agatcaattg ccggcggccg gacgatggga acctcttccc cctcctggaa
 481 gggcaggatt cagggaacag caatgggaac gccagtatca acatcacgga catctcaagg
 541 aatatcactt ccatacacat agagaactgg cgcagtcttc acacgctcaa cgccgtggac
 601 atggagctct acaccggact tcaaaagctg accatcaaga actcaggact tcggagcatt
 661 cagcccagag cctttgccaa gaacccccat ttgcgttata taaacctgtc aagtaaccgg
 721 ctcaccacac tctcgtggca gctcttccag acgctgagtc ttcgggaatt gcagttggag
 781 cagaactttt tcaactgcag ctgtgacatc cgctggatgc agctctggca ggagcagggg
 841 gaggccaagc tcaacagcca gaacctctac tgcatcaacg ctgatggctc ccagcttcct
 901 ctcttccgca tgaacatcag tcagtgtgac cttcctgaga tcagcgtgag ccacgtcaac
 961 ctgaccgtac gagagggtga caatgctgtt atcacttgca atggctctgg atcacccctt
1021 cctgatgtgg actggatagt cactgggctg cagtccatca acactcacca gaccaatctg
1081 aactggacca atgttcatgc catcaacttg acgctggtga atgtgacgag tgaggacaat
1141 ggcttcaccc tgacgtgcat tgcagagaac gtggtgggca tgagcaatgc cagtgttgcc
1201 ctcactgtct actatccccc acgtgtggtg agcctggagg agcctgagct gcgcctggag
```

```
1261 cactgcatcg agtttgtggt gcgtggcaac cccccaccaa cgctgcactg gctgcacaat
1321 gggcagcctc tgcgggagtc caagatcatc catgtggaat actaccaaga gggagagatt
1381 tccgagggct gcctgctctt caacaagccc acccactaca acaatggcaa ctataccctc
1441 attgccaaaa acccactggg cacagccaac cagaccatca atggccactt cctcaaggag
1501 ccctttccag agagcacgga taactttatc ttgtttgacg aagtgagtcc cacacctcct
1561 atcactgtga cccacaaacc agaagaagac acttttgggg tatccatagc agttggactt
1621 gctgcttttg cctgtgtcct gttggtggtt ctcttcgtca tgatcaacaa atatggtcga
1681 cggtccaaat ttggaatgaa gggtcccgtg gctgtcatca gtggtgagga ggactcagcc
1741 agcccactgc accacatcaa ccacggcatc accacgccct cgtcactgga tgccgggccc
1801 gacactgtgg tcattggcat gactcgcatc cctgtcattg agaaccccca gtacttccgt
1861 cagggacaca actgccacaa gccggacacg tatgtgcagc acattaagag gagagacatc
1921 gtgctgaagc gagaactggg tgagggagcc tttggaaagg tcttcctggc cgagtgctac
1981 aacctcagcc cgaccaagga caagatgctt gtggctgtga aggccctgaa ggatcccacc
2041 ctggctgccc ggaaggattt ccagagggag gccgagctgc tcaccaacct gcagcatgag
2101 cacattgtca agttctatgg agtgtgcggc gatggggacc ccctcatcat ggtctttgaa
2161 tacatgaagc atggagacct gaataagttc ctcagggccc atgggccaga tgcaatgatc
2221 cttgtggatg gacagccacg ccaggccaag ggtgagctgg ggctctccca aatgctccac
2281 attgccagtc agatcgcctc gggtatggtg tacctggcct cccagcactt tgtgcaccga
2341 gacctggcca ccaggaactg cctggttgga gcgaatctgc tagtgaagat tggggacttc
2401 ggcatgtcca gagatgtcta cagcacggat tattacaggc tctttaatcc atctggaaat
2461 gatttttgta tatggtgtga ggtgggagga cacaccatgc tccccattcg ctggatgcct
2521 cctgaaagca tcatgtaccg gaagttcact acagagagtg atgtatggag cttcggggtg
2581 atcctctggg agatcttcac ctatggaaag cagccatggt tccaactctc aaacacggag
2641 gtcattgagt gcattaccca aggtcgtgtt ttggagcggc cccgagtctg ccccaaagag
2701 gtgtacgatg tcatgctggg gtgctggcag agggaaccac agcagcggtt gaacatcaag
2761 gagatctaca aaatcctcca tgctttgggg aaggccaccc caatctacct ggacattctt
2821 ggctagtggt ggctggtggt catgaattca tactctgttg cctcctctct ccctgcctca
2881 catctccctt ccacctcaca actccttcca tccttgactg aagcgaacat cttcatataa
2941 actcaagtgc ctgctacaca tacaacactg aaaaaaggaa aaaaaaagaa agaaaaaaaa
3001 accc
```

By "Neurotrophic Receptor Tyrosine Kinase 3 (NTRK3)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAH13693, version AAH13693.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 45):

```
  1 mdvslcpakc sfwrifllgs vwldyvgsvl acpancvcsk teincrrpdd gnlfpllegq
 61 dsgnsngnas initdisrni tsihienwrs lhtlnavdme lytglqklti knsglrsiqp
121 rafaknphlr yinlssnrlt tlswqlfqtl slrelqleqn ffncscdirw mqlwqeqgea
181 klnsqnlyci nadgsqlplf rmnisqcdlp eisvshvnlt vregdnavit cngsgsplpd
241 vdwivtglqs inthqtnlnw tnvhainltl vnvtsednqf tltciaenvv gmsnasvalt
301 vyypprvvsl eepelrlehc iefvvrgnpp ptlhwlhngq plreskiihv eyyqegeise
361 gcllfnkpth ynngnytlia knplgtanqt inghflkepf pestdnfilf devsptppit
421 vthkpeedtf gvsiavglaa facvllvvlf vminkygrrs kfgmkgpvav isgeedsasp
481 lhhinhgitt pssldagpdt vvigmtripv ienpqyfrqg hnchkpdtwv fsnidnhgil
541 nlkdnrdhlv psthyiyeep evqsgevsyp rshgfreiml npislpghsk plnhgiyved
601 vnvyfskgrh gf
```

By "Androgen Receptor (AR) nucleic acid molecule" is meant a polynucleotide encoding a AR polypeptide. An exemplary AR nucleic acid molecule is provided at NCBI Accession No. NM_000044, version NM_000044.4, incorporated herein by reference, and reproduced below (SEQ ID NO: 46):

```
   1 gcggagagaa ccctctgttt tcccccactc tctctccacc tcctcctgcc ttccccaccc
  61 cgagtgcgga gccagagatc aaaagatgaa aaggcagtca ggtcttcagt agccaaaaaa
 121 caaaacaaac aaaaacaaaa aagccgaaat aaaagaaaaa gataataact cagttcttat
 181 ttgcacctac ttcagtggac actgaatttg gaaggtggag gattttgttt ttttctttta
 241 agatctgggc atcttttgaa tctacccttc aagtattaag agacagactg tgagcctagc
 301 agggcagatc ttgtccaccg tgtgtcttct tctgcacgag actttgaggc tgtcagagcg
 361 ctttttgcgt ggttgctccc gcaagtttcc ttctctggag cttcccgcag gtgggcagct
 421 agctgcagcg actaccgcat catcacagcc tgttgaactc ttctgagcaa gagaagggga
 481 ggcggggtaa gggaagtagg tggaagattc agccaagctc aaggatggaa gtgcagttag
 541 ggctgggaag ggtctaccct cggccgccgt ccaagaccta ccgaggagct ttccagaatc
 601 tgttccagag cgtgcgcgaa gtgatccaga acccgggccc caggcaccca gaggccgcga
 661 gcgcagcacc tcccggcgcc agtttgctgc tgctgcagca gcagcagcag cagcagcagc
 721 agcagcagca gcagcagcag cagcagcagc agcagcagca gcaagagact agccccaggc
 781 agcagcagca gcagcagggt gaggatggtt ctccccaagc ccatcgtaga ggccccacag
 841 gctacctggt cctggatgag gaacagcaac cttcacagcc gcagtcggcc ctggagtgcc
 901 accccgagag aggttgcgtc ccagagcctg gagccgccgt ggccgccagc aaggggctgc
 961 cgcagcagct gccagcacct ccggacgagg atgactcagc tgccccatcc acgttgtccc
1021 tgctgggccc cactttcccc ggcttaagca gctgctccgc tgaccttaaa gacatcctga
1081 gcgaggccag caccatgcaa ctccttcagc aacagcagca ggaagcagta tccgaaggca
1141 gcagcagcgg gagagcgagg gaggcctcgg gggctcccac ttcctccaag gacaattact
1201 taggggggcac ttcgaccatt tctgacaacg ccaaggagtt gtgtaaggca gtgtcggtgt
1261 ccatgggcct gggtgtggag gcgttggagc atctgagtcc aggggaacag cttcgggggg
1321 attgcatgta cgccccactt ttgggagttc cacccgctgt gcgtcccact ccttgtgccc
1381 cattggccga atgcaaaggt tctctgctag acgacagcgc aggcaagagc actgaagata
1441 ctgctgagta ttccccttcc aagggaggtt acaccaaagg gctagaaggc gagagcctag
1501 gctgctctgg cagcgctgca gcagggagct ccgggacact tgaactgccg tctaccctgt
1561 ctctctacaa gtccggagca ctggacgagg cagctgcgta ccagagtcgc gactactaca
1621 actttccact ggctctggcc ggaccgccgc cccctccgcc gcctccccat ccccacgctc
1681 gcatcaagct ggagaacccg ctggactacg gcagcgcctg ggcggctgcg gcggcgcagt
1741 gccgctatgg ggacctggcg agcctgcatg gcgcgggtgc agcgggaccc ggttctgggt
1801 caccctcagc cgccgcttcc tcatcctggc acactctctt cacagccgaa gaaggccagt
1861 tgtatggacc gtgtggtggt ggtgggggtg gtggcggcgg cggcggcggc ggcggcggcg
1921 gcggcggcgg cggcggcggc ggcgaggcgg gagctgtagc cccctacggc tacactcggc
1981 cccctcaggg gctggcgggc caggaaagcg acttcaccgc acctgatgtg tggtaccctg
2041 gcggcatggt gagcagagtg ccctatccca gtcccacttg tgtcaaaagc gaaatgggcc
2101 cctggatgga tagctactcc ggaccttacg gggacatgcg tttggagact gccagggacc
2161 atgttttgcc cattgactat tactttccac cccagaagac ctgcctgatc tgtggagatg
2221 aagcttctgg gtgtcactat ggagctctca catgtggaag ctgcaaggtc ttcttcaaaa
2281 gagccgctga agggaaacag aagtacctgt gcgccagcag aaatgattgc actattgata
2341 aattccgaag gaaaaattgt ccatcttgtc gtcttcggaa atgttatgaa gcagggatga
2401 ctctgggagc ccggaagctg aagaaacttg gtaatctgaa actacaggag gaaggagagg
2461 cttccagcac caccagcccc actgaggaga caacccagaa gctgacagtg tcacacattg
2521 aaggctatga atgtcagccc atctttctga atgtcctgga agccattgag ccaggtgtag
2581 tgtgtgctgg acacgacaac aaccagcccg actcctttgc agccttgctc tctagcctca
2641 atgaactggg agagagacag cttgtacacg tggtcaagtg ggccaaggcc ttgcctggct
2701 tccgcaactt acacgtggac gaccagatgg ctgtcattca gtactcctgg atggggctca
2761 tggtgtttgc catgggctgg cgatccttca ccaatgtcaa ctccaggatg ctctacttcg
```

2821 cccctgatct ggttttcaat gagtaccgca tgcacaagtc ccggatgtac agccagtgtg
2881 tccgaatgag gcacctctct caagagtttg gatggctcca aatcaccccc caggaattcc
2941 tgtgcatgaa agcactgcta ctcttcagca ttattccagt ggatgggctg aaaaatcaaa
3001 aattctttga tgaacttcga atgaactaca tcaaggaact cgatcgtatc attgcatgca
3061 aaagaaaaaa tcccacatcc tgctcaagac gcttctacca gctcaccaag ctcctggact
3121 ccgtgcagcc tattgcgaga gagctgcatc agttcacttt tgacctgcta atcaagtcac
3181 acatggtgag cgtggacttt ccggaaatga tggcagagat catctctgtg caagtgccca
3241 agatcctttc tgggaaagtc aagcccatct atttccacac ccagtgaagc attggaaacc
3301 ctatttcccc accccagctc atgccccctt tcagatgtct tctgcctgtt ataactctgc
3361 actactcctc tgcagtgcct tggggaattt cctctattga tgtacagtct gtcatgaaca
3421 tgttcctgaa ttctatttgc tgggcttttt ttttctcttt ctctcctttc tttttcttct
3481 tccctcccta tctaaccctc ccatggcacc ttcagacttt gcttcccatt gtggctccta
3541 tctgtgtttt gaatggtgtt gtatgccttt aaatctgtga tgatcctcat atggcccagt
3601 gtcaagttgt gcttgtttac agcactactc tgtgccagcc acacaaacgt ttacttatct
3661 tatgccacgg gaagtttaga gagctaagat tatctgggga aatcaaaaca aaaacaagca
3721 aacaaaaaaa aaaagcaaaa acaaaacaaa aaataagcca aaaaaccttg ctagtgtttt
3781 ttcctcaaaa ataaataaat aaataaataa atacgtacat acatacacac atacatacaa
3841 acatatagaa atccccaaag aggccaatag tgacgagaag gtgaaaattg caggcccatg
3901 gggagttact gatttttttca tctcctccct ccacgggaga ctttattttc tgccaatggc
3961 tattgccatt agagggcaga gtgaccccag agctgagttg ggcaggggggg tggacagaga
4021 ggagaggaca aggagggcaa tggagcatca gtacctgccc acagccttgg tccctggggg
4081 ctagactgct caactgtgga gcaattcatt atactgaaaa tgtgcttgtt gttgaaaatt
4141 tgtctgcatg ttaatgcctc acccccaaac ccttttctct ctcactctct gcctccaact
4201 tcagattgac tttcaatagt ttttctaaga cctttgaact gaatgttctc ttcagccaaa
4261 acttggcgac ttccacagaa aagtctgacc actgagaaga aggagagcag agatttaacc
4321 ctttgtaagg ccccatttgg atccaggtct gctttctcat gtgtgagtca gggaggagct
4381 ggagccagag gagaagaaaa tgatagcttg gctgttctcc tgcttaggac actgactgaa
4441 tagttaaact ctcactgcca ctacctttc cccaccttta aaagacctga atgaagtttt
4501 ctgccaaact ccgtgaagcc acaagcacct tatgtcctcc cttcagtgtt ttgtgggcct
4561 gaatttcatc acactgcatt tcagccatgg tcatcaagcc tgtttgcttc ttttgggcat
4621 gttcacagat tctctgttaa gagcccccac caccaagaag gttagcaggc caacagctct
4681 gacatctatc tgtagatgcc agtagtcaca aagatttctt accaactctc agatcgctgg
4741 agcccttaga caaactggaa agaaggcatc aaagggatca ggcaagctgg gcgtcttgcc
4801 cttgtccccc agagatgata ccctcccagc aagtggagaa gttctcactt ccttctttag
4861 agcagctaaa ggggctaccc agatcagggt tgaagagaaa actcaattac cagggtggga
4921 agaatgaagg cactagaacc agaaaccctg caaatgctct tcttgtcacc cagcatatcc
4981 acctgcagaa gtcatgagaa gagagaagga acaaagagga gactctgact actgaattaa
5041 aatcttcagc ggcaaagcct aaagccagat ggacaccatc tggtgagttt actcatcatc
5101 ctcctctgct gctgattctg ggctctgaca ttgcccatac tcactcagat tccccacctt
5161 tgttgctgcc tcttagtcag agggaggcca aaccattgag actttctaca gaaccatggc
5221 ttctttcgga aaggtctggt tggtgtggct ccaatacttt gccacccatg aactcagggt
5281 gtgccctggg acactggttt tatatagtct tttggcacac ctgtgttctg ttgacttcgt
5341 tcttcaagcc caagtgcaag ggaaaatgtc cacctacttt ctcatcttgg cctctgcctc
5401 cttacttagc tcttaatctc atctgttgaa ctcaagaaat caagggccag tcatcaagct
5461 gcccatttta attgattcac tctgtttgtt gagaggatag tttctgagtg acatgatatg
5521 atccacaagg gtttccttcc ctgatttctg cattgatatt aatagccaaa cgaacttcaa
5581 aacagcttta aataacaagg gagaggggaa cctaagatga gtaatatgcc aatccaagac
5641 tgctggagaa aactaaagct gacaggttcc ctttttgggg tgggatagac atgttctggt
5701 tttctttatt attacacaat ctggctcatg tacaggatca cttttagctg ttttaaacag
5761 aaaaaaatat ccaccactct tttcagttac actaggttac attttaatag gtcctttaca
5821 tctgttttgg aatgattttc atcttttgtg atacacagat tgaattatat cattttcata
5881 tctctccttg taaatactag aagctctcct ttacatttct ctatcaaatt tttcatcttt
5941 atgggtttcc caattgtgac tcttgtcttc atgaatatat gtttttcatt tgcaaaagcc
6001 aaaaatcagt gaaacagcag tgtaattaaa agcaacaact ggattactcc aaatttccaa
6061 atgacaaaac tagggaaaaa tagcctacac aagcctttag gcctactctt tctgtgcttg
6121 ggtttgagtg aacaaaggag attttagctt ggctctgttc tcccatggat gaaaggagga

6181 ggatttttt tttcttttgg ccattgatgt tctagccaat gtaattgaca gaagtctcat
6241 tttgcatgcg ctctgctcta caaacagagt tggtatggtt ggtatactgt actcacctgt
6301 gagggactgg ccactcagac ccacttagct ggtgagctag aagatgagga tcactcactg
6361 gaaaagtcac aaggaccatc tccaaacaag ttggcagtgc tcgatgtgga cgaagagtga
6421 ggaagagaaa aagaaggagc accagggaga aggctccgtc tgtgctgggc agcagacagc
6481 tgccaggatc acgaactctg tagtcaaaga aaagagtcgt gtggcagttt cagctctcgt
6541 tcattgggca gctcgcctag gcccagcctc tgagctgaca tgggagttgt tggattcttt
6601 gtttcatagc tttttctatg ccataggcaa tattgttgtt cttggaaagt ttattatttt
6661 tttaactccc ttactctgag aaagggatat tttgaaggac tgtcatatat ctttgaaaaa
6721 agaaaatctg taatacatat atttttatgt atgttcactg gcactaaaaa atatagagag
6781 cttcattctg tcctttgggt agttgctgag gtaattgtcc aggttgaaaa ataatgtgct
6841 gatgctagag tccctctctg tccatactct acttctaaat acatataggc atacatagca
6901 agttttattt gacttgtact ttaagagaaa atatgtccac catccacatg atgcacaaat
6961 gagctaacat tgagcttcaa gtagcttcta agtgtttgtt tcattaggca cagcacagat
7021 gtggcctttc ccccttctc tcccttgata tctggcaggg cataaaggcc caggccactt
7081 cctctgcccc ttcccagccc tgcaccaaag ctgcatttca ggagactctc tccagacagc
7141 ccagtaacta cccgagcatg gcccctgcat agccctggaa aaataagagg ctgactgtct
7201 acgaattatc ttgtgccagt tgcccaggtg agagggcact gggccaaggg agtggttttc
7261 atgtttgacc cactacaagg ggtcatggga atcaggaatg ccaaagcacc agatcaaatc
7321 caaaacttaa agtcaaaata agccattcag catgttcagt ttcttggaaa aggaagtttc
7381 tacccctgat gcctttgtag gcagatctgt tctcaccatt aatctttttg aaaatctttt
7441 aaagcagttt ttaaaaagag agatgaaagc atcacattat ataaccaaag attacattgt
7501 acctgctaag ataccaaaat tcataagggc agggggggag caagcattag tgcctctttg
7561 ataagctgtc caaagacaga ctaaaggact ctgctggtga ctgacttata agagctttgt
7621 gggtttttt ttccctaata atatacatgt ttagaagaat tgaaaataat ttcgggaaaa
7681 tgggattatg ggtccttcac taagtgattt tataagcaga actggctttc cttttctcta
7741 gtagttgctg agcaaattgt tgaagctcca tcattgcatg gttggaaatg gagctgttct
7801 tagccactgt gtttgctagt gcccatgtta gcttatctga agatgtgaaa cccttgctga
7861 taagggagca tttaaagtac tagattttgc actagaggga cagcaggcag aaatccttat
7921 ttctgcccac tttggatggc acaaaaagtt atctgcagtt gaaggcagaa agttgaaata
7981 cattgtaaat gaatatttgt atccatgttt caaaattgaa atatatatat atatatatat
8041 atatatatat atatatatat agtgtgtgtg tgtgttctga tagctttaac tttctctgca
8101 tctttatatt tggttccaga tcacacctga tgccatgtac ttgtgagaga ggatgcagtt
8161 ttgttttgga agctctctca gaacaaacaa gacacctgga ttgatcagtt aactaaaagt
8221 tttctcccct attgggtttg acccacaggt cctgtgaagg agcagaggga taaaaagagt
8281 agaggacatg atacattgta ctttactagt tcaagacaga tgaatgtgga aagcataaaa
8341 actcaatgga actgactgag atttaccaca gggaaggccc aaacttgggg ccaaaagcct
8401 acccaagtga ttgaccagtg gccccctaat gggacctgag ctgttggaag aagagaactg
8461 ttccttggtc ttcaccatcc ttgtgagaga agggcagttt cctgcattgg aacctggagc
8521 aagcgctcta tctttcacac aaattccctc acctgagatt gaggtgctct tgttactggg
8581 tgtctgtgtg ctgtaattct ggttttggat atgttctgta aagattttga caaatgaaaa
8641 tgtgtttttc tctgttaaaa cttgtcagag tactagaagt tgtatctctg taggtgcagg
8701 tccatttctg cccacaggta gggtgttttt ctttgattaa gagattgaca cttctgttgc
8761 ctaggacctc ccaactcaac catttctagg tgaaggcaga aaaatccaca ttagttactc
8821 ctcttcagac atttcagctg agataacaaa tcttttggaa ttttttcacc catagaaaga
8881 gtggtagata tttgaattta gcaggtggag tttcatagta aaaacagctt ttgactcagc
8941 tttgatttat cctcatttga tttggccaga aagtaggtaa tatgcattga ttggcttctg
9001 attccaattc agtatagcaa ggtgctaggt tttttccttt ccccacctgt ctcttagcct
9061 ggggaattaa atgagaagcc ttagaatggg tggcccttgt gacctgaaac acttcccaca
9121 taagctactt aacaagattg tcatggagct gcagattcca ttgcccacca aagactagaa
9181 cacacacata tccatacacc aaaggaaaga caattctgaa atgctgtttc tctggtggtt
9241 ccctctctgg ctgctgcctc acagtatggg aacctgtact ctgcagaggt gacaggccag
9301 atttgcatta tctcacaacc ttagcccttg gtgctaactg tcctacagtg aagtgcctgg
9361 ggggttgtcc tatcccataa gccacttgga tgctgacagc agccaccatc agaatgaccc
9421 acgcaaaaaa aagaaaaaaa aaattaaaaa gtcccctcac aacccagtga caccttttctg
9481 ctttcctcta gactggaaca ttgattaggg agtgcctcag acatgacatt cttgtgctgt

```
 9541 ccttggaatt aatctggcag caggagggag cagactatgt aaacagagat aaaaattaat
 9601 tttcaatatt gaaggaaaaa agaaataaga agagagagag aaagaaagca tcacacaaag
 9661 attttcttaa aagaaacaat tttgcttgaa atctctttag atggggctca tttctcacgg
 9721 tggcacttgg cctccactgg gcagcaggac cagctccaag cgctagtgtt ctgttctctt
 9781 tttgtaatct tggaatcttt tgttgctcta aatacaatta aaaatggcag aaacttgttt
 9841 gttggactac atgtgtgact ttgggtctgt ctctgcctct gctttcagaa atgtcatcca
 9901 ttgtgtaaaa tattggctta ctggtctgcc agctaaaact tggccacatc ccctgttatg
 9961 gctgcaggat cgagttattg ttaacaaaga gacccaagaa aagctgctaa tgtcctctta
10021 tcattgttgt taatttgtta aaacataaag aaatctaaaa tttcaaaaaa
```

By "Androgen Receptor (AR)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. AAA51771, version AAA51771.1, incorporated herein by reference, as reproduced below (SEQ ID NO: 47):

```
  1 mevqlglgrv yprppsktyr gafqnlfqsv reviqnpgpr hpeaasaapp gasllllqqq
 61 qqqqqqqqqq qqqqqqqets prqqqqqqge dgspqahrrg ptgylvldee qqpsqpqsal
121 echpergcvp epgaavaask glpqqlpapp deddsaapst lsllgptfpg lsscsadlkd
181 ilseastmql lqqqqqeavs egsssgrare rsgaptsskd nylggtstis dnakelckav
241 svsmglgvea lehlspgeql rgdcmyapll gvppavrptp caplaeckgs llddsagkst
301 edtaeyspfk ggytkglege slgcsgsaaa gssgtlelps tlslyksgal deaaayqsrd
361 yynfplalag pppppppphp hariklenpl dygsawaaaa aqcrygdlas lhgagaagpg
421 sgspsaaass swhtlftaee gqlygpcggg gggggggggg gggggggggg eagavapygy
481 trppqglagq esdftapdvw ypggmvsrvp ypsptcvkse mgpwmdsysg pygdmrleta
541 rdhvlpidyy fppqktclic gdeasgchyg altcgsckvf fkraaegkqk ylcasrndct
601 idkfrrkncp scrlrkcyea gmtlgarklk klgnlklqee geassttspt eettqkltvs
661 hiegyecqpi flnvleaiep gvvcaghdnn qpdsfaalls slnelgerql vhvvkwakal
721 pgfrnlhvdd qmaviqyswm glmvfamgwr sftnvnsrml yfapdlvfne yrmhksrmys
781 qcvrmrhlsq efgwlqitpq eflcmkalll fsiipvdglk nqkffdelrm nyikeldrii
841 ackrknptsc srrfyqltkl ldsvqpiare lhqftfdlli kshmvsvdfp emmaeiisvq
901 vpkilsgkvk piyfhtq
```

The following is a detailed description provided to aid those skilled in the art in practicing the present disclosure. Those of ordinary skill in the art may make modifications and variations in the embodiments described herein without departing from the spirit or scope of the present disclosure. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. The terminology used in the description of the disclosure herein is for describing particular embodiments only and is not intended to be limiting of the disclosure. All publications, patent applications, patents, figures and other references mentioned herein are expressly incorporated by reference in their entirety.

The practice of the present subject matter may employ, unless otherwise indicated, conventional techniques and descriptions of organic chemistry, molecular biology (including

recombinant techniques), cell biology, and biochemistry, which are within the skill of the art. Such conventional techniques include, but are not limited to, preparation of synthetic polynucleotides, polymerization techniques, chemical and physical analysis of polymer particles, preparation of nucleic acid libraries, nucleic acid sequencing and analysis, and the like. Specific illustrations of suitable techniques can be used by reference to the examples provided herein. Other equivalent conventional procedures can also be used. Such conventional techniques and descriptions can be found in standard laboratory manuals such as Genome Analysis: A Laboratory Manual Series (Vols. I-IV), PCR Primer: A Laboratory Manual, and Molecular Cloning: A Laboratory Manual (all from Cold Spring Harbor Laboratory Press), Hermanson, Bioconjugate Techniques, Second Edition (Academic Press, 2008); Merkus, Particle Size Measurements (Springer, 2009); Rubinstein and Colby, Polymer Physics (Oxford University Press, 2003); "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (Gait, ed., 1984); "Animal Cell Culture" (Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Handbook of Experimental Immunology" (Wei & Blackwell, eds.); "Gene Transfer Vectors for Mammalian Cells" (Miller & Calos, eds., 1987); "Current Protocols in Molecular Biology" (Ausubel et al., eds., 1987); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994); and "Current Protocols in Immunology" (Coligan et al., eds., 1991). These techniques are applicable to the production of the polynucleotides and polypeptides, and, as such, can be considered in making and practicing the disclosure.

The primers of the disclosure and their functional derivatives can include any suitable polynucleotide that can hybridize to a target sequence of interest. The primers can serve to prime nucleic acid synthesis, e.g., in a PCR reaction. Typically, the primer functions as a substrate onto which nucleotides can be polymerized by a polymerase; in some embodiments, however, the primer can become incorporated into the synthesized nucleic acid strand and provide a site to which another primer can hybridize to prime synthesis of a new strand that is complementary to the synthesized nucleic acid molecule. The primers of the disclosure may be comprised of any combination of nucleotides or analogs thereof, which may be optionally linked to form a linear polymer of any suitable length. In some embodiments, the primers are single-stranded oligonucleotides or polynucleotides. In some embodiments, the primers are single-stranded. The primers can also be double-stranded.

The primers optionally occur naturally, as in a purified restriction digest, or can be produced synthetically. In some embodiments, the primers act as a point of initiation for amplification or synthesis when exposed to amplification or synthesis conditions; such amplification or synthesis can occur in a template-dependent fashion and optionally results in formation of a primer extension product that is complementary to at least a portion of the target sequence.

Exemplary amplification or synthesis conditions can include contacting the primer with a polynucleotide template (e.g., a template including a target SLGI sequence or sequences), nucleotides and an inducing agent such as a polymerase at a suitable temperature and pH to induce polymerization of nucleotides onto an end of the target-specific primer. If double-stranded, the primer can optionally be treated to separate its strands before being used to prepare primer extension products. In some embodiments, the primer is an oligodeoxyribonucleotide or an oligoribonucleotide. In some embodiments, the primer can include one or more nucleotide analogs. The exact length and/or composition, including sequence, of the target-specific primer can influence many properties, including melting temperature (Tm), GC content, formation of secondary structures, repeat nucleotide motifs, length of predicted primer extension products, extent of coverage across a nucleic acid molecule of interest, number of primers present in a single amplification or synthesis reaction, presence of nucleotide analogs or modified nucleotides within the primers, and the like.

In some embodiments, a primer can be paired with a compatible primer within an amplification or synthesis reaction to form a primer pair consisting or a forward primer and a reverse primer. In some embodiments, the forward primer of the primer pair includes a sequence that is substantially complementary to at least a portion of a strand of a nucleic acid molecule, and the reverse primer of the primer of the primer pair includes a sequence that is substantially identical to at least of portion of the strand. In some embodiments, the forward primer and the reverse primer are capable of hybridizing to opposite strands of a nucleic acid duplex. Optionally, the forward primer primes synthesis of a first nucleic acid strand, and the reverse primer primes synthesis of a second nucleic acid strand, wherein the first and second strands are substantially complementary to each other, or can hybridize to form a double-stranded nucleic acid molecule.

In some embodiments, one end of an amplification or synthesis product is defined by the

forward primer and the other end of the amplification or synthesis product is defined by the reverse primer. In some embodiments, where the amplification or synthesis of lengthy primer extension products is required, such as amplifying an exon, coding region, or gene, several primer pairs can be created than span the desired length to enable sufficient amplification of the region. In some embodiments, a primer can include one or more cleavable groups.

In some embodiments, primer lengths are in the range of about 10 to about 60 nucleotides, about 12 to about 50 nucleotides and about 15 to about 40 nucleotides in length. Typically, a primer is capable of hybridizing to a corresponding target sequence and undergoing primer extension when exposed to amplification conditions in the presence of dNTPs and a polymerase. In some instances, the particular nucleotide sequence or a portion of the primer is known at the outset of the amplification reaction or can be determined by one or more of the methods disclosed herein. In some embodiments, the primer includes one or more cleavable groups at one or more locations within the primer.

In the various disclosed embodiments, any suitable length primers are contemplated. The length of the primers may be limited by a minimum primer length threshold and a maximum primer length, and a length score for the primers may be set so as to decrease as the length gets shorter than the minimum primer length threshold and to decrease as the length gets longer than the maximum primer length threshold. In an embodiment, the minimum primer length threshold may be 16. In other embodiments, the minimum primer length threshold may be 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, or 5, for example, and may also be 17, 18, 19, 20, 21, 22, 23, and 24, for example. In an embodiment, the maximum primer length threshold may be 28. In other embodiments, the maximum primer length threshold may be 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, and 40, for example, and may also be 27, 26, 25, 24, 23, 22, 21, and 20, for example. In an embodiment, the primer length criterion may be given a score of 1.0 if the length thresholds are satisfied, for example, and that score may go down to 0.0 as the primer length diverges from the minimum or maximum length threshold. For example, if the maximum primer length threshold were set to 28, then the score could be set to 1.0 if the length does not exceed 28, to 0.7 if the length is 29, to 0.6 if the length is 30, to 0.5 if the length is 31, to 0.3 if the length is 32, to 0.1 if the length is 33, and to 0.0 if the length is 34 or more. The attribute/score could be scaled between values other than 0.0 and 1.0, of course, and the function defining how the score varies with an increase difference relative to the threshold could be any other or more complex linear or

non-linear function that does not lead to increases in score for primer that further diverge from length thresholds.

In various embodiments, the method of the disclosure preferably utilizes wildtype primer sets that are modified to prevent their extension by a polymerase in a PCR reaction or in a PCR-based assay. Such modification can be any known in the art. For example, the wildtype primers can be modified with a 3' end blocking group which prevents extension by DNA polymerase. One such blocking group can include a 3'-end dideoxyCytosine (ddC), which is covalently modified on the 3' terminal phosphate and prevents extension by DNA polymerase. Any other suitable blocking group known in the art is contemplated which blocks DNA polymerase extension.

In various embodiments, the detection of PCR products resulting from the methods of the disclosure may be performed by any known read-out methodology, such as by nucleotide sequence, gel-based detection, or by molecular reporter system. Such read-out methodologies are well-known in the art and the skilled person will understand how to use such read-out techniques to in the disclosed detection methods.

In various aspects, the read-out methods may be conducted with the aid of a computer-based system configured to execute machine-readable instructions, which, when executed by a processor of the system causes the system to perform steps including determining the identity, size, nucleotide sequence or other measurable characteristics of the amplicons produced in the method of the disclosure. One or more features of any one or more of the above-discussed teachings and/or exemplary embodiments may be performed or implemented using appropriately configured and/or programmed hardware and/or software elements. Determining whether an embodiment is implemented using hardware and/or software elements may be based on any number of factors, such as desired computational rate, power levels, heat tolerances, processing cycle budget, input data rates, output data rates, memory resources, data bus speeds, etc., and other design or performance constraints.

Examples of hardware elements may include control units, processors, microprocessors, input(s) and/or output(s) (I/O) device(s) (or peripherals) that are communicatively coupled via a local interface circuit, circuit elements (e.g., transistors, resistors, capacitors, inductors, and so forth), integrated circuits, application specific integrated circuits (ASIC), programmable logic

devices (PLD), digital signal processors (DSP), field programmable gate array (FPGA), logic gates, registers, semiconductor device, chips, microchips, chip sets, and so forth. The local interface may include, for example, one or more buses or other wired or wireless connections, controllers, buffers (caches), drivers, repeaters and receivers, etc., to allow appropriate communications between hardware components. A processor is a hardware device for executing software, particularly software stored in memory. The processor can be any custom made or commercially available processor, a central processing unit (CPU), an auxiliary processor among several processors associated with the computer, a semiconductor-based microprocessor (e.g., in the form of a microchip or chip set), a macroprocessor, or generally any device for executing software instructions. A processor can also represent a distributed processing architecture. The I/O devices can include input devices, for example, a keyboard, a mouse, a scanner, a microphone, a touch screen, an interface for various medical devices and/or laboratory instruments, a bar code reader, a stylus, a laser reader, a radio-frequency device reader, etc. Furthermore, the I/O devices also can include output devices, for example, a printer, a bar code printer, a display, etc. Finally, the I/O devices further can include devices that communicate as both inputs and outputs, for example, a modulator/demodulator (modem; for accessing another device, system, or network), a radio frequency (RF) or other transceiver, a telephonic interface, a bridge, a router, etc. It is expressly contemplated that the components and/or elements described herein can be implemented as software being stored on a tangible (non-transitory) computer-readable medium (e.g., disks/CDs/etc.) having program instructions executing on a computer, hardware, firmware, or a combination thereof.

Examples of software may include software components, programs, applications, computer programs, application programs, system programs, machine programs, operating system software, middleware, firmware, software modules, routines, subroutines, functions, methods, procedures, software interfaces, application program interfaces (API), instruction sets, computing code, computer code, code segments, computer code segments, words, values, symbols, or any combination thereof. A software in memory may include one or more separate programs, which may include ordered listings of executable instructions for implementing logical functions. The software in memory may include a system for identifying data streams in accordance with the present teachings and any suitable custom made or commercially available operating system (O/S), which may control the execution of other computer programs such as the

system, and provides scheduling, input-output control, file and data management, memory management, communication control, etc.

According to various exemplary embodiments, one or more features of any one or more of the above-discussed teachings and/or exemplary embodiments may be performed or implemented at least partly using a distributed, clustered, remote, or cloud computing resource.

According to various exemplary embodiments, one or more features of any one or more of the above-discussed teachings and/or exemplary embodiments may be performed or implemented using a source program, executable program (object code), script, or any other entity comprising a set of instructions to be performed. When using a source program, the program can be translated via a compiler, assembler, interpreter, etc., which may or may not be included within the memory, so as to operate properly in connection with the O/S. The instructions may be written using (a) an object-oriented programming language, which has classes of data and methods, or (b) a procedural programming language, which has routines, subroutines, and/or functions, which may include, for example, C, C++, Pascal, Basic, Fortran, Cobol, Pert, Java, and Ada.

According to various exemplary embodiments, one or more of the above-discussed exemplary embodiments may include transmitting, displaying, storing, printing or outputting to a user interface device, a computer readable storage medium, a local computer system or a remote computer system, information related to any information, signal, data, and/or intermediate or final results that may have been generated, accessed, or used by such exemplary embodiments. Such transmitted, displayed, stored, printed or outputted information can take the form of searchable and/or filterable lists of runs and reports, pictures, tables, charts, graphs, spreadsheets, correlations, sequences, and combinations thereof, for example.

Various additional exemplary embodiments may be derived by repeating, adding, or substituting any generically or specifically described features and/or components and/or substances and/or steps and/or operating conditions set forth in one or more of the above-described exemplary embodiments. Further, it should be understood that an order of steps or order for performing certain actions is immaterial so long as the objective of the steps or action remains achievable, unless specifically stated otherwise. Furthermore, two or more steps or actions can be conducted simultaneously so long as the objective of the steps or action remains

achievable, unless specifically stated otherwise. Moreover, any one or more feature, component, aspect, step, or other characteristic mentioned in one of the above-discussed exemplary embodiments may be considered to be a potential optional feature, component, aspect, step, or other characteristic of any other of the above-discussed exemplary embodiments so long as the objective of such any other of the above-discussed exemplary embodiments remains achievable, unless specifically stated otherwise.

The term "cancer," as used herein, may include, but is not limited to: biliary tract cancer; bladder cancer; brain cancer including glioblastomas and medulloblastomas; breast cancer; cervical cancer; choriocarcinoma; colon cancer; endometrial cancer; esophageal cancer; gastric cancer; hematological neoplasms including acute lymphocytic and myelogenous leukemia; multiple myeloma; AIDS-associated leukemia and adult T-cell leukemia lymphoma; intraepithelial neoplasms including Bowen's disease and Paget's disease; liver cancer; lung cancer; lymphomas including Hodgkin's disease and lymphocytic lymphomas; neuroblastomas; oral cancer including squamous cell carcinoma; ovarian cancer including those arising from epithelial cells, stromal cells, germ cells and mesenchymal cells; pancreatic cancer; prostate cancer; rectal cancer; sarcomas including leiomyosarcoma, rhabdomyosarcoma, liposarcoma, fibrosarcoma, and osteosarcoma; skin cancer including melanoma, Kaposi's sarcoma, basocellular cancer, and squamous cell cancer; testicular cancer including germinal tumors such as seminoma, non-seminoma, teratomas, choriocarcinomas; stromal tumors and germ cell tumors; thyroid cancer including thyroid adenocarcinoma and medullar carcinoma; and renal cancer including adenocarcinoma and Wilms' tumor. Commonly encountered cancers include breast, prostate, lung, ovarian, colorectal, and brain cancer. In general, an effective amount of the compositions of the disclosure for treating cancer will be that amount necessary to inhibit mammalian cancer cell proliferation *in situ*. Those of ordinary skill in the art are well-schooled in the art of evaluating effective amounts of anti-cancer agents.

Reference will now be made in detail to exemplary embodiments of the disclosure. While the disclosure will be described in conjunction with the exemplary embodiments, it will be understood that it is not intended to limit the disclosure to those embodiments. To the contrary, it is intended to cover alternatives, modifications, and equivalents as may be included within the spirit and scope of the disclosure as defined by the appended claims.

EXAMPLES

The present disclosure is further illustrated by the following examples, which should not be construed as limiting. The contents of all references, GenBank Accession and Gene numbers, and published patents and patent applications cited throughout the application are hereby incorporated by reference. Those skilled in the art will recognize that the disclosure may be practiced with variations on the disclosed structures, materials, compositions and methods, and such variations are regarded as within the scope of the disclosure.

According to the techniques herein, the simultaneous lentiviral delivery of paired guide RNAs (pgRNAs) targeting two separate genes in a CRISPR/Cas9 knockout (KO) screen may provide a cost-effective approach for high throughput identification of SLGIs. The present disclosure provides experimental technologies and computational methods to conduct large-scale prediction, identification, and validation of synthetic lethal gene interaction (SLGIs) involved in cancer. In particular, the below Examples describe a novel pgRNA CRISPR vector system, vector library, screening techniques and integrative algorithms to find novel therapies targeting cancers with tumor suppressor gene (TSG) loss. Prior art SLGI studies in humans have either focused on a single SLGI pair or compared essential genes between cancer cell lines where one anchor gene is wild-type or mutant (e.g., a "1 x n" design) or via combinatorial pairs (e.g., an "a x b" design), which drastically limits the number of effective SLGI pairs that can be investigated. Due to these limitations, the current collection of human SLGI pairs that have a high degree of confidence is only about 100. The present disclosure provides cutting-edge and cost-effective technologies for high throughput identification, prediction, and validation of SLGIs in individual cell lines. First, the techniques herein provide a novel pooled CRISPR/Cas9 double KO screening technique in which each lentivirus carries pgRNAs designed to simultaneously KO specific pairs of SLGI partners. Second, the techniques herein provide a novel computational algorithm that integrates pgRNA screening data, available single guide RNA (sgRNA) CRISPR screening data, and The Cancer Genome Atlas (TCGA) tumor profiling data, to predict SLGI pairs. Third, the techniques herein provide large-scale pgRNA CRISPR screens across different cancer cell lines to identify and characterize cancer-specific SLGIs. The techniques herein will enable comprehensive identification of therapeutic targets for cancers with TSG loss, and will inform better development of precision cancer medicine.

Example 1: CRISPR Screens with a "1 x n" Design Identified P21 (RAC1) Activated Kinase 2 (PAK2) as a C-Src Tyrosine Kinase (CSK) SLGI Partner in Breast Cancers

CRISPR/Cas9 KO libraries with a sgRNA per vector targeting exons have been proven to be a powerful genetic screen platform (see e.g., reference 7). The techniques herein expand sgRNA screening to a pgRNA modality. As shown in FIGS. 2A-2F, initial experiments have shown that two rounds of CRISPR screening using a "1 x n" design identified a unique synthetic lethal pair that drives hormone independent cell growth in breast cancer models. In particular, these CRISPR screens identified PAK2 and CSK as a SLGI pair in breast cancer cells.

As shown in FIG. 2A, a genome-wide sgRNA CRISPR knockout screen was first conducted in the T47D and MCF7 breast cancer cell lines to search for key genes whose loss would specifically drive estrogen-independent growth. CSK was identified as the strongest positively-selected hit in both T47D and MCF7 cell lines (FIGS. 2A-C). CSK knockout confers hormone independent growth, which could be fully reversed by the overexpression of a human CSK cDNA (FIG. 2D).

To identify key genes that drive hormone independent growth upon CSK loss, a second round of genome-wide CRISPR screen was performed to compare the T47D-CSK null vs T47D-CSK wild type cells (FIG. 2E). This secondary screen identified PAK2 as possibly having a SLGI in combination with CSK because PAK2 is uniquely essential in the CSK-null cells (FIG. 2F). Based on this method, a series of genome-wide CRISPR screens were conducted by simultaneously knocking out another positively-selected gene(s) such as Tuberous Sclerosis 1/2 (TSC1/2) in T47D, which provides multiple "1 x n" design SLGI pairs with which to train the algorithms described below.

Example 2: A pgRNA Library Enables CRISPR Deletion Screens to Find Functional lncRNAs in Human Cancers

The simultaneous expression of two gRNAs targeting two different genes in the genome may introduce indels to KO both genes. Alternatively, if the two targeting sites are close to each other, the fragment in between could be deleted (see e.g., reference 26). Therefore, with a reliable cloning method to construct pgRNA CRISPR libraries, a high-throughput SLGI screen(s) or deletion screen(s) may be conducted.

A two-step pgRNA library (see e.g., reference 27) was capable of delivering the expression of two gRNAs per lentiviral vector and building the cell library pool in a similar way as in single gene CRISPR KO libraries (FIGS. 3A-3B) and screening methods (FIG. 3C) as described in Zhu et al. (Nat Biotechnol. 2016 Dec;34(12):1279-1286). FIG. 3B shows DNA sequences of the engineered oligo and linker between the two gRNAs of each pair, which sequence is set forth below (SEQ ID NO: 29):

5'-ATCTTGTGGAAAGGACGAAACACCG
[+guide1+]
GTTTAGAGACGAGCCTCTATACTTACTAAACGTGATCGTCTCAACCG
[+guide2+]
GTTTAAGAGCTATGCTGGAAACAGC-3'

In this screen, the same U6 was used in front of both gRNAs; therefore, it was only possible to sequence the first gRNA as a barcode for each pgRNA pair and decode the screen results. Unfortunately, this sequencing strategy could not assay whether the pairs swapped during the library construction, screening, or sequencing preparation processes because it only decodes the first half of pgRNA information. Additionally, this strategy also limits the choices of pgRNA design by requiring the first gRNA to be unique in every pair. This screening strategy also suffered a relatively high false negative rate, potentially due to PCR swapping/recombination that disrupting the designed pgRNA pairing.

Example 3: Novel pgRNA Oligo Design with a Unique Linker Improves the Quality of the pgRNA Library

According to the techniques herein, paired-end sequencing could decode both pgRNAs in each pair and reveal a substantial portion of the swapped pairs in the library. To reduce the swapping rate, the present disclosure provides a novel pgRNA expression system design in which two different U6 promoters (e.g., a human U6 promoter and mouse U6 promoter) are used to drive expression of two gRNAs, each of which is followed sequentially by a different scaffold sequence that includes a tracrRNA sequence. Advantageously, this design minimizes the

possibility of lentiviral replication-generated recombination (see e.g., references 28 and 29), and it decreases the swapping rate at the cell library level.

As shown in FIG. 1 and FIG. 3H, paired-end sequencing analysis of swapped pairs generated in prior art pgRNA library design revealed that the first amplification step of the oligo library may generate around 50% of all swapped pairs in the library, and also that these swapped pairs are preserved in later plasmid vector and cell libraries. It was believed that the common linker between the two gRNAs resulted in the PCR-generated swapping events. In a pilot 7.5K pgRNA library construction experiment in which two gRNAs flank a cis-element for deletion, this hypothesis was confirmed when an altered oligo design in which every pair contains a unique linker completely eliminated the swapping issue during the first PCR step. However, in the second cloning step, the tracrRNA-U6 promoter sequence is inserted between the first gRNA sequence and the second gRNA sequence, and the inserted tracrRNA-U6 fragment then becomes a common linker. As shown in FIG. 3I, the analysis of the colony PCR amplicons from the complete vector library, in which the PCR-related recombination events are eliminated because each colony has only one pgRNA vector, 12/12 of the pgRNAs are correct pairs.

To prepare the deep sequencing samples from the vector pool or the genomic DNA of the cell library, it was necessary to PCR amplify the pgRNA sequence and add sequencing adaptors, which again created swapped/recombined pairs at a frequency of about 50%. However, since the screening was still done with the correct pairing and swapping only happened during the final step of preparing the library before sequencing, it was possible to filter the pgRNAs with the wrong pairing from the sequencing data. A pilot 7.5K library screen yielded good results with low false discovery rate even with a single replicate, demonstrating the ability of the techniques herein to conduct robust and cost-effective pgRNA CRISPR screens in cancer cell lines.

The techniques herein provide, in part, a pgRNA library vector including two gRNA cassettes and a Cas9 expression cassette (see e.g., FIG. 3D) and methods for constructing the same (FIG. 3E). In the library vector design, two different U6 promoters (e.g., a human U6 promoter and mouse U6 promoter) may be used to drive expression of two different gRNAs in conjunction with two different gRNA scaffolds. It is contemplated with the scope of the disclosure that any of a variety of different promoters may be used, and one of skill in the art will appreciate that the choice of promoters may vary depending upon a variety of factors such as the cell type and/or disease state of the cell line that is being screened. For example, alternate

promoters may include, but are not limited to, the H1 promoter (see e.g., Myslinski, E., Ame, J.C., Krol, A. and Carbon, P. (2001) An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. Nucleic Acids Res., 29, 2502–2509), the 7SK promoter (see e.g., Murphy, S., Di Liegro, C. and Melli, M. (1987) The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. Cell, 51, 81–87), or a modified bovine U6 promoter (see e.g., Adamson et al. (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. Volume 167, Issue 7, p1867–1882).

Library Construction: Design and synthesis of the oligo library

FIG. 3E shows a method of making the present pgRNA vector that greatly reduces, or eliminates, internal recombination between pgRNAs, thereby increasing the fidelity of resulting pgRNA libraries.

In an exemplary embodiment shown in FIG. 3F, the design of the oligo may be as follows (SEQ ID NO: 16): 5'-GTGGAAAGGACGAAACACCG+guide1+GTTTNGAGACGNNNNNNNNNNNNNNNNCGTCTCNGTTG+guide2+GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGG-3'. It is contemplated within the scope of the invention that each gRNA pair may have a different linker (e.g., a unique linker that may be randomly designed and assigned to a given gRNA pair), in sharp contrast to prior art methods. In this regards, the specific linker used for a given gRNA pair does not matter so long as each gRNA pair has a different linker.

While the above exemplary embodiment discloses a 16 nucleotide linker (NNNNNNNNNNNNNNNN (SEQ ID NO: 17)), it is contemplated within the scope of the disclosure that the linker may range from 10-30 nucleotides in length. In exemplary embodiments, the GC content of the linker may be less than or equal to 40% (e.g., 40%, 39%, 38%, 37%, 36%, 35%, 34%, 33%, 32%, 31%, 30%, 29%, 28%, 27%, 26%, 25%, 24%, 23%, 22%, 21%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%).

Exemplary gRNAs may be selected from any genomic regions of interest that match the PAM requirement (e.g., a trailing or leading NGG) and/or the guide efficiency model. In an exemplary embodiment, the length of both gRNAs may be 19 nucleotides, so the total length of the product is 130 nucleotides. One of skill in the art will appreciate that the length of the gRNA

may be slightly longer or shorter (e.g., the gRNA length may range from about 17-27 nucleotides in length).

The manufacture of the oligo pool may be conducted by Agilent Technologies Inc. or Twist Biosciences, Inc.

An exemplary forward oligo (e.g., oligo_F) may have the following sequence (SEQ ID NO: 18):

TAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG

An exemplary reverse oligo (e.g., oligo_R) may have the following sequence (SEQ ID NO: 19):

ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

Library Construction: Two-step cloning of the oligo pool into a lentiCRISPRv2 vector

FIG. 3E depicts an exemplary two step cloning process that may be used to make the library vectors disclosed herein. In a first step, a Gibson assembly reaction may be applied to an exemplary linearized (e.g., enzymatically digested) lentiCRISPRv2 vector backbone (e.g., including in 5' to 3' order: a human U6 promoter, a vector linker, and a second gRNA scaffold; see e.g., FIG. 3E, top panel) in which the vector linker has been removed and an amplified oligonucleotide library having the general structure, in 5' to 3' order, first gRNA-unique linker-second gRNA (FIG. 3F), to create an intermediate nucleic acid sequence having the following exemplary structure in 5' to 3' order: a human U6 promoter, first gRNA, a unique linker (e.g., randomized linker), second gRNA, second gRNA scaffold (see e.g., FIG. 3E, middle panel).

In an exemplary embodiment, the vector linker may have the following sequence (SEQ ID NO: 20):

GAGACGGTTGTAAATGAGCACACAAAATACACATGCTAAAATATTATATTCTATGACCTTTATAAAATCAACCAAAATCTTCTTTTTAATAACTTTAGTATCAATAATTAGAATTTTTATGTTCCTTTTTGCAAACTTTTAATAAAAATGAGCAAAATAAAAAAACGCTAGTTTTAGTAACTCGCGTTGTTTTCTTCACCTTTAATAATAGCTACTCCACCACTTGTTCCTAAGCGGTCAGCTCCTGCTTCAATCATTTTTTGAGCATCTTCAAATGTTCTAACTCCACCAGCTGCTTTAACTAAAGCATTGTCTTTAACAACTGACTTCATTAGTTTAACATCTTCAAATGTTGCACCTGATTTTGAAAATCCTGTTGATGTTTTAACAAATTCTAATCCAGCT

TCAACAGCTATTTCACAAGCTTTCATGATTTCTTCTTTTGTTAATAAACAATTTTCCA
TAATACATTTAACAACATGTGATCCAGCTGCTTTTTTTACAGCTTTCATGTCTTCTAA
AACTAATTCATAATTTTTGTCTTTTAATGCACCAATATTTAATACCATATCAATTTCT
GTTGCACCATCTTTAATTGCTTCAGAAACTTCGAATGCTTTTGTAGCTGTTGTGCATG
CACCTAGAGGAAAACCTACAACATTTGTTATTCCTACATTTGTGCCTTTTAATAATTC
TTTACAATAGCTTGTTCAATATGAATTAACACAAACTGTTGCAAAATCAAATTCAAT
TGCTTCATCACATAATTGTTTAATTTCAGCTTTCGTAGCATCTTGTTTTAATAATGTGT
GATCTATATATTTGTTTAGTTTCATTTTTTCTCCTATATATTCATTTTTAATTTTAATTC
TTTAATAATTTCGTCTACTTTAACTTTAGCGTTTTGAACAGATTCACCAACACCTATA
AAATAAATTTTTAGTTTAGGTTCAGTTCCACTTGGGCGAACAGCAAATCATGACTTA
TCTTCTAAATAAAATTTTAGTAAGTCTTGTCCTGGCATATTATACATTCCATCGATGT
AGTCTTCAACATTAACAACTTTAAGTCCAGCAATTTGAGTTAAGGGTGTTGCTCTCA
ATGATTTCATTAATGGTTCAATTTTTAATTTCTTTTCTTCTGGTTTAAAATTCAAGTTT
AAAGTGAAAGTGTAATATGCACCCATTTCTTTAAATAAATCTTCTAAATAGTCTACT
AATGTTTTATTTTGTTTTTTATAAAATCAAGCAGCCTCTGCTATTAATATAGAAGCTT
GTATTCCATCTTTATCTCTAGCTGAGTCATCAATTACATATCCATAACTTTCTTCATA
AGCAAAAACAAAATTTAATCCGTTATCTTCTTCTTTAGCAATTTCTCTACCCATTCAT
TTAAATCCAGTTAAAGTTTTTACAATATTAACTCCATATTTTTCATGAGCGATTCTAT
CACCCAAATCACTTGTTACAAAACTTGAATATAGAGCCGGATTTTTTGGAATGCTAT
TTAAGCGTTTTAGATTTGATAATTTTCAATCAATTAAAATTGGTCCTGTTTGATTTCC
ATCTAATCTTACAAAATGACCATCATGTTTTATTGCCATTCCAAATCTGTCAGCATCT
GGGTCATTCATAATAATAATATCTGCATCATGTTTAATACCATATTCAAGCGGTATTT
TTCATGCAGGATCAAATTCTGGATTTGGATTTACAACATTTTTAAATGTTTCATCTTC
AAATGCATGCTCTTCAACCTCAATAACGTTATATCCTGATTCACGTAATATTTTTGGG
GTAAATTTAGTTCCTGTTCCATTAACTGCGCTAAAAATAATTTTTAAATCTTTTTTAG
CTTCTTGCTCTTTTTTGTACGTCTCT

(see e.g., the world wide web at (www)addgene.org/52961/).

It is also contemplated within the scope of the disclosure that the region of sequence overlap for the Gibson reaction may be at least 30 nucleotides in length.

In a second step, the intermediate nucleic acid sequence may be linearized by removing the unique linker, and a ligation reaction may then occur between the linearized intermediate

nucleic acid sequence and a linker block having the structure, in 5' to 3' order: a first gRNA scaffold, a unique linker sequence, and a mouse U6 promoter.

An exemplary linker block may contain a first gRNA scaffold and mouse U6 promotor (shown in bold)(SEQ ID NO: 21):

GTTTAAGAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATC
AACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTCGAGTACTAGGATCCATTA
GGCGGCCGCGTCGACAAGCTTTCTAGAGAATTC**GATCCGACGCGCCATCTCTAGG**
CCCGCGCCGGCCCCCTCGCACGGACTTGTGGGAGAAGCTCGGCTACTCCCCTG
CCCCGGTTAATTTGCATATAATATTTCCTAGTAACTATAGAGGCTTAATGTGCG
ATAAAAGACAGATAATCTGTTCTTTTTAATACTAGCTACATTTTACATGATAGG
CTTGGATTTCTATAACTTCGTATAGCATACATTATACGAAGTTATAAACAGCAC
AAAAGGAAACTCACCCTAACTGTAAAGTAATTGTGTGTTTTGAGACTATAAGTA
TCCCTTGGAGAACCACCTTGTTG

A complete exemplary linker sequence including leading and trailing sequences may contain the following sequence (SEQ ID NO: 22):

TATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAgaattaatttgactg
taaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatat
gcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacaccg**CCTCCCGCTCCTGGAGCGG***GTTTAA*
GAGCTAAGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTC
*GGTGCTTTTTTT*CTCGAGTACTAGGATCCATTAGGCGGCCGCGTCGACAAGCTTTCTAGAGAATTC**gatccgacgcgcc**
atctctaggcccgcgccggccccctcgcacggacttgtgggagaagctcggctactcccctgccccggttaatttgcatataatatttcctagtaacta
tagaggcttaatgtgcgataaaagacagataatctgttctttttaatactagctacattttacatgataggcttggatttctataacttcgtatagcata
cattatacgaagttataaacagcacaaaaggaaactcaccctaactgtaaagtaattgtgtgttttgagactataagtatcccttggagaaccacct
tgttg*GATATTCACCATTATAGGT**GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGA*
*AAAAGTGGCACCGAGTCGGTGCTTTTTT*GAATTCTAGACTTGATGCTAACTAGGTCTTGAAAGGAGTGGGAATTG
GCTCCGGTGCCCGTCAGT

(The human U6 promoter is shown in lowercase, mouse U6 promoter is shown in bold lowercase, gRNA1 is shown in uppercase bold, gRNA2 is shown in uppercase bold italic, and the first and second scaffold sequences, respectively, are shown in uppercase italic).

Once the ligation reaction between the linearized intermediate nucleic acid sequence and a linker block is complete, a pgRNA library vector having a nucleic acid sequence including, in

5' to 3' order, a human U6 promoter, a first gRNA, a first gRNA scaffold, a unique linker, a mouse U6 promoter, a second gRNA, and a second gRNA scaffold is constructed (see e.g., FIG. 3E, lower panel).

Decoding the pgRNA libraries

The pgRNA libraries may be decoded by amplifying the pgRNA region from the plasmid or genomic DNA samples with the following exemplary primers:

pgRNA_Lib_F (SEQ ID NO: 23):
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC TTTGTGGAAAGGACGAAACACCG

pgRNA_Lib_R1 (SEQ ID NO: 24):
TCTACTATTCTTTCCCCTGCACTGTACCCGGACTAGCCTTATTTTAACTTGCTATTTCT AGCTCTAAAAC

pgRNA_Lib_R2 (SEQ ID NO: 25):
CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTC CGATCTNNNNNNNNTCTACTATTCTTTCCCCTGCACTGTACC (N(8) is the specific index sequences)

The amplified pgRNA library may then be sequenced using any of a variety of high throughput sequencing techniques known in the art such as, for example, the Illumina high-throughput platform.

read1_seq (SEQ ID NO: 26):
GAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (for the 1stgRNA)

read2_seq (SEQ ID NO: 27):
TGCACTGTACCCGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC (for the 2stgRNA)

index_seq (SEQ ID NO: 28):
GCTAGTCCGGGTACAGTGCAGGGGAAAGAATAGTAGA

In a pilot scale pgRNA CRISPR screen using the above pgRNA library vector, a 7.5k pgRNA library was used to delete regulatory cis-elements in a human breast cancer line T47D. The sequencing data of the vector library and cell library by our new paired-end sequencing method demonstrated that that library quality was very high and that there was minimal recombination between the two gRNAs.

As shown in FIG. 3G, the method of vector construction depicted in FIG. 3E reduces frequencies of recombination/swapping of pgRNAs during library construction.

Example 4: Design and Construction of the SLGI pgRNA CRISPR Library

A pgRNA CRISPR library was synthesized in an "a x b" design to explore all genetic interactions between anchors (i.e., part "a") and partners (i.e., part "b") using an improved oligo design with the following general structure: "gRNA1 + unique linker + gRNA2". Part "a" may include four TSGs including Phosphatase and Tensin Homolog (PTEN), Neurofibromin 1 (NF1), RB Transcriptional Corepressor 1 (RB1), C-Src Tyrosine Kinase (CSK), as well as one control anchor, AAVS1, that has no function in the genome. Part "b" may include 121 genes that encode kinases and are targets of approved drugs according to annotations in the OASIS database (see e.g., reference 30), as well as AAVS1 as a control.

In an exemplary embodiment, the screen was carried out in a breast cancer cell line, T47D, in which no mutations are detected in any of the four TSG anchors. Between each anchor-partner pair, 21 pgRNA pairs may be designed. Advantageously, this number of pgRNA pairs conveniently fit in one 15K Agilent oligo synthesis order (21 * (4+1) * (121+1) < 15K). Each gene has 7 unique CRISPR gRNAs designed from an efficiency model (see e.g., reference 31) and validated recent screens. 21 pgRNA pairs were then selected according to the selection matrix from all 49 possible pairwise gRNA combinations (FIG. 14). The 15K pgRNA vector library was then constructed from the faithfully amplified oligo pool using the two-step cloning described in detail above. The lentivirus was packaged from the vector library and the four cell lines was infected at low MOI (~0.3) with 500-fold coverage to build the cell libraries with biological replicates.

Quality control was assessed for both plasmid and cell libraries by paired-end pgRNA sequencing to ensure the coverage and evenness of all designed pgRNAs and to check for swapping/recombination events (FIG. 15). The frequency of such swapping/recombination events were addressed by sequencing the library samples deeper to ensure sufficient coverage of the library after the swapped products have been eliminated.

Example 5: SLGI pgRNA CRISPR Screen

To screen for SLGI pairs that play key roles in cell growth, library cells were cultured at over 500-fold coverage of the library size for 11-12 population doubling times (~3 weeks for T47D cells) and the genomic DNA was harvested on day 0 and the end time point. After amplifying the pgRNA from all the samples, deep-sequencing libraries were prepared to submit for paired-end sequencing to decode the pgRNA information. The functional positive control SLGI pairs were confirmed in the screen, indicating that the screen works well (FIG. 16).

To call SLGI genes anchored on each TSG, the method based on regression residual was used, which is similar to the approach used in shRNA screens (see e.g., reference 9). The phenotype for each CRISPR gRNA in either the single (e.g., targeting gene X as a partner to AAVS1) or double (e.g., targeting gene X as a partner to TSG) KO was quantified as the fold change in gRNA abundance between selection and the day 0 control. For most of gRNAs, a linear relationship between the phenotype of the single and double KO is expected. Each gRNA on the partner and paired with a TSG gRNA may be ranked by the p-value (fold-change determines rank directions) of its deviations from the linear fit between double KO and single KO phenotype (FIG. 4; FIG. 17A-FIG. 17D). The top ranked SLGI pairs include RB1_MAPK8, RB1_JAK3, PTEN_CDK12, PTEN_AKT3, NF1_TYR03, NF1_EPHA5, CSK_NTRK3 and CSK_AR. Another method may adopt the BLISS independence model (see e.g., reference 32, incorporated herein by reference).

The techniques herein provide a robust pgRNA CRISPR screening technique, as well as a data analysis pipeline for SLGI identification.

The pgRNA CRISPR screening techniques described herein have the potential to create segmental genomic deletions in the situation where two gRNAs target a pair of genes that are in close proximity to one another. To avoid this confounding issue, all gene pairs that are within 1 mega base pair of one another in the library design may generally be excluded. An alternative strategy to study genetic interactions between proximal gene pairs is to use a CRISPR interference screening technology that avoids genome cutting.

Another potential issue of an “a x b” paired design is that paired-end sequencing of the pgRNA may underestimate pgRNA swapping frequency from the sequencing preparation PCR step. However, as discussed above, use of an exo-polymerase may reduce the swapping rate by about 25% and top pgRNA hits can still be reliably identified. Even at swapping rate of about

50%, top pgRNA hits may still be identified because a particular swapped pair will only happen at a very low frequency, which is unlikely to overwhelm the frequency of the correct pgRNA pair.

Another potential issue may arise in the circumstance in which copy number alteration (CNA) confounds gene essentiality in a CRISPR screen (see e.g., reference 34). CNA may be addressed by the techniques here by assessing CNA profiles of screened cell lines in view of databases such as, for example, CCLE35 and GDSC databases (see e.g., reference 36), which may be used to reduce or eliminate the impact of CNA in determined pgRNA essentiality scores.

Example 6: Optimized sgRNA Design and Gene Calling for Genome-wide CRISPR Screens

Recent studies of CRISPR guide efficiency have analyzed the growth effects of different sgRNAs targeting genes that are essential for cell growth, and identified DNA sequence features that contribute to sgRNA efficiency in CRISPR-based screens (see e.g., reference 31). Leveraging the information from multiple sgRNA library designs (see e.g., references 7, 21, and 22), the techniques herein provide a new sequence model for predicting sgRNA efficiency in CRISPR/Cas9 KO experiments. This model confirms known features and suggests new features that include, but are not limited to, a preference for cytosine at the cleavage site (FIG. 5A). The model was experimentally validated for sgRNA-mediated mutation rate and gene KO efficiency (FIG. 5B) in that it achieved significant results under both positive and negative selection conditions, and clearly outperformed existing models (such as, e.g., those described in reference 37).

The ability to use CRISPR screen technology to identify CRISPR screen hits from sequencing data, a statistical algorithm, MAGeCK, Model-based Analysis of Genome-wide CRISPR/Cas9 KO has been developed (see e.g., reference 33). The MAGeCK algorithm was expanded via an updated algorithm, MAGeCK-VISPR, which provides a comprehensive quality control (QC), analysis, and visualization workflow for CRISPR screen analysis (see e.g., reference 38). Given the design matrix annotating the different screen conditions (FIG. 5C), MAGeCK first uses the sequence model to estimate sgRNA efficiency. It then iteratively updates each sgRNA efficiency based on whether the sgRNA behavior follows the selection of the gene across conditions (see, e.g., the E step in FIG. 5D), and uses the updated sgRNA efficiency to estimate the level of gene selection in different samples (see, e.g., the M step in FIG. 5D).

Example 7: Novel Algorithm to Predict SLGI Pairs

The present disclosure provides a new algorithm for SLGI prediction. About 5,000 experimentally validated SLGI pairs in yeast (see e.g., reference 1) were assembled and their corresponding orthologous human genes were identified. The patterns of gene mutation, expression in TCGA, and protein-protein-interactions (PPI) of these orthologous genes were then examined. Using these yeast-to-human SLGI pairs as positive controls and 5,000 randomly selected gene pairs as negative controls, a feature selection and regression model was constructed to predict whether a pair of human genes will have SLGI. In this model, the response variable is whether the pair has SLGI, whereas the independent variables include expression, mutation, and CNV features of the two interacting genes in TCGA molecular profiles and PPI.

In TCGA breast, prostate, lung, and colon cancer data that was examined, the regression model consistently found the following factors to show statistical significance in predicting SLGI:

1) better overall positive expression correlation in the tumor samples;

2) more PPI;

3) better positive fold change (tumor-to-normal) correlation than random pairs;

4) when one SLGI gene is frequently deleted in cancer, the expression of the other often significantly increases; and

5) when one SLGI gene shows down-expression in tumor over normal, the expression correlations of the pairs tend to be negative.

Using the yeast SLGI and TCGA data for training and one of the very few available mammalian high throughput experimental genetic interaction (GI) screens for testing (see e.g., reference 11), it was found that the new algorithm disclosed herein provides a statistically significant separation between the genetically interacting (GI) pairs and non-GI pairs (FIG. 6).

The techniques herein provide that the new SLGI prediction algorithm may be refined/improved in a variety of ways. For example, more independent variables (features) for testing and selection may be included in our regression model. Such independent variables may include, but are not limited to, correlations of expression and mutations (including CNA) in different TCGA cancer types, frequency of mutations or differential expression in TCGA, as well as the association of a gene's expression or mutation with patient prognosis. This may allow SLGI pairs that have robust relationship to be identified across most TCGA cancer types, as well

as those unique to certain cancer types. The RABIT method may be to select those independent variables (features) that are predictive of SLGI (see e.g., reference 39). RABIT utilizes the efficient Frisch-Waugh-Lovell theorem to correct confounding effects in linear models for fast stepwise feature selection.

As another example, efficiency of the prediction algorithm may be increased by using more SLGI data, which may include pgRNA CRISPR SLGI screening data and "1 x n" design CRISPR SLGI screening data. Additionally, efficiency of the prediction algorithm may be increased by adding known SLGI pairs in yeast and C. elegans that have orthologous genes in human, literature-reported SLGI individual genes in mammalian genomes, as well as the previous shRNA screens for SLGI (e.g., SynLethDB40). The regression model may be trained on each known SLGI dataset separately, evaluated for its performance using 10-fold cross validation (CV), and each dataset may be assigned a specific weight based on the CV R2 metric. Then, all the known SLGI datasets may be combined into one feature selection and regression model, with weights assigned to each dataset proportional to its cross-validation performance (FIG. 7). Preliminary testing conducted by adding new features (e.g., PPI) or data (e.g., combining yeast SLGI pairs with human colon cancer shRNA screen), the new algorithm may improve the area under the curve (AUC) on the receiver operating characteristic (ROC) curve by > 0.1 to final AUC > 0.7.

The above described SLGI algorithm may predict a likelihood of SLGI between every pair of human genes in each cancer type. However, the specific expression and mutation profiles in a particular patient tumor or cancer cell line dictate a tumor- or cell-line specific prediction of SLGI. For each cell line or tumor sample of a specific cancer type, the molecular profiles may be examined and an activity score for each gene may be computed based on its molecular profiles in the tumor. Low activity scores reflect copy number deletion, nonsense/frameshift mutations, or lower expression level, while high activity scores represent copy number amplification, known gain-of-function mutations, or higher expression level. Then, for each SLGI, its predicted likelihood may be re-weighted by the minimum activity score of the two partner genes. The accuracy of this tumor-specific SLGI prediction may be evaluated by cross validation as described below.

The present computational algorithm provides significant advantages over prior art SLGI prediction algorithms (see e.g., reference 20) in a number of ways. First, the regression model

may consider many more public data and features and use feature selection to select those that are associated with SLGI. Second, weights may be given to the response variable in the different training data based on the confidence and strength of the observed SLGI. Finally, instead of using a number of Wilcoxon rank sum tests to filter gene pairs which could falsely remove promising pairs on one specific feature (as described in reference 20), the present multiple regression model automatically assigns feature weights, removes redundant features, and assigns a quantitative confidence for each prediction.

Example 8: Cross-validation to Systematically Evaluate New Algorithm Performance

Data has been collected for yeast-to-human SLGI pairs, as well as human SLGI pairs identified in previous literature studies, shRNA screens, and CRISPR screens on isogenic cell lines. The above-described TSG anchored SLGI genome-wide screening data may provide one additional high quality dataset with which to further evaluate the new SLGI prediction algorithm. The performance of the new algorithm may be systematically validated through a three-fold cross-validation (CV) procedure. The algorithm may initially be trained based on two-third SLGI pairs and used to predict the likelihood of SLGIs for the one-third held-out data and to then evaluate the prediction accuracy. In addition, CV may also be done by leaving one data set (e.g., an isogenic cell line screen for one TSG) out to validate the models trained on all other data sets. Based on the CV R2 metric, the SLGI prediction performance may be further compared between the new algorithm disclosed herein and previous algorithms (see e.g., reference 20 and 16).

In addition to evaluating SLGI prediction performance, the CV R2 metric may also be used to estimate the effect of down-sampling pgRNA pair number. Using the above-described pgRNA screening data, the number of pgRNAs for each gene pairs may be down-sampled and used to compute the CV R2 metric. If a significant deterioration of CV R2 is observed at certain pgRNA number, a higher number of pgRNA may be used in a design for large scale validation.

The new computational algorithm described above may be further refined to predict SLGI pairs in the human genome by integrating existing SLGI knowledge, high throughput SLGI identification data from previous literature and CRISPR screens, as well as TCGA data. The above described techniques may also be used for high throughput experimental validation of predicted SLGI pairs, without anchoring on one TSG in isogenic cell lines. It should be noted that many cancer cell lines harbor mutations and CNVs already, and thus SLGI pairs with one

gene already mutated in these cell lines might display an unexpected behavior. For example, PTEN has a heterozygous deletion in the LNCaP cell line, so genes with SLGI with PTEN might not show a strong difference in phenotype between single KO and double KO (targeting PTEN and its SLGI partners) screens. Similarly, unique SLGI behavior may be observed between LNCaP (prostate) and ZR-75-1 (breast), not due to their tissue of origin, but due to the unique mutations intrinsic to these two cell lines. Thus, when using the cell line screening data to either train or validate the new computational algorithm, it is necessary to consider the confounding effects of cell line specific genetic backgrounds. Since the somatic mutation and copy number information for most COSMIC cell lines are measured (see e.g., reference 36), it may be necessary to remove genes mutated or deleted in a cell line in the process of computational method training and validation.

Example 9: Expanded SLGI Knowledge Base

As described above, initial screens only tested the potential SLGI between 4 tumor suppressor genes (TSGs) and about 700 druggable genes. Many other TSGs are frequently lost as a result of mutation/deletion/inactivation in many cancers, and it has not been possible so far to restore their functions in the clinic. Therefore, it is critical to identify the SLGI partners of TSGs, which may enable therapies to treat cancers with TSG loss. The novel TSG SLGI partners identified without available inhibitors may be important new targets for drug development. Furthermore, different cancers and different tumors of the same cancer type likely have distinct transcriptome and mutation profiles, which may lead to cancer- or tumor-specific SLGI pairs. The above described SLGI-prediction algorithm has the advantage of being able to account for these differences by integrating cancer-specific and cell-specific genetic alteration and gene expression, among other factors, into the prediction of new SLGI pairs.

Example 10: Large Scale SLGI Screening Across Five Cancer Types

The techniques described herein may generate pan-cancer, cancer-specific as well as cell line-specific SLGI across all the human genome across all TCGA cancer types. To systematically evaluate the predictions, especially the novel ones, a CRISPR SLGI screening strategy targeting specific gene pairs predicted by our algorithm may be used in about 20 cancer cells across about 5 cancer types. The pgRNA screening library may include candidate pan-

cancer, cancer-specific, as well as cell-specific SLGI pairs involving ~50 TSGs, consisting of ~4K pairs across different scores of prediction confidence. More pgRNA pairs may be designed to target the more confident predictions, and the specific number of pgRNA pairs as well as the number of pgRNAs / pair in the CRISPR library design may be based on the power analysis described above. pgRNA CRISPR library construction and screening may be done as described above. The analysis to call SLGI depends on the number of predicted SLGI partners tested in the pgRNA CRISPR screen: a regression residual approach may be used for TSGs with many tested partners, while a BLISS independence model may be used for TSGs with fewer tested partners.

The results of these screens may significantly expand our knowledge of SLGI in different cancers and reveal potential novel therapy targets in cancers with non-targetable loss-of-function mutations. Additionally, examining the SLGI hits within the predicted pan-cancer SLGI, cell-specific SLGI, and non-SLGI may further evaluate the sensitivity and specificity of the new prediction algorithm, and assess its general applicability in target identification of cancer. Furthermore, the data generated herein may also serve as new training data to refine our algorithm.

Example 11: Characterizing the Mechanisms of Pan-Cancer and Cell-Specific SLGIs

Based on the above-validated SLGIs, two SLGI pairs each in the pan-cancer or cell-specific categories (FIG. 8A) may be selected and assessed for their respective mechanisms. Priority for selection may be given to novel SLGI pairs with frequent TSG loss in cancers and partners with available inhibitors. For the selected SLGI pairs with TSG "A" and druggable gene "B," small molecule inhibitors against B may be tested to determine if they have stronger killing in the cells harboring inactivating mutations in TSG "A." In addition, RNA-seq may be performed on unperturbed, gene "A" single KO, gene "B" single KO, or double "A+B" KO in two cell lines of different cancer type, respectively. Analysis of the RNA-seq may identify the transcriptome programs uniquely altered in the double KO condition, which might underlie the SLGI in different cancers or cell lines. Some pathways essential for cell survival or proliferation may remain unaffected or even activated with single gene KO, but be inactivated or inhibited with double KO in the SLGI pair. This may be assessed by validation assays. For example, in the case of a specific pan-cancer SLGI pair with TSG A and partner B, literature and pathway analysis may be conducted to examine whether the two genes share downstream pathways. If so,

such pathway activity may be tested to determine if it is significantly altered only when both A and B are deficient and whether modulating its activity can influence the synthetic lethality (FIG. 8B). From the RNA-seq profiles above, perturbed pathways may be assessed by enrichment algorithms such as GSEA (see e.g., reference 41), GO analysis (see e.g., reference 42), and GREAT (see e.g., reference 43). Usually many of the downstream genes will also be SLGI hits, albeit weaker, which may be confirmed either from predictions or from available CRISPR screening results. NEST (see e.g., reference 44) analysis may be applied to determine whether SLGI prediction or differentially expressed genes are enriched for PPI members. The identified pathways serve as putative mediator(s) of SLGI, and may be assessed by genetic or pharmacological modulations.

For a cancer type-specific SLGI pair with TSG "C" and partner "D" genes, there are two general scenarios: the "CD" downstream pathway is differentially expressed (FIG. 8C left), or they are similarly expressed but differentially required (FIG. 8C right). Published CRISPR screens have shown that if members of a protein complex involving gene "D" are all up-regulated in expression, "D" can be more essential without being differentially expressed (see e.g., reference 44). RNA-seq profiles may pinpoint the underlying scenario. For "CD" downstream pathways differentially required, a NEST analysis may be applied to the expression data to examine whether the differential expression of the PPI partners of "C" or "D" cause them to be differentially required. For "CD" downstream pathways differentially expressed, the expression profile and transcriptional regulatory network may be used to identify their upstream regulators that are differentially expressed in different cancers. These techniques may utilize any of a variety of algorithms (e.g., MACS 45, Cistrome AP 46, RABIT 39, MARGE 47, and the like) and databases (e.g., Cistrome DB 48) for transcription regulation. Identified transcriptional regulators that underlie the differential pathway may be verified by using genetic perturbation to verify their role in mediating the cancer type-specific SLGI relationship.

It should be noted that double KO of SLGI genes may lead to dramatic cell death and/or senescence by definition. Consequently, gene expression profiling of such double KO cells may become technically infeasible with limited cell number. To overcome this caveat, alternative approaches may be used to perturb the candidate genes for expression profiling such as, for example, by (inducible) RNAi or small molecule inhibition. However, it is not uncommon for RNAi and small molecule inhibitors to have pleiotropic or off-target effects, so it is possible that

different phenotypes may be observed between functional validations using shRNA and/or small molecule inhibitors versus pgRNA-mediated double KO. To ensure robustness of the validation, multiple small molecule inhibitors or multiple shRNAs may be tested against the candidate genes. Additionally, exome and cistrome genotypes in these cancer cell lines may be the confounding factors that affect the interpretation of the SLGI screening data, so choosing cancer cell lines that have exome sequencing and copy number variation data available from COSMIC and CCLE to ensure that this information could be taken into consideration.

Example 11: A paired-guide (pgRNA) CRISPR Library for Functional Enhancer Screen

The techniques herein also provide that a paired-guide CRISPR library may be used to conduct functional enhancer screen(s). As shown in FIG. 9A, the rationale of the strategy is that two gRNAs may be introduced into a single cell, and if the two targeting loci are close to each other, then the fragment in-between has a high probability of being deleted, rather than having two indels mutation at each of the two loci separately. Because the deletion could affect larger regions than small indel mutations, the techniques herein provide that a small number of pgRNAs may be used to cover much larger regions of the genome than sgRNA libraries. Furthermore, since the deletion could completely knock out the putative functional motifs of an enhancer, the efficiency is also higher.

For an enhancer screen experiment, a small pgRNA library containing 7500 pairs of guide RNAs was designed for use in screening in an ER+ breast cancer cell line: T47D. This line had previously been used to conduct a genome-wide CRISPR screens. In an exemplary embodiment, the distance range between the two gRNAs was between 150-300 bp.

This library was designed to target three groups of predicted cis-elements: 1) Enhancers and promoters of positively-selected genes: PTEN, TSC1, RB1, CSK (tilling arrays); 2) Enhancers and promoters of negatively-selected genes: ESR1, MYC, GATA3, FOXA1; and 3) A short list of CTCF and FOXA1 binding sites from the sgRNA CRISPR library. An overview of the screening procedure is shown in FIG. 9B, in which the cell libraries were cultured for 30 days under three conditions: full medium, white medium and white medium + Estrogen (E2) before harvested for genomic DNA and sequencing of the pgRNAs together with the Day 0 cell library sample as control. Negative controls used in the enhancer screen included double cuts on

AAVS1, whereas positive controls used in the enhancer screen included double cuts on an essential gene + AAVS1.

As shown in FIG. 10, CSK is an important positively-selected gene in T47D and MCF7 cell lines under hormone-depleted growth condition (also shown in FIG.2C). Knockout of the putative CSK enhancer with ER binding and DNase-I/H3K27ac mark totally abolished CSK expression upon estrogen stimulus (FIG. 10 right panel). Therefore, CSK enhancer loss reconstructs the CSK-knockout phenotype under estrogen-depleted growth condition.

FIG. 11 shows an exemplary tilling design to target the CSK enhancer, in which more than 1,300 pgRNAs were designed in a tilling format to cover the CSK enhancer region in which each pgRNA flanks 150-300 bp locus to search for novel and unknown CSK enhancers.

The CSK enhancer tilling design shown in FIG. 11 was analyzed by a modified MAGeCK algorithm with conversion of pgRNAs into consecutive bins of DNA locus, result in a representative p-value plot of each bin to show a potential functional enhancer, as shown in FIG. 12.

As shown in FIG. 13, the functional enhancer screen successfully identified known CSK enhancers, as well as potentially novel enhancer elements. As the positive selection p-value plot shows, the three peaks represent one functionally validated CSK enhancer co-localized with DNase-I/H3K27ac mark and ESR1-binding peak (FIG. 10) and two previously unknown enhancers with only H3K27ac marks.

REFERENCES

1. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science 2010; 327:425–31.
2. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science 2008;320:362–5.
3. Srivas R, Shen JP, Yang CC, Sun SM, Li J, Gross AM, et al. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Mol Cell 2016;63:514–25.
4. Steckel M, Molina-Arcas M, Weigelt B, Marani M, Warne PH, Kuznetsov H, et al. Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. Cell Res 2012;22:1227–45.

5. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell 2009;137:835–48.
6. Bommi-Reddy A, Almeciga I, Sawyer J, Geisen C, Li W, Harlow E, et al. Kinase requirements in human cells: III. Altered kinase requirements in VHL-/- cancer cells detected in a pilot synthetic lethal screen. Proceedings of the National Academy of Sciences 2008;105:16484–9.
7. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, , et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014;343:84–7.
8. Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, et al. Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell 2017. https://doi.org/10.1016/j.cell.2017.01.013.
9. Bassik MC, Kampmann M, Lebbink RJ, Wang S, Hein MY, Poser I, et al. A systematic mammalian genetic interaction map reveals pathways underlying ricin susceptibility. Cell 2013;152:909–22.
10. Wong ASL, Choi GCG, Cui CH, Pregernig G, Milani P, Adam M, et al. Multiplexed barcoded CRISPR-Cas9 screening enabled by CombiGEM. Proc Natl Acad Sci U S A 2016;113:2544–9.
11. Laufer C, Fischer B, Billmann M, Huber W, Boutros M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. Nat Methods 2013;10:427–31.
12. Chipman KC, Singh AK. Predicting genetic interactions with random walks on biological networks. BMC Bioinformatics 2009;10:17.
13. Kelley R, Ideker T. Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol 2005;23:561–6.
14. Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet 2011;43:656–62.
15. Wong SL, Zhang LV, Tong AHY, Li Z, Goldberg DS, King OD, et al. Combining biological networks to predict genetic interactions. Proc Natl Acad Sci U S A 2004;101:15682–7.

16. Conde-Pueyo N, Munteanu A, Sole RV, Rodríguez-Caso C. Human synthetic lethal inference as potential anti-cancer target gene detection. BMC Syst Biol 2009;3:116.
17. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. Mol Syst Biol 2011;7:501.
18. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, et al. Haem oxygenase is synthetically lethal with the tumor suppressor fumarate hydratase. Nature 2011;477:225–8.
19. Lu X, Kensche PR, Huynen MA, Notebaart RA. Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. Nat Commun 2013;4:2124.
20. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell 2014;158:1199–209.
21. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. Science 2014;343:80–4.
22. Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC, Yusa K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat Biotechnol 2014;32:267–73.
23. Zhu S, Li W, Liu J, Chen C-H, Liao Q, Xu P, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nat Biotechnol 2016;34:1279–86.
24. Vidigal JA, Ventura A. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. Nat Commun 2015;6:8083.
25. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell 2016;167:1867– 82.e21.
26. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science 2013;339:819–23.
27. Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature 2014;509:487–91.
28. Sack LM, Davoli T, Xu Q, Li MZ, Elledge SJ. Sources of Error in Mammalian Genetic Screens. G3 2016;6:2781–90.

29. Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. Virus Res 2012;169:415–29.
30. Fernandez-Banet J, Esposito A, Coffin S, Horvath IB, Estrella H, Schefzick S, et al. OASIS: web-based platform for exploring cancer multi-omics data. Nat Methods 2016;13:9–10.
31. Xu H, Xiao T, Chen C-H, Li W, Meyer CA, Wu Q, et al. Sequence determinants of improved CRISPR sgRNA design. Genome Res 2015;25:1147–57.
32. Bliss CI. THE TOXICITY OF POISONS APPLIED JOINTLY1. Ann Appl Biol 1939;26:585–615.
33. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 2014;15:554.
34. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, et al. Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. Cancer Discov 2016;6:914–29.
35. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603–7.
36. Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. Cell 2016;166:740–54.
37. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. Nat Biotechnol 2014;32:1262–7.
38. Li W, Köster J, Xu H, Chen C-H, Xiao T, Liu JS, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol 2015;16:281.
39. Jiang P, Freedman ML, Liu JS, Liu XS. Inference of transcriptional regulation in cancers. Proceedings of the National Academy of Sciences 2015;112:7731–6.
40. Guo J, Liu H, Zheng J. SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. Nucleic Acids Res 2016;44:D1011–7.
41. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.

42. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.
43. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 2010;28:495–501.
44. Jiang P, Wang H, Li W, Zang C, Li B, Wong YJ, et al. Network analysis of gene essentiality in functional genomics experiments. Genome Biol 2015;16:239.
45. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
46 Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: an integrative platform for transcriptional regulation studies. Genome Biol 2011;12:R83.
47. Wang S, Zang C, Xiao T, Fan J, Mei S, Qin Q, et al. Modeling cis-regulation with a compendium of genome-wide histone H3K27ac profiles. Genome Res 2016;26:1417–29.
48. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C, et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res 2017;45:D658–62.

INCORPORATION BY REFERENCE

All documents cited or referenced herein and all documents cited or referenced in the herein cited documents, together with any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated by reference, and may be employed in the practice of the disclosure.

EQUIVALENTS

It is understood that the detailed examples and embodiments described herein are given by way of example for illustrative purposes only, and are in no way considered to be limiting to the disclosure. Various modifications or changes in light thereof will be suggested to persons skilled in the art and are included within the spirit and purview of this application and are considered within the scope of the appended claims. Additional advantageous features and functionalities associated with the systems, methods, and processes of the present disclosure will

be apparent from the appended claims. Moreover, those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the disclosure described herein. Such equivalents are intended to be encompassed by the following claims.

CLAIMS

We claim:

1. A paired-guide ribonucleic acid (pgRNA) vector, comprising:

a first guide RNA (gRNA) cassette;

a second gRNA cassette; and

a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated protein 9 (Cas9) expression cassette;

wherein the second gRNA cassette is positioned between the first gRNA cassette and the Cas9 expression cassette.

2. The pgRNA vector of claim 1, wherein the first gRNA cassette includes a first nucleic acid sequence including, in 5' to 3' order, a first gRNA promoter, a first gRNA, and a first gRNA scaffold, and the second gRNA cassette includes a second nucleic acid sequence including, in 5' to 3' order, a second gRNA promoter, a second gRNA, and a second gRNA scaffold.

3. The pgRNA vector of claim 2, wherein the first gRNA promoter is selected from the group consisting of a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and a modified bovine 7SK promoter.

4. The pgRNA vector of claim 2, wherein the second gRNA promoter is selected from the group consisting of a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and a modified bovine 7SK promoter.

5. The pgRNA vector of claim 2, wherein the second gRNA promoter is different than the first gRNA promoter.

6. The pgRNA vector of claim 2, wherein the first gRNA and the second gRNA are each between about 17 and 27 nucleotides in length.

7. The pgRNA vector of claim 2, wherein the first gRNA and the second gRNA are each about 19 nucleotides in length.

8. The pgRNA vector of claim 1, wherein the pgRNA vector is constructed by using an intermediate pgRNA nucleic acid, comprising:

a first guide RNA (gRNA);

a unique linker; and

a second gRNA;

wherein the unique linker is positioned between the first gRNA and the second gRNA .

9. The pgRNA vector of claim 8, wherein the unique linker is about 16 nucleotides in length.

10. The pgRNA vector of claim 1, wherein the Cas9 cassette includes a promoter, a Cas9 coding sequence, and a P2A sequence.

11. A method of making a paired-guide RNA (pgRNA) library vector, comprising:

obtaining a first nucleic acid sequence including, in 5' to 3' order, a first guide RNA (gRNA) cassette promoter, a vector linker, and a second gRNA cassette scaffold;

removing the vector linker to create a double strand break (DSB) between a 3' end of the first gRNA cassette promoter and a 5' end of the second gRNA cassette scaffold;

inserting into the DSB a second nucleic acid sequence including, in 5' to 3' order, a first guide RNA (gRNA) sequence, a unique linker, and a second gRNA sequence to create an intermediate nucleic acid sequence;

removing the unique linker to create a DSB in the intermediate nucleic acid sequence between a 3' end of the first gRNA sequence and a 5' end of the second gRNA sequence;

inserting into the DSB in the intermediate nucleic acid sequence a third nucleic acid sequence including, in 5' to 3' order, a first gRNA cassette scaffold, a spacer, and a second guide RNA (gRNA) cassette promoter, thereby creating the pgRNA vector.

12. The method of claim 11, wherein the first gRNA cassette promoter is selected from the group consisting of a mouse U6 promoter and a human U6 promoter.

13. The method of claim 11, wherein the second gRNA cassette promoter is selected from the group consisting of a mouse U6 promoter and a human U6 promoter.

14. The method of claim 13, wherein the second gRNA cassette promoter is different than the first gRNA cassette promoter.

15. The method of claim 11, wherein the first gRNA sequence and the second gRNA sequence are each between about 17 and 27 nucleotides in length.

16. The method of claim 15, wherein the first gRNA sequence and the second gRNA sequence are each about 19 nucleotides in length.

17. The method of claim 11, wherein the unique linker is between about 12 and 24 nucleotides in length.

18. The method of claim 11, wherein the unique linker is about 16 nucleotides in length.

19. The method of claim 11, wherein the first nucleic acid sequence further includes a Cas9 cassette.

20. The method of claim 19, wherein the Cas9 cassette includes a promoter, a Cas9 coding sequence, and a P2A sequence.

21. A paired-guide RNA (pgRNA)/Clustered Regularly-Interspaced Short Palindromic Repeats (CRISPR) library, comprising: a plurality of pgRNA sequence pairs capable of targeting a plurality of target sequence pairs in a target genome via a CRISPR/Cas9 system to knockout function of a first target sequence and a second target sequence in the target sequence pair, wherein pgRNA vector is constructed by using an intermediate pgRNA nucleic acid, that

includes a first guide RNA (gRNA); a unique linker; and a second gRNA; wherein the unique linker is positioned between the first gRNA and the second gRNA.

22. The pgRNA/CRISPR library of claim 21, wherein each of the plurality of pgRNA sequence pairs includes a first guide RNA (gRNA) cassette and a second gRNA cassette.

23. The pgRNA/CRISPR library of claim 22, wherein the first gRNA cassette includes a first nucleic acid sequence including, in 5' to 3' order, a first gRNA promoter, a first gRNA sequence, and a first gRNA scaffold, and the second gRNA cassette includes a second nucleic acid sequence including, in 5' to 3' order, a second gRNA promoter, a second gRNA sequence, and a second gRNA scaffold.

24. The pgRNA/CRISPR library of claim 23, wherein the first gRNA promoter is selected from the group consisting of a mouse U6 promoter and a human U6 promoter.

25. The pgRNA/CRISPR library of claim 23, wherein the second gRNA promoter is selected from the group consisting of a mouse U6 promoter and a human U6 promoter.

26. The pgRNA/CRISPR library of claim 25, wherein the second gRNA promoter is different than the first gRNA promoter.

27. The pgRNA/CRISPR library of claim 23, wherein the first gRNA sequence and the second gRNA sequence are each between about 17 and 27 nucleotides in length.

28. The pgRNA/CRISPR library of claim 23, wherein the first gRNA sequence and the second gRNA sequence are each about 19 nucleotides in length.

29. The pgRNA/CRISPR library of claim 21, wherein the unique linker is between about 12 and 24 nucleotides in length.

30. The pgRNA/CRISPR library of claim 21, wherein the unique linker is about 16 nucleotides in length.

31. A method of identifying synthetic lethal genetic interactions (SLGIs) or enhancers within a genome, comprising:

contacting a population of cells with one or more pgRNA vectors of claim 1;

selecting successfully transduced cells;

culturing the population of cells for a plurality of population doubling times, wherein genomic DNA is harvested on a first day of culture and on a last day of culture;

deep sequencing the genomic DNA harvested on the first day of culture and on the last day of culture;

quantifying abundance of the pgRNAs at the first day of culture and the last day of culture;

analyzing an abundance fold change of the pgRNAs between the first day of culture and the last day of culture; and

identifying, based on the abundance fold change; a SLGI or enhancer.

32. The method of claim 31, wherein analyzing further includes a regression residual analysis.

33. The method of claim 31, wherein analyzing further includes a BLISS independence model analysis.

34. The method of claim 31, wherein the plurality of population doubling times is between about 8 and 16.

35. The method of claim 31, wherein the plurality of population doubling times is about 12.

36. A tangible, non-transitory, computer-readable media having software encoded thereon, the software, when executed by a processor on a particular device, operable to:

identify a plurality of gene pairs;

determine a response variable;

analyze, by a feature selection and regression model, the plurality of gene pairs; and

determine, based on the response variable and the analysis, that one or more gene pairs within the plurality of gene pairs interact genetically.

37. An intermediate paired-guide RNA (pgRNA) nucleic acid, comprising:

a first guide RNA (gRNA) cassette;

a unique linker; and

a second gRNA cassette;

wherein the unique linker is positioned between the first gRNA cassette and the second gRNA cassette.

38. The intermediate pgRNA nucleic acid of claim 37, wherein the first gRNA cassette includes a first nucleic acid sequence including, in 5' to 3' order, a first gRNA promoter, a first gRNA, and a first gRNA scaffold, and the second gRNA cassette includes a second nucleic acid sequence including, in 5' to 3' order, a second gRNA promoter, a second gRNA, and a second gRNA scaffold.

39. The intermediate pgRNA nucleic acid of claim 37, wherein the first gRNA promoter is selected from the group consisting of a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and a modified bovine 7SK promoter.

40. The intermediate pgRNA nucleic acid of claim 37, wherein the second gRNA promoter is selected from the group consisting of a mouse U6 promoter, a human U6 promoter, a modified bovine U6 promoter, a mouse H1 promoter, a human H1 promoter, a mouse 7SK promoter, and a human 7SK promoter, and a modified bovine 7SK promoter.

41. The intermediate pgRNA nucleic acid of claim 37, wherein the second gRNA promoter is different than the first gRNA promoter.

42. The intermediate pgRNA nucleic acid of claim 37, wherein the first gRNA and the second gRNA are each between about 17 and 27 nucleotides in length.

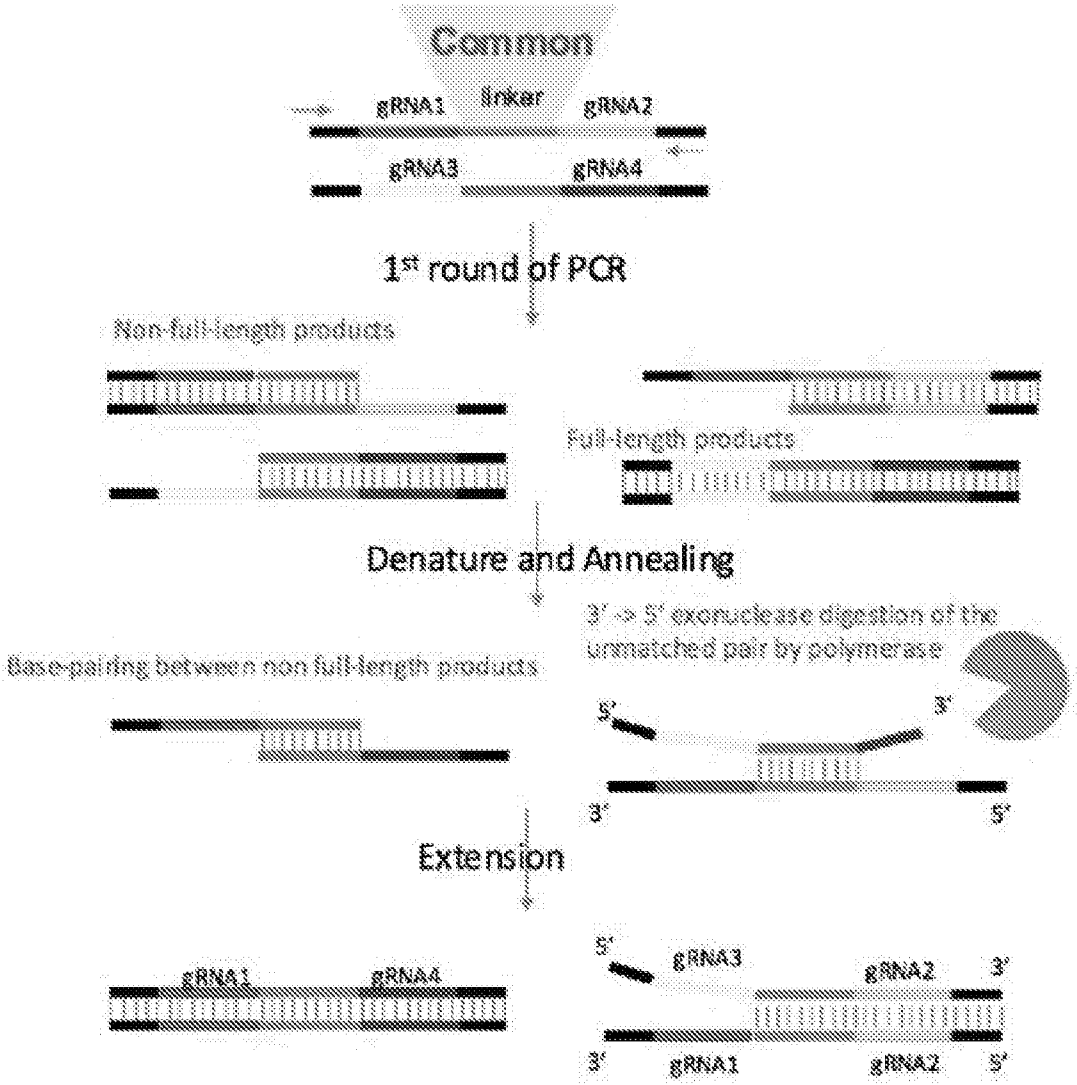
43. The intermediate pgRNA nucleic acid of claim 37, wherein the first gRNA and the second gRNA are each about 19 nucleotides in length.

44. The intermediate pgRNA nucleic acid of claim 37, wherein the unique linker is between about 10 and 30 nucleotides in length.

45. The intermediate pgRNA nucleic acid of claim 37, wherein the unique linker is about 16 nucleotides in length.

46. The intermediate pgRNA nucleic acid of claim 37, wherein the unique linker has a GC content less than or equal to 40%.

FIG. 1



Swapped pairs created

FIG. 2A

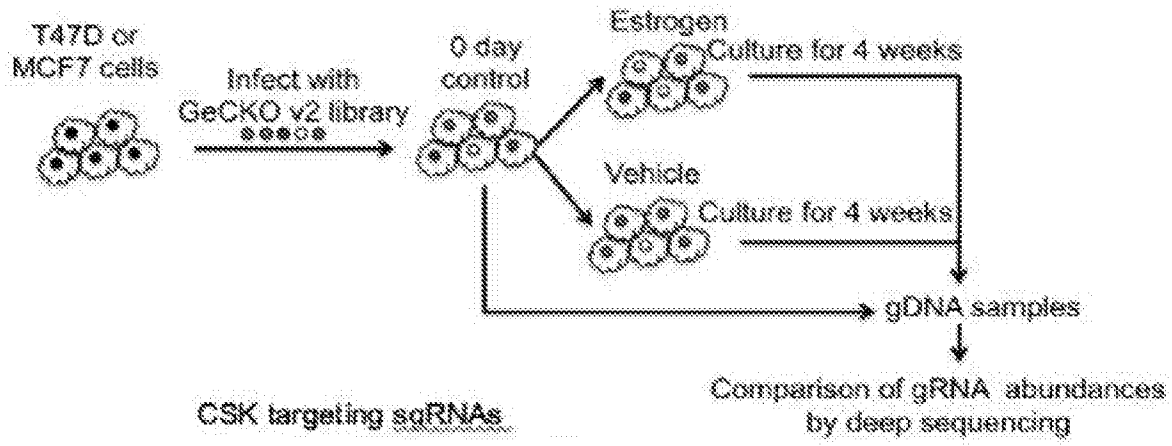

T47D or
MCF7 cells
Infect with
GeCKO v2 library
0 day
control
Estrogen
Culture for 4 weeks
Vehicle
Culture for 4 weeks
gDNA samples
Comparison of gRNA abundances
by deep sequencing
CSK targeting sgRNAs

FIG. 2B

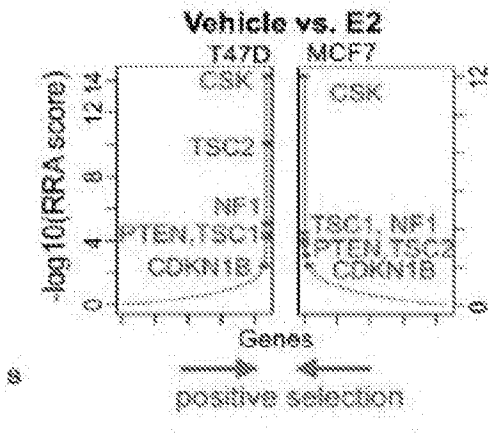

Vehicle vs. E2
T47D
MCF7
-log10(RRA score)
CSK
TSC2
NF1
PTEN,TSC1
CDKN1B
TSC1, NF1
PTEN,TSC2
Genes
positive selection

FIG. 2C

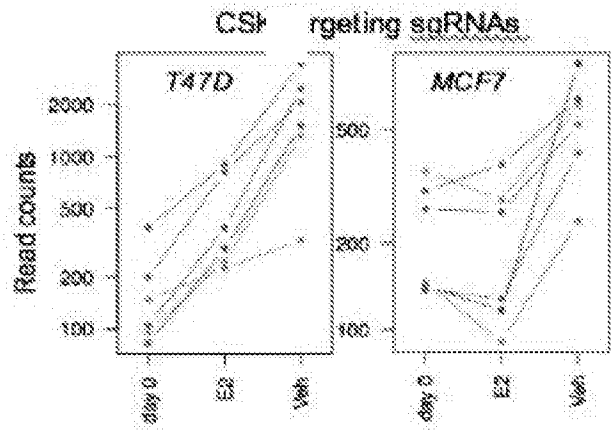

CSh rgeting saRNAs
T47D
MCF7
Read counts
2000
1000
500
200
100
500
200
100
day 0
day 3
day 6

FIG. 2D

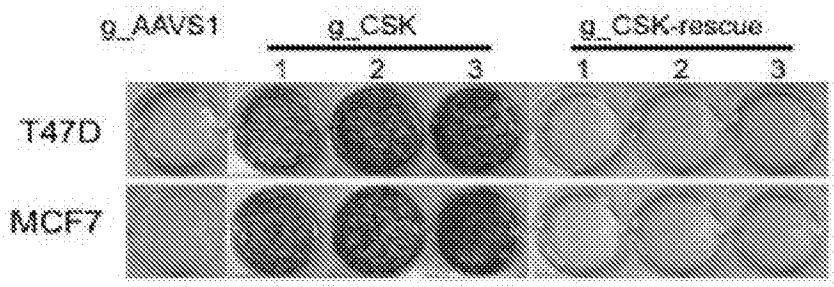


FIG. 2E

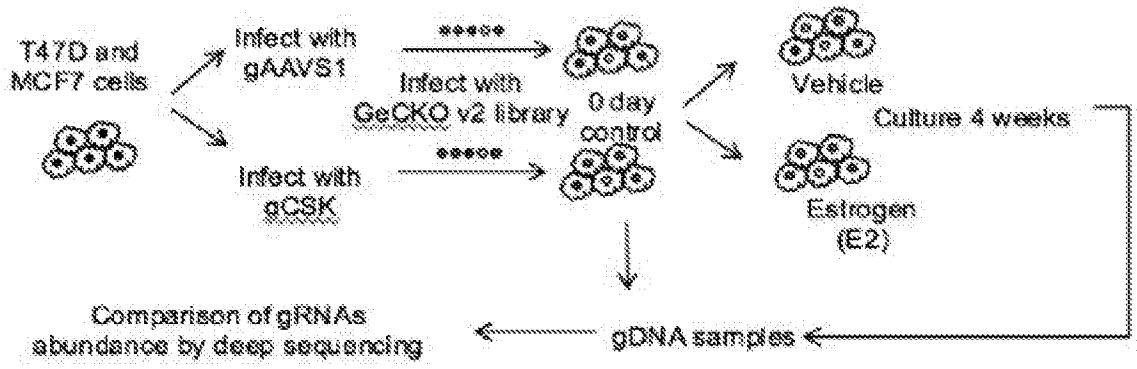

T47D and MCF7 cells
Infect with gAAVS1
Infect with GeCKO v2 library
Infect with gCSK
0 day control
Vehicle
Culture 4 weeks
Estrogen (E2)
gDNA samples
Comparison of gRNAs abundance by deep sequencing

FIG. 2F

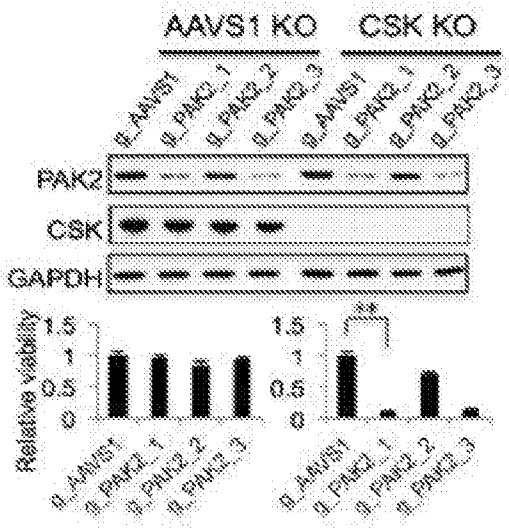

AAVS1 KO
CSK KO
g_AAVS1
g_PAK2_1
g_PAK2_2
g_PAK2_3
g_AAVS1
g_PAK2_1
g_PAK2_2
g_PAK2_3
PAK2
CSK
GAPDH
Relative viability
1.5
1
0.5
0
**
1.5
1
0.5
0
g_AAVS1
g_PAK2_1
g_PAK2_2
g_PAK2_3
g_AAVS1
g_PAK2_1
g_PAK2_2
g_PAK2_3

FIG. 3A

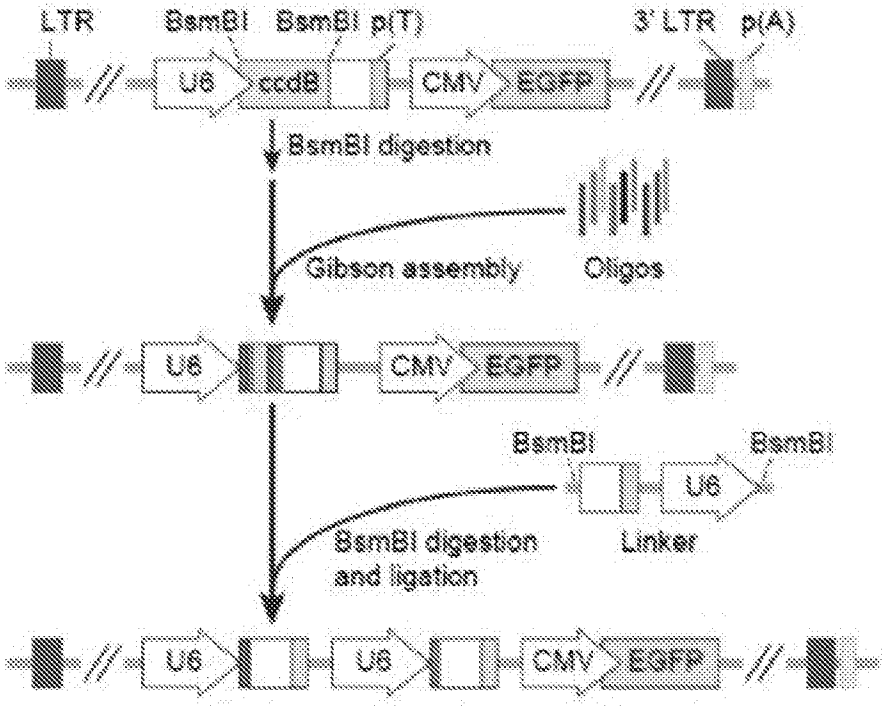

LTR
BsmBI
BsmBI
p(T)
3' LTR
p(A)
U6
ccdB
CMV
EGFP
BsmBI digestion
Gibson assembly
Oligos
U6
CMV
EGFP
BsmBI
BsmBI
U6
BsmBI digestion
and ligation
Linker
U6
U6
CMV
EGFP

FIG. 3B

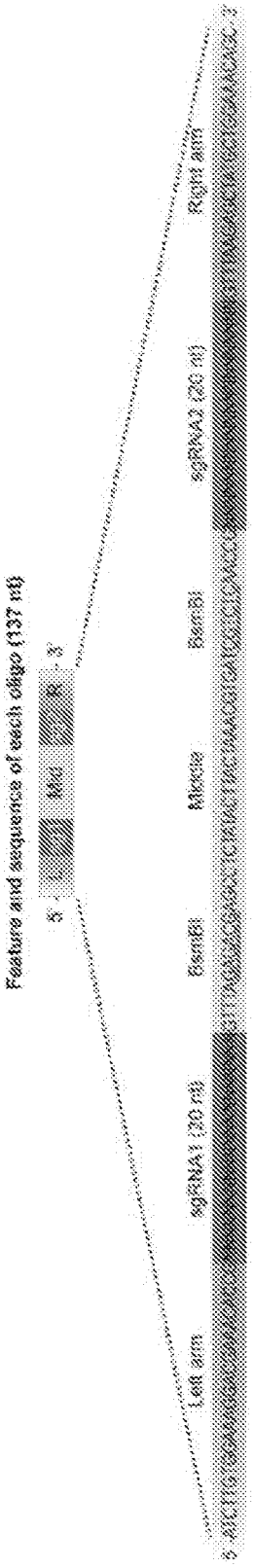

Feature and sequence of each oligo (137 nt)
5'
Mid
R
3'
Left arm
sgRNA1 (20 nt)
BsmBI
Mlucia
BsmBI
sgRNA2 (20 nt)
Right arm

FIG. 3C

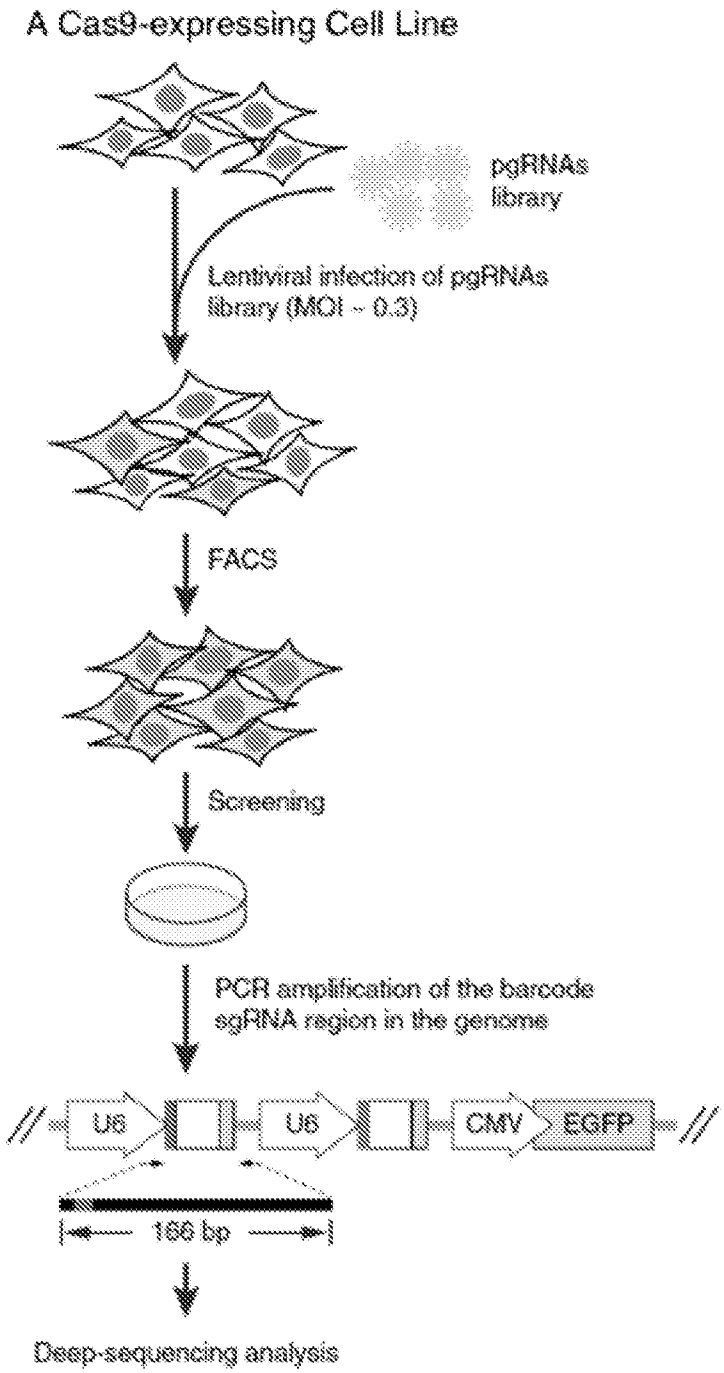

A Cas9-expressing Cell Line
pgRNAs library
Lentiviral infection of pgRNAs library (MOI ~ 0.3)
FACS
Screening
PCR amplification of the barcode sgRNA region in the genome
U6
U6
CMV
EGFP
166 bp
Deep-sequencing analysis

FIG. 3D

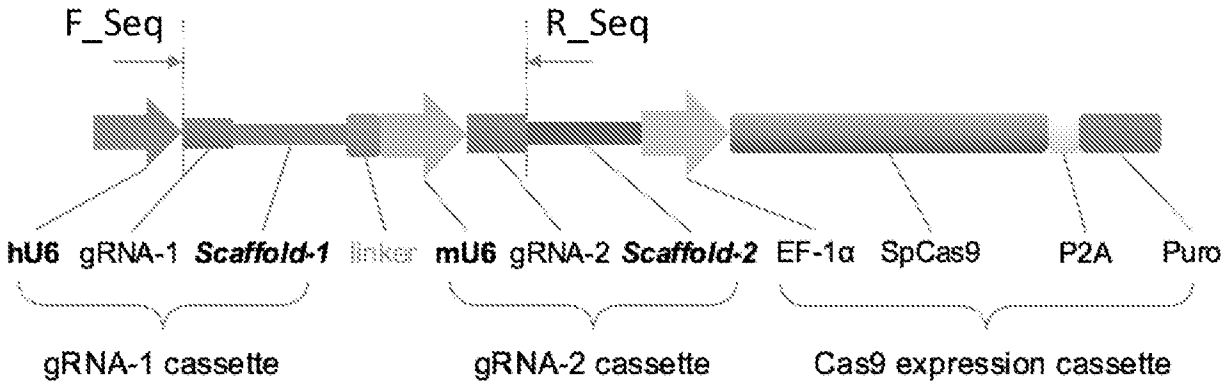

F_Seq
R_Seq
hU6 gRNA-1 Scaffold-1
linker
mU6 gRNA-2 Scaffold-2
EF-1α SpCas9
P2A
Puro
gRNA-1 cassette
gRNA-2 cassette
Cas9 expression cassette

FIG. 3E

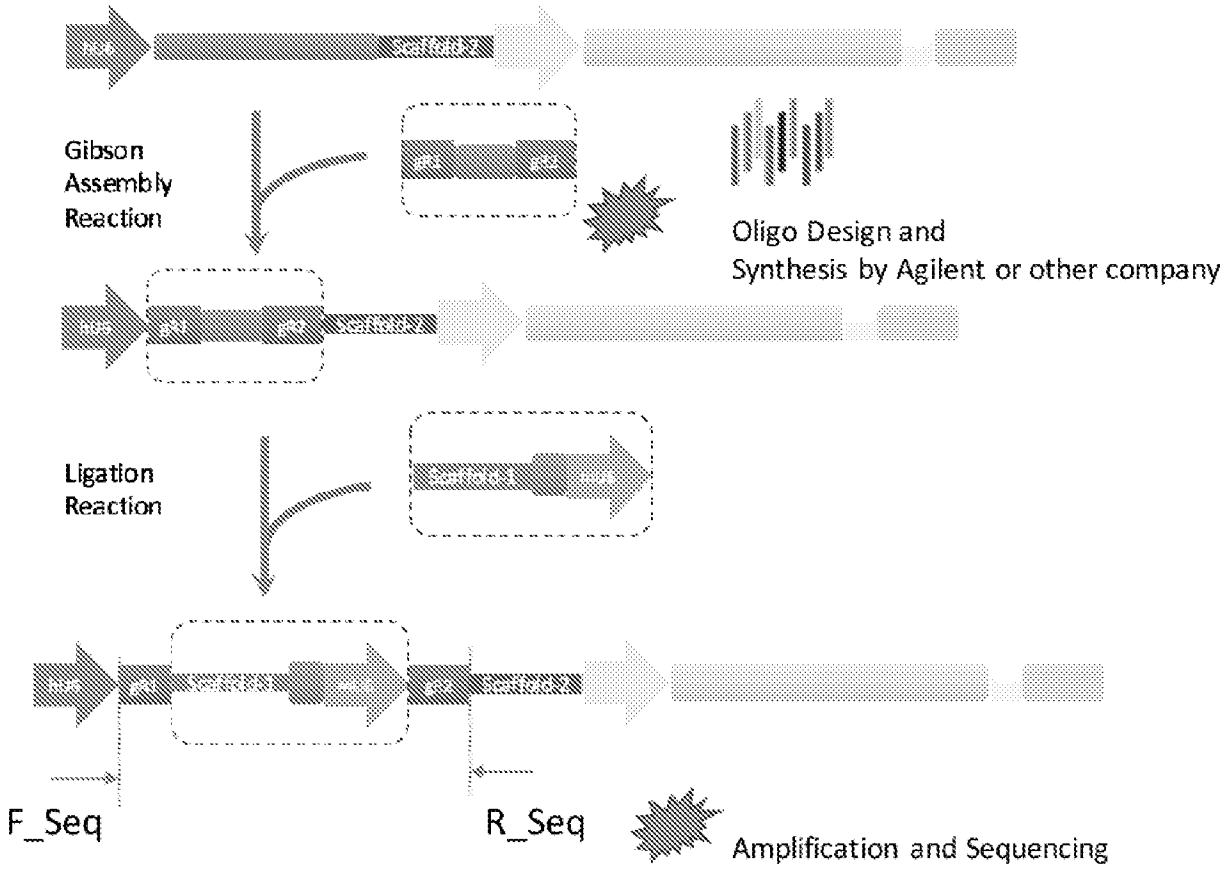

Gibson
Assembly
Reaction
Oligo Design and
Synthesis by Agilent or other company
Ligation
Reaction
F_Seq
R_Seq
Amplification and Sequencing

FIG. 3F

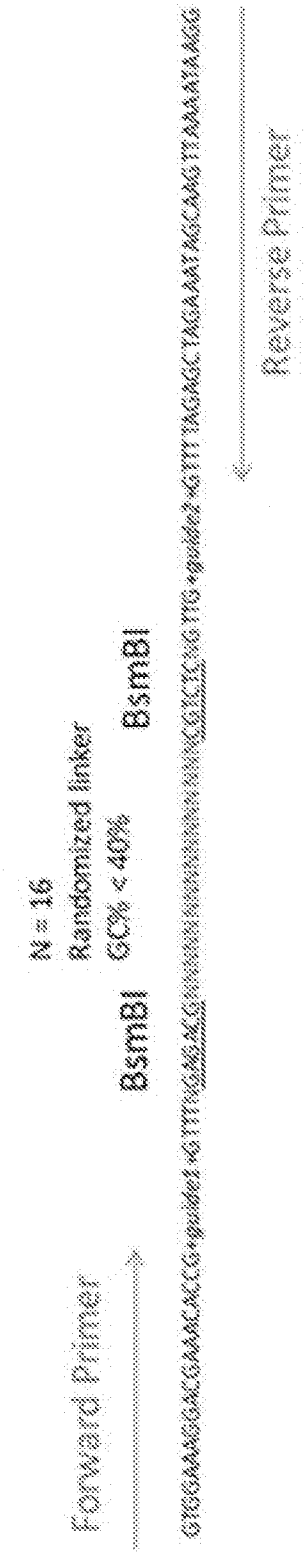
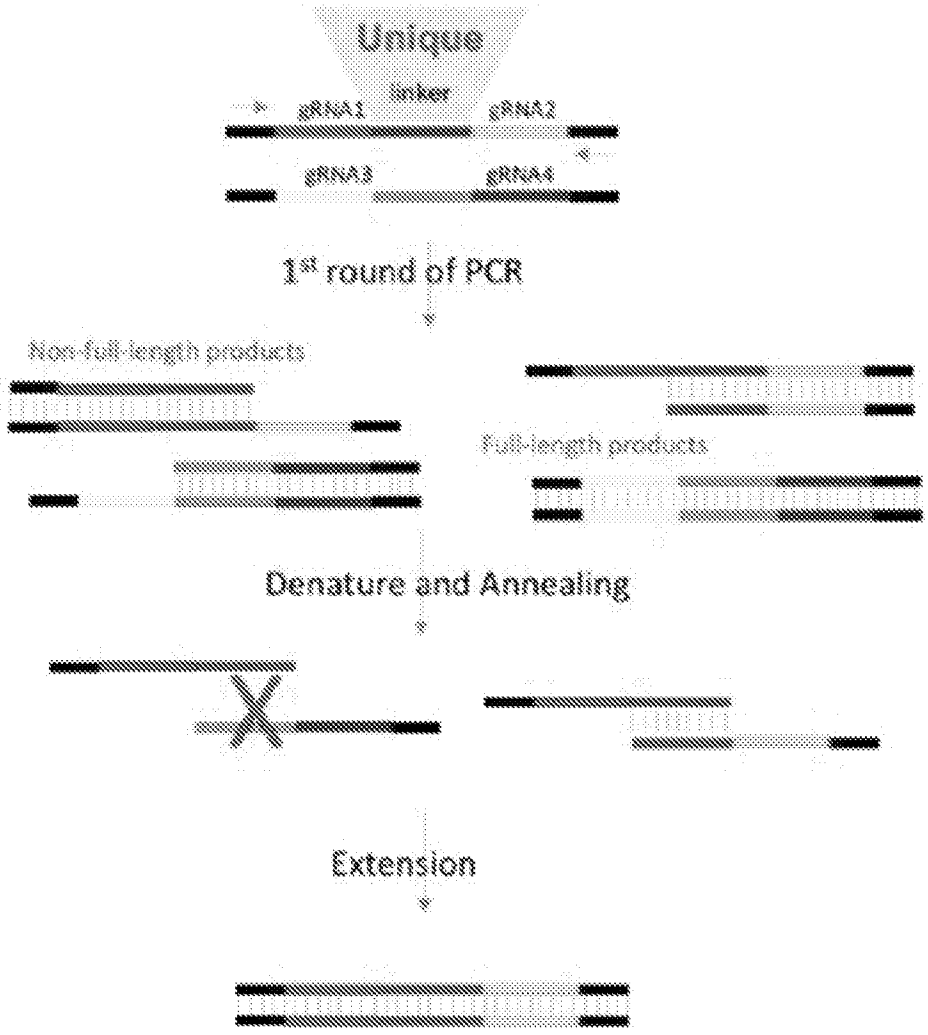

N = 16
Randomized linker
GC% < 40%
BsmBI
BsmBI
Forward Primer
Reverse Primer

FIG. 3G



Faithful Amplification of the Oligo Pool

FIG. 3H

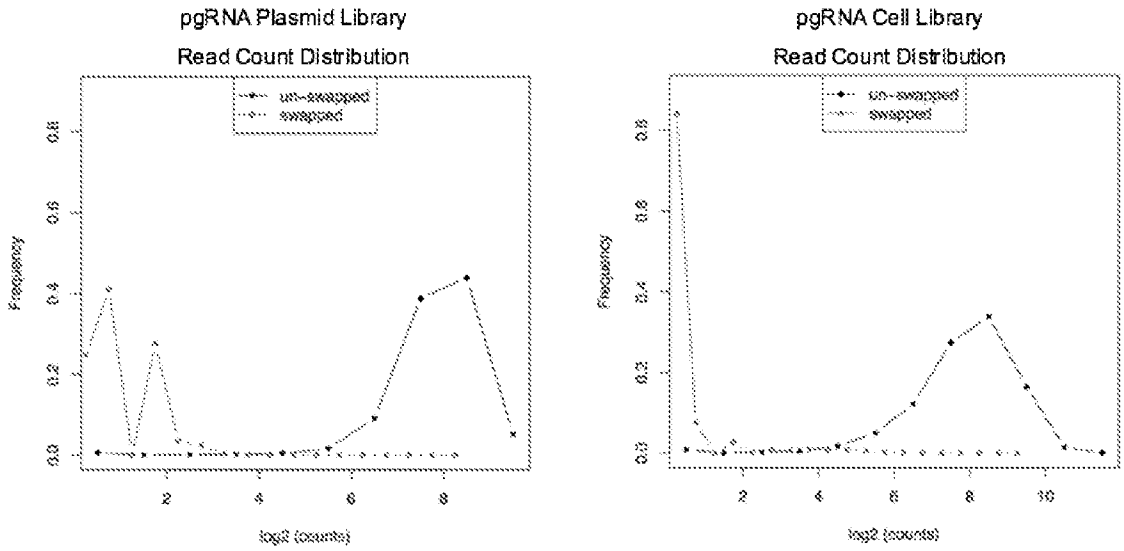

pgRNA Plasmid Library
Read Count Distribution
un-swapped
swapped
Frequency
log2 (counts)
pgRNA Cell Library
Read Count Distribution
un-swapped
swapped
Frequency
log2 (counts)

FIG. 3I

Sanger Sequencing Summary of the Colony PCR Amplicons from the pgRNA Library

Category	Pairs with Both Correct gRNAs	Correct Pairs	Swapped Pairs	Swapping Rate
Number or Ratio	12	12	0	0%

FIG. 4

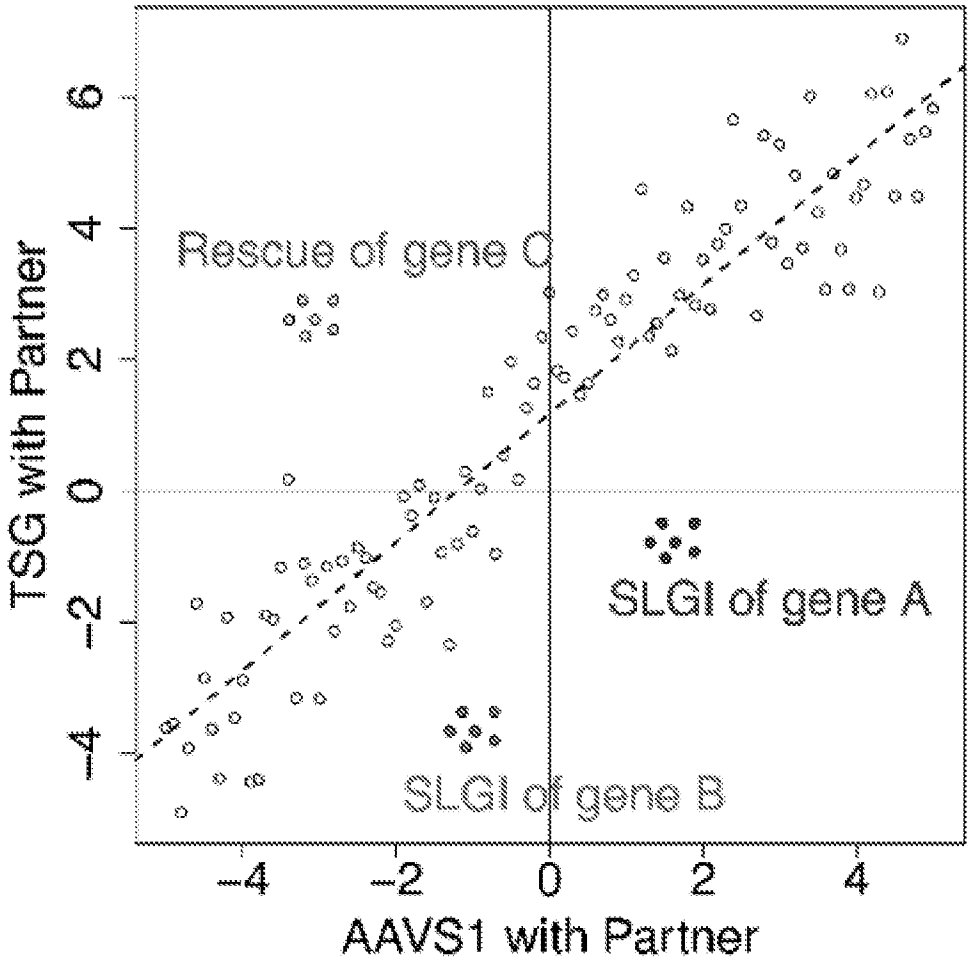

Rescue of gene C
SLGI of gene A
SLGI of gene B
TSG with Partner
AAVS1 with Partner
6
4
2
0
−2
−4
−4
−2
0
2
4

FIG. 5A

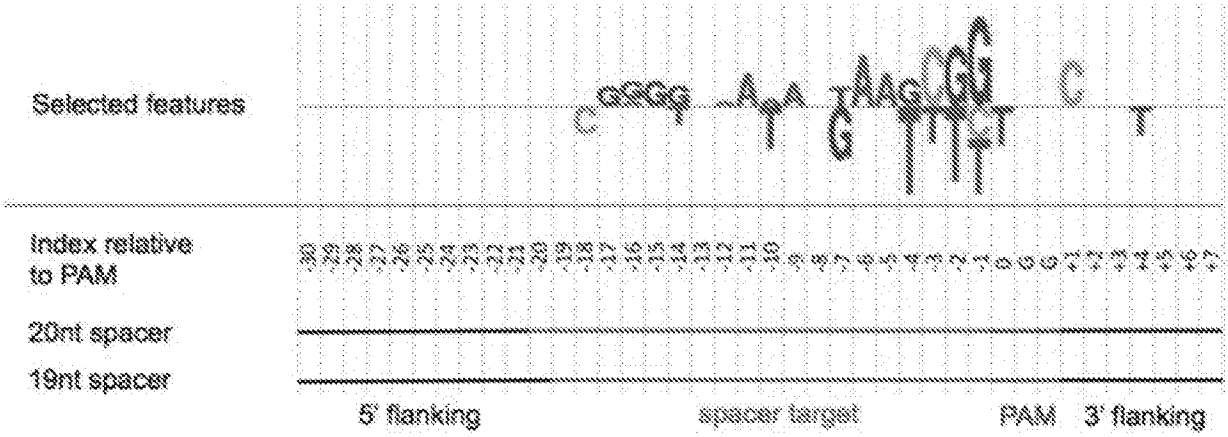

Selected features
Index relative to PAM
20nt spacer
19nt spacer
5' flanking
spacer target
PAM
3' flanking

FIG. 5B

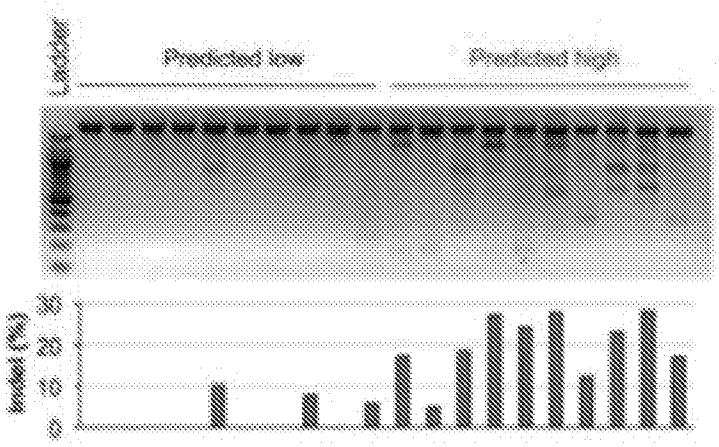


FIG. 5C

Sample	0 day	X	Y	Z	
A	1	0	0	0	...
B	1	1	0	0	...
C	1	1	0	0	...
D	1	0	1	0	...
E	1	0	1	0	...
F	1	0	1	1	...
G	1	0	1	1	...
H	...	...	...	...	...
I	...	...	...	...	

FIG. 5D

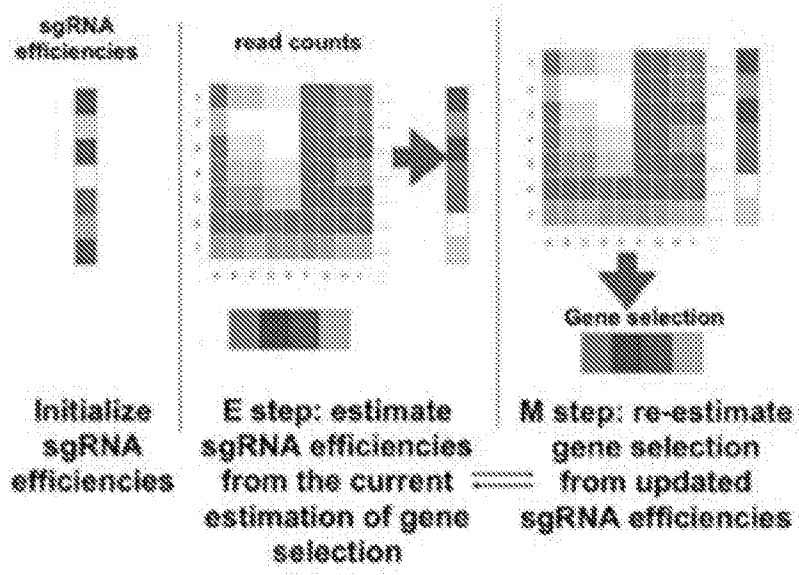

sgRNA efficiencies
read counts
Gene selection
Initialize sgRNA efficiencies
E step: estimate sgRNA efficiencies from the current estimation of gene selection
M step: re-estimate gene selection from updated sgRNA efficiencies

FIG. 6

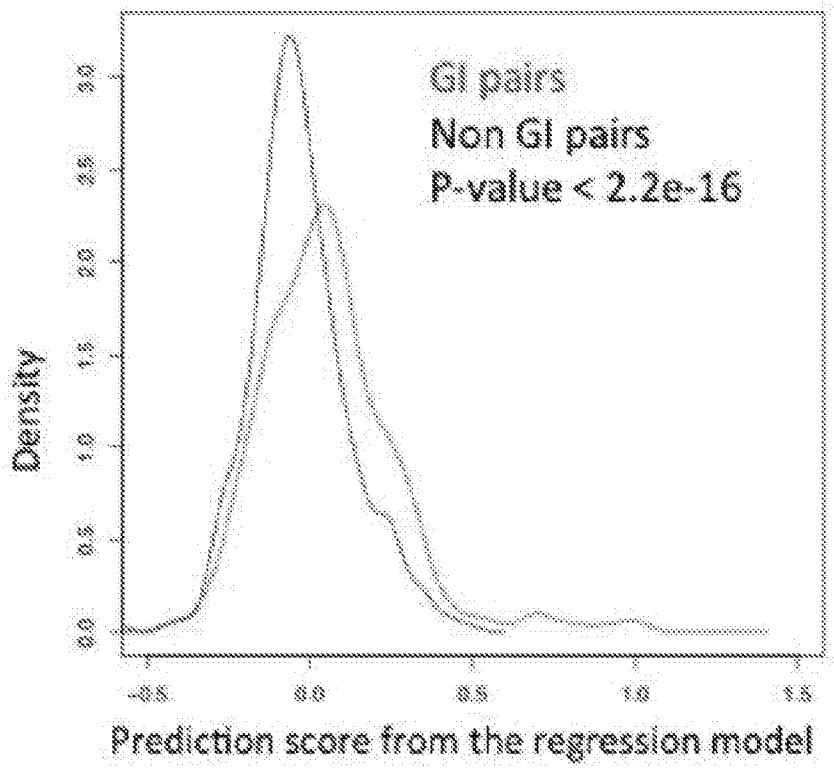

GI pairs
Non GI pairs
P-value < 2.2e-16
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-0.5
0.0
0.5
1.0
1.5
Prediction score from the regression model

FIG. 7

$$Y = X\beta + \epsilon$$

$$W = \begin{bmatrix} w_1 I_1 & & \\ & \ddots & \\ & & w_d I_d \end{bmatrix}$$

$$\hat{\beta}_W = (X^T W X)^{-1} X^T W Y$$

FIG. 8A

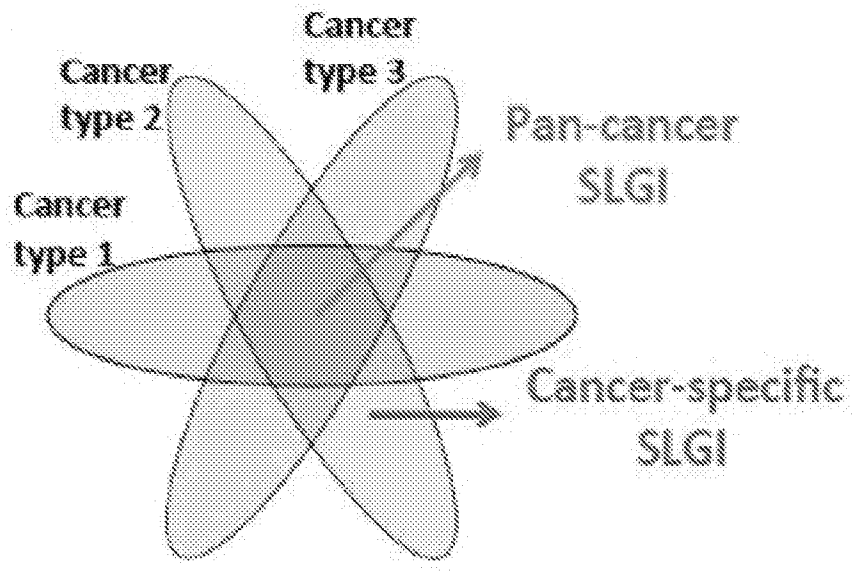

Cancer type 3
Cancer type 2
Pan-cancer SLGI
Cancer type 1
Cancer-specific SLGI

FIG. 8B

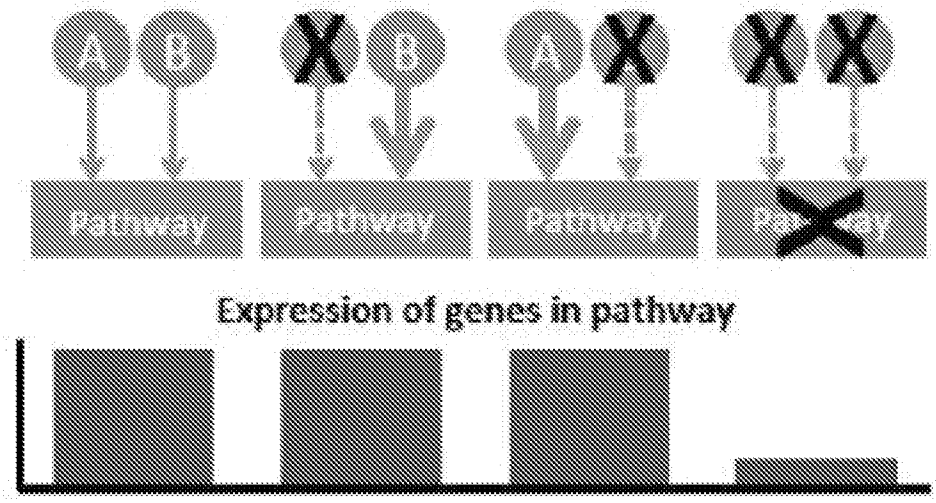

A
B
B
A
Pathway
Pathway
Pathway
Expression of genes in pathway

FIG. 8C

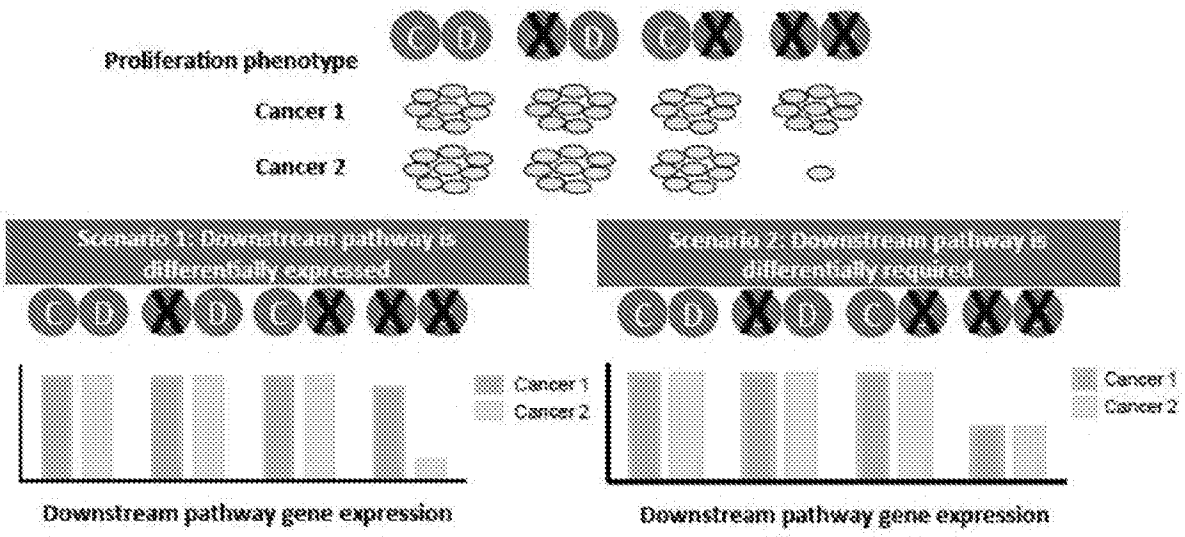
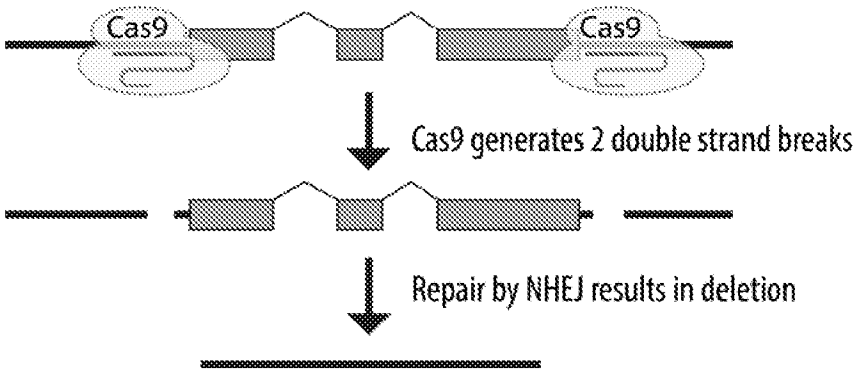

Proliferation phenotype
Cancer 1
Cancer 2
Scenario 1: Downstream pathway is differentially expressed
Scenario 2: Downstream pathway is differentially required
Cancer 1
Cancer 2
Downstream pathway gene expression
Cancer 1
Cancer 2
Downstream pathway gene expression

FIG. 9A

The paired-guide (pgRNA) CRISPR Library for Functional Enhancer Screen



A more efficient way to do enhancer screens:

1. smaller library for higher coverage
2. higher efficiency through fragment deletion

FIG. 9B

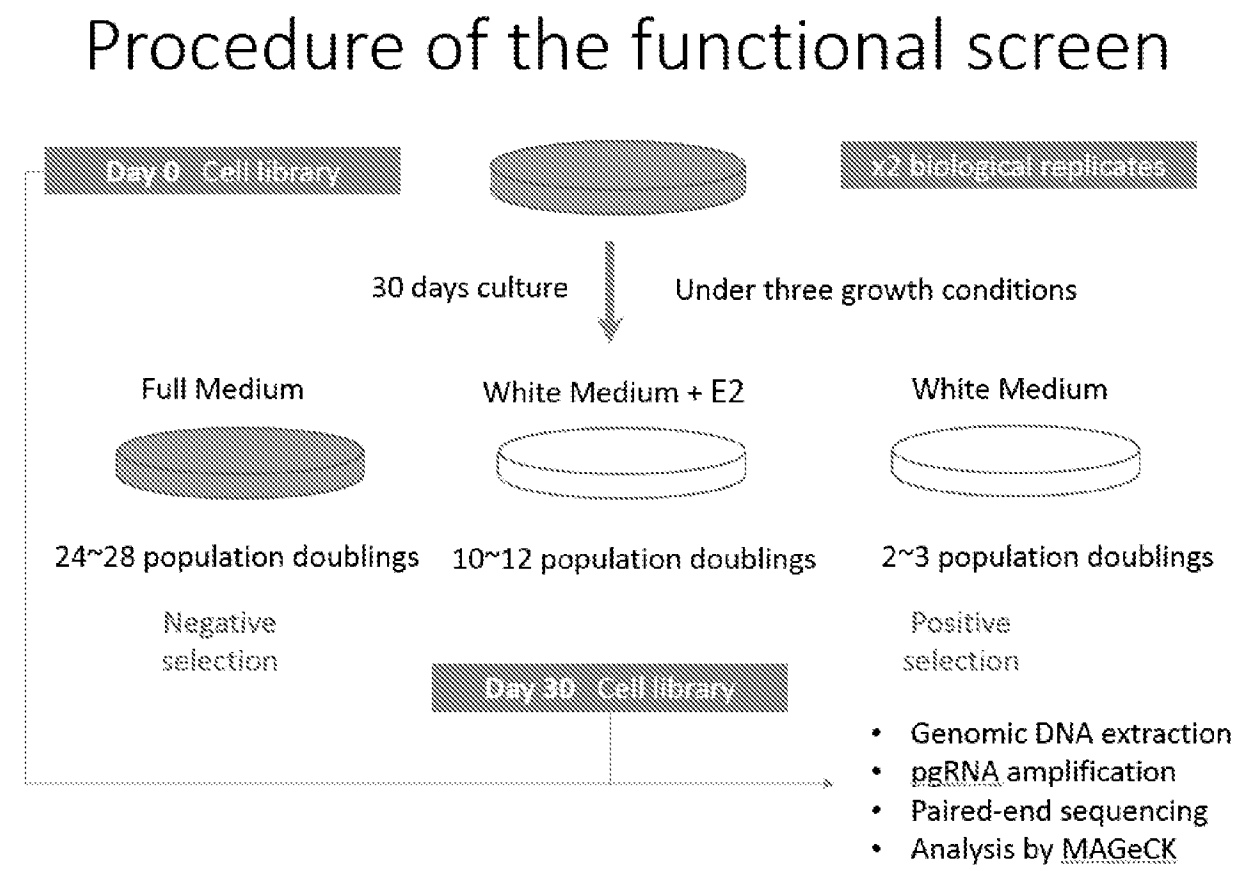

Procedure of the functional screen
Day 0 Cell library
X2 biological replicates
30 days culture
Under three growth conditions
Full Medium
White Medium + E2
White Medium
24~28 population doublings
10~12 population doublings
2~3 population doublings
Negative selection
Positive selection
Day 30 Cell library
Genomic DNA extraction
pgRNA amplification
Paired-end sequencing
Analysis by MAGeCK

FIG. 10

Deletion of the CSK enhancer by pgRNA reconstructs the CSK knockout phenotype

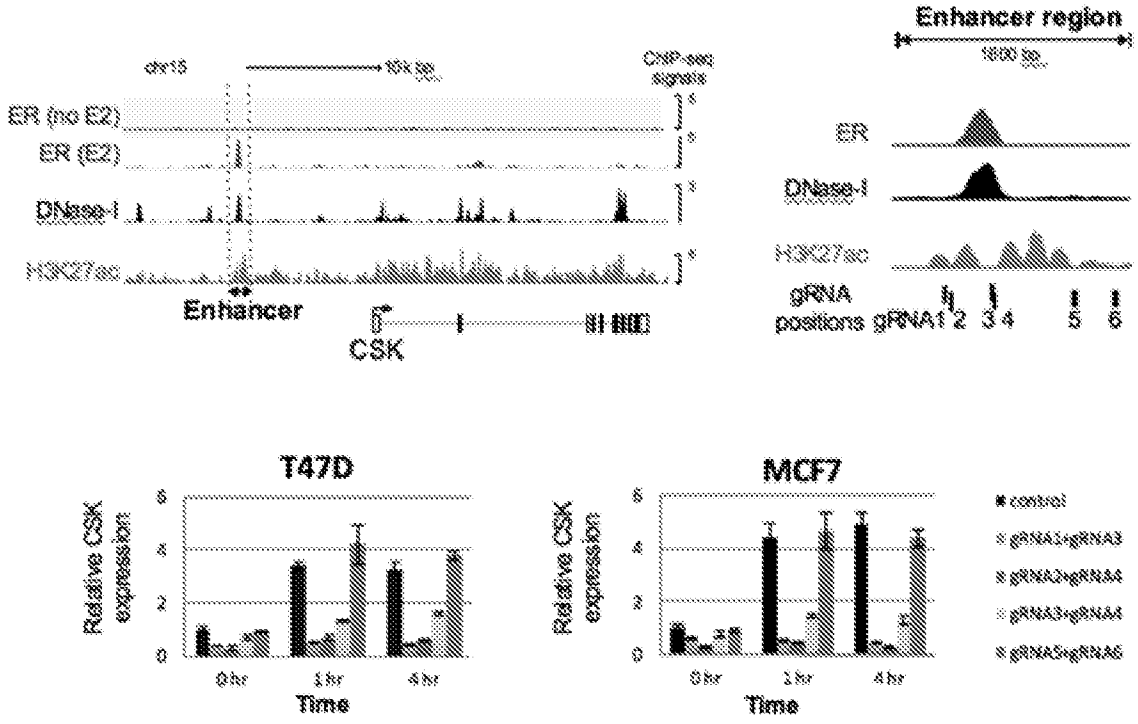


FIG. 11

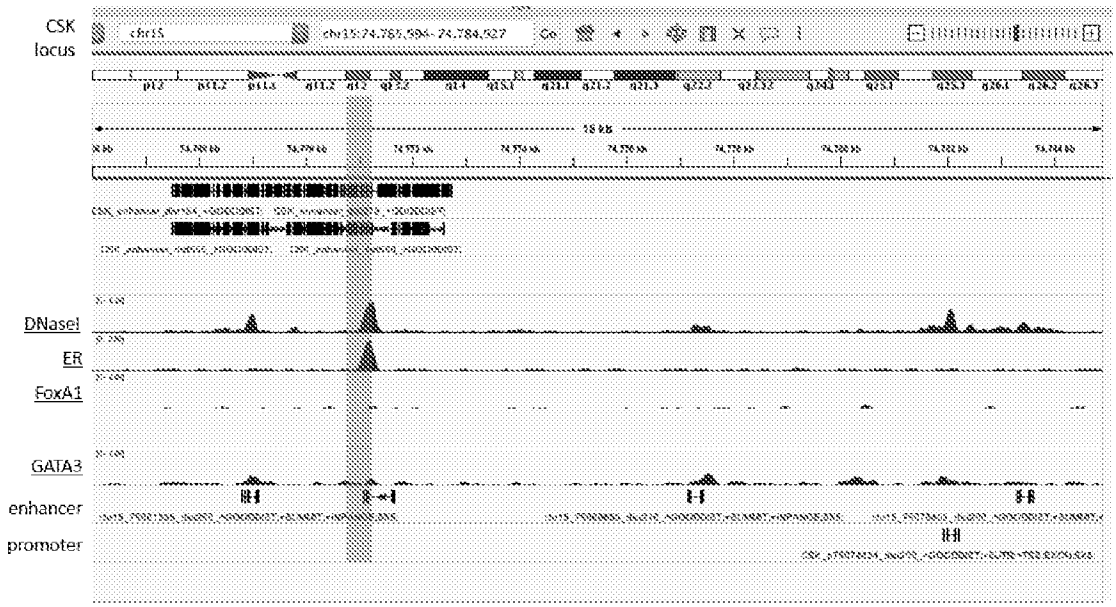
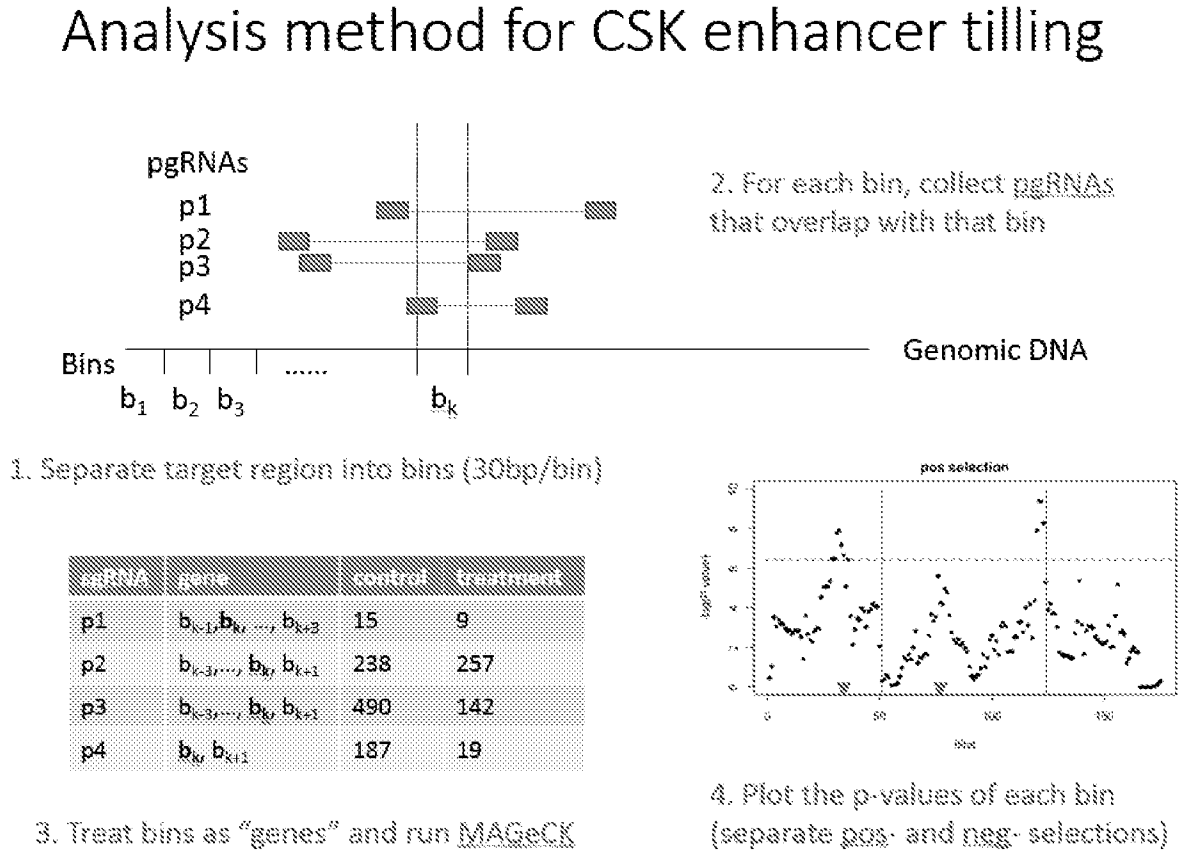

CSK Enhancer Tilling Design
CSK locus
DNaseI
ER
FoxA1
GATA3
enhancer
promoter

FIG. 12



sgRNA	gene	control	treatment
p1	b_{k-1},b_k, ..., b_{k+3}	15	9
p2	b_{k-3},..., b_k, b_{k+1}	238	257
p3	b_{k-3},..., b_k, b_{k+1}	490	142
p4	b_k, b_{k+1}	187	19

FIG. 13

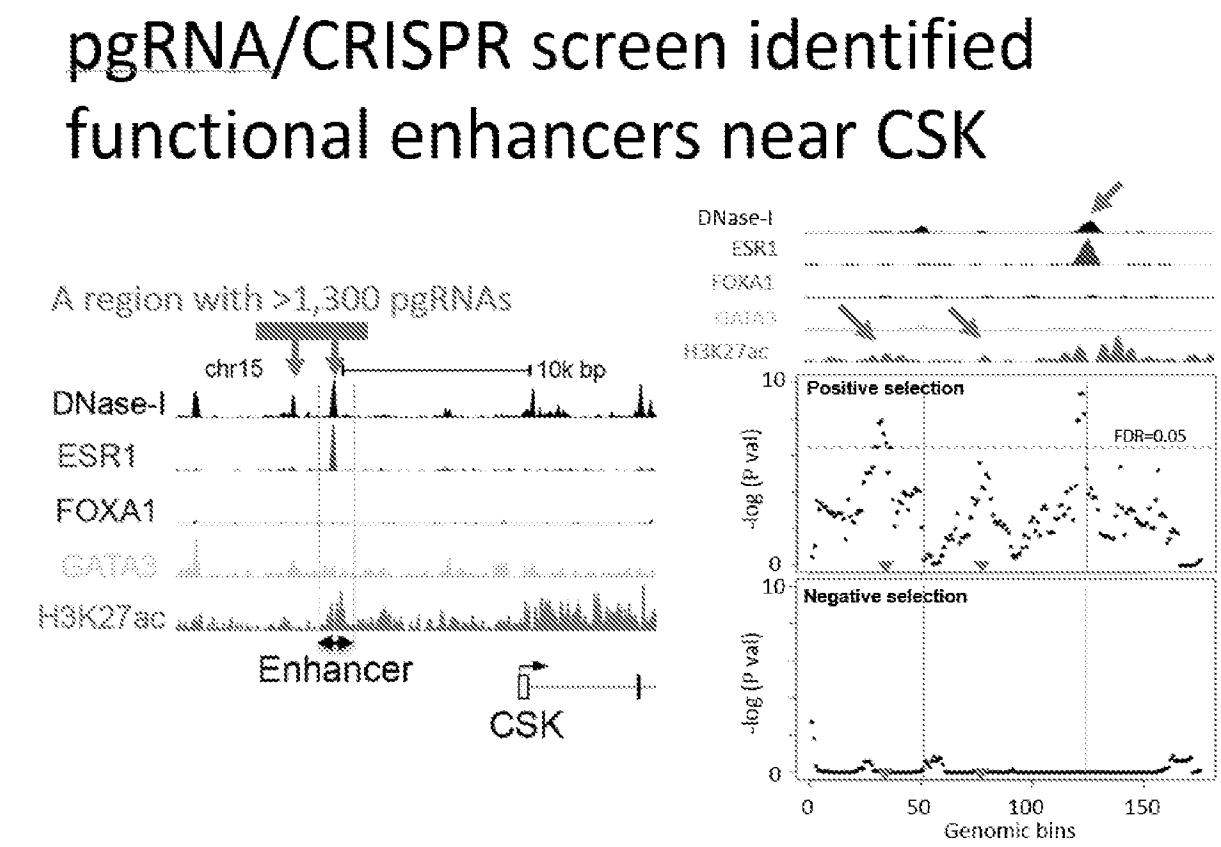

pgRNA/CRISPR screen identified functional enhancers near CSK
A region with >1,300 pgRNAs
chr15
10k bp
DNase-I
ESR1
FOXA1
GATA3
H3K27ac
Enhancer
CSK
DNase-I
ESR1
FOXA1
GATA3
H3K27ac
Positive selection
FDR=0.05
-log (P val)
Negative selection
-log (P val)
0
50
100
150
Genomic bins

FIG. 14

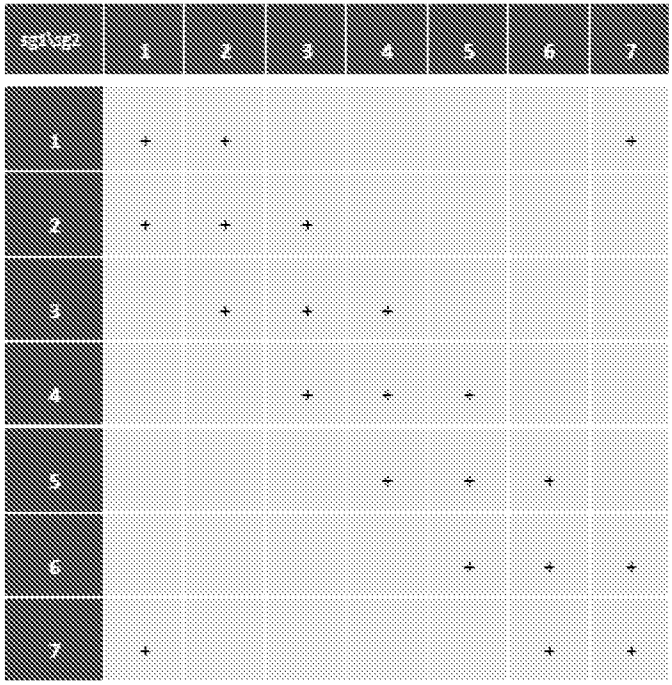


FIG. 15

Sample	# of reads with both correct guides	Mapped pairs	Percentage	Total sgRNAs	Zero counts	Gini Index
vitro	1,593,218	1,582,308	99.94%	15000	0	0.0216
(2) plasmid	1,200,978	1,134,716	92.90%	15000	0	0.0369
(1) plasmid	1,468,788	1,321,885	90.01%	15000	22	0.0470
Day 0 Cell Library	1,445,524	795,038	54.98%	15000	59	0.0959

FIG. 16

id	num	neg|score	neg|p.value	neg|fdr	neg|rank	neg|goodsgrna	neg|lfc
ARID1A_ARID1B	21	6.08E-08	4.96E-06	0.000339	5	19	-3.2304
AAVS1_ARID1A	21	0.2414	0.60183	0.999856	336	4	-0.043453
AAVS1_ARID1B	21	0.0065296	0.045652	0.34662	81	13	-2.1224
AKT1_AKT3	21	0.00010342	0.0012443	0.023374	32	16	-2.7892
AAVS1_AKT1	21	0.175	0.50607	0.999856	298	10	-0.58018
AAVS1_AKT3	21	0.10467	0.36856	0.999856	223	11	-1.1019
BRCA2_PARP1	21	0.00025144	0.0024539	0.044387	34	14	-1.7575
AAVS1_BRCA2	21	0.0010887	0.0098899	0.132224	46	13	-0.9687
AAVS1_PARP1	21	0.043893	0.20864	0.788575	162	12	-0.54089

FIG. 17A

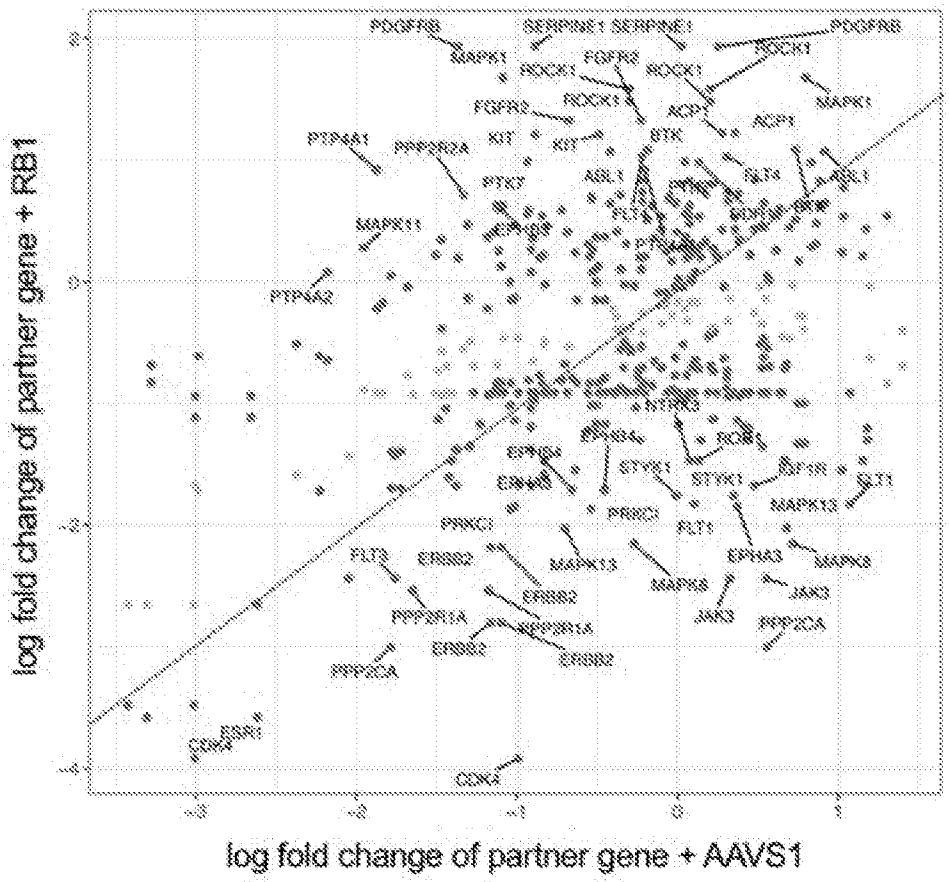

log fold change of partner gene + RB1
log fold change of partner gene + AAVS1
PDGFRB
SERPINE1
MAPK1
ROCK1
FGFR2
KIT
PTK7
ABL1
FLT1
MAPK11
PTPN2
STYK1
PRKCI
ERBB2
PPP2R1A
PPP2CA
CDK2
JAK3
MAPK13
MAPK8
EPHA3

FIG. 17B

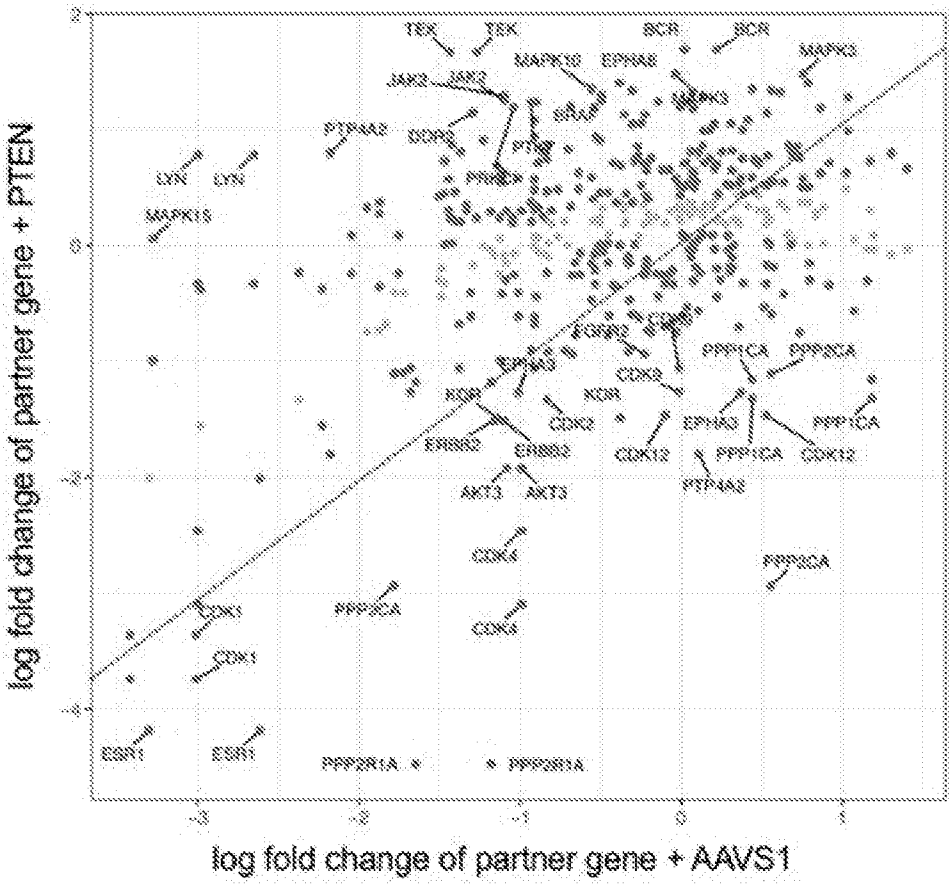

log fold change of partner gene + PTEN
log fold change of partner gene + AAVS1
TEK
BCR
MAPK10
EPHA6
JAK2
PTPN22
CDK1
MAPK13
PPP1CA
PPP3CA
CDK2
EPHA3
ERBB2
AKT3
CDK12
PTPN22
CDK4
CDK6
PPP2R1A

FIG. 17C

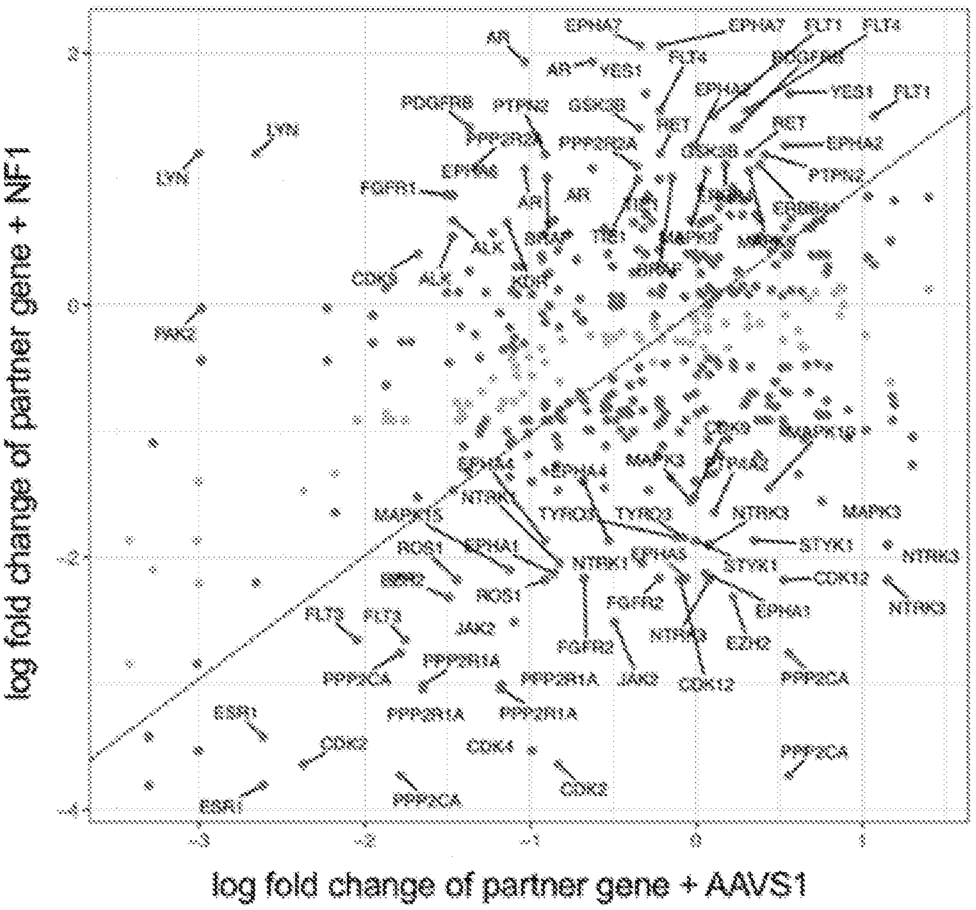
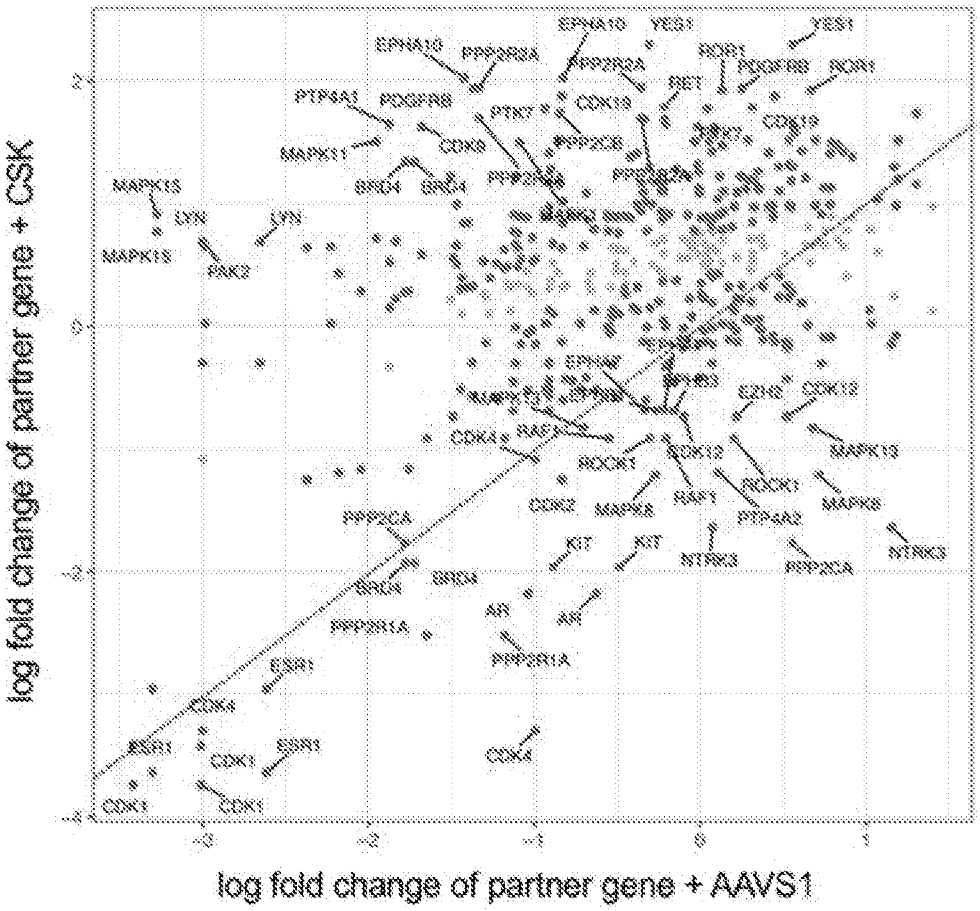

log fold change of partner gene + NF1
log fold change of partner gene + AAVS1
EPHA7
FLT1
FLT4
AR
YES1
LYN
PDGFRB
PTPN2
GSK3B
RET
EPHA2
FGFR1
ALK
CDK6
PAK2
MAPK15
ROS1
EPHA1
NTRK1
TYRO3
EPHA5
STYK1
CDK12
NTRK3
MAPK3
FGFR2
JAK2
EZH2
FLT3
PPP2R1A
PPP2CA
ESR1
CDK2
CDK4
2
0
-2
-4
-3
-1
1

FIG. 17D


log fold change of partner gene + CSK
log fold change of partner gene + AAVS1
EPHA10
YES1
PPP2R2A
ROR1
PDGFRB
RET
CDK19
PTP4A1
PTK7
MAPK11
MAPK15
LYN
PAK2
BRD4
CDK4
RAF1
ROCK1
CDK12
EZH2
CDK13
MAPK13
MAPK8
PTP4A2
NTRK3
CDK2
KIT
PPP2CA
AR
PPP2R1A
ESR1
CDK1