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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 May 2013 (16.05.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/071295 A2

(51) International Patent Classification:
C12N 15/09 (2006.01)

(21) International Application Number:
PCT/US2012/064836

(22) International Filing Date:
13 November 2012 (13.11.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/558,277 10 November 2011 (10.11.2011) US

(71) Applicant: **RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY** [US/US]; Old Queens Building, Somerset and George Streets, New Brunswick, New Jersey 08903 (US).

(72) Inventors; and

(71) Applicants : **ACTON, Thomas B.** [US/US]; Old Queens Building, Somerset and George Streets, New Brunswick, New Jersey 08903 (US). **ANDERSON, Stephen** [US/US]; Old Queens Building, Somerset and George Streets, New Brunswick, New Jersey 08903 (US). **HUANG, Yuanpeng Janet** [US/US]; Old Queens Building, Somerset and George Streets, New Brunswick, New Jersey 08903 (US). **MONTELIONE, Gaetano** [US/US]; Old Queens Building, Somerset and George Streets, New Brunswick, New Jersey 08903 (US).

(74) Agents: **KRUEGER, Katherine A.D.** et al.; 7900 International Drive, Suite 670, Bloomington, Minnesota 55425 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

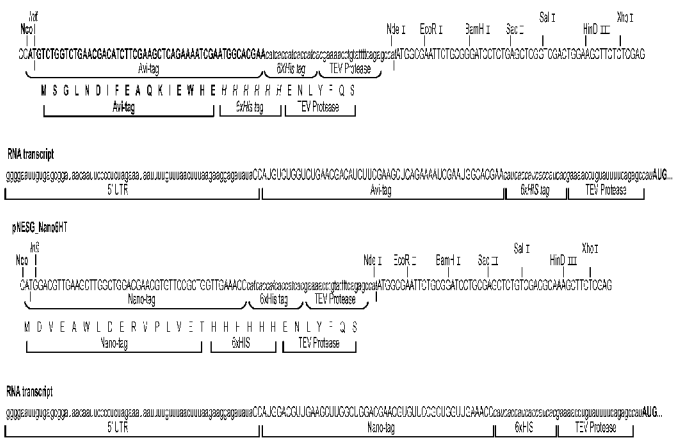
— *of inventorship (Rule 4.17(iv))*

Published:

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

(54) **Title**: TRANSCRIPT OPTIMIZED EXPRESSION ENHANCEMENT FOR HIGH-LEVEL PRODUCTION OF PROTEINS AND PROTEIN DOMAINS

FIG. 1



(57) **Abstract**: The present invention relates to a system for high-level production of recombinant proteins and protein domains.

WO 2013/071295 A2

TRANSCRIPT OPTIMIZED EXPRESSION ENHANCEMENT FOR HIGH-LEVEL PRODUCTION OF PROTEINS AND PROTEIN DOMAINS

CROSS-REFERENCE TO RELATED APPLICATION(S)

This application claims priority from United States Provisional Application Number 61/558277, filed November 10, 2011, which application is herein incorporated by reference.

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with government support under Grant # U54-GM074958 awarded by the National Institute of General Medical Sciences Protein Structure Initiative and Grant # U01-DC011485 awarded by the National Institute on Deafness and other Communication Disorders under the auspices of the NIH Common Fund. The government has certain rights in the invention.

BACKGROUND

The production of recombinant proteins and protein domains as reagents is extremely valuable to biomedical researchers and the entire biotechnology industry. *Escherichia coli* expression systems are the most cost effective and widely utilized expression systems for this task. However, production of certain proteins can be challenging in this bacterial system. Often proteins or protein domains fail to express at sufficient levels to allow for the purification of the protein reagents. This is especially true of the protein coding sequences derived from higher eukaryotes (such as humans). For example, using a standard pET *E. coli* expression system (Acton et al., 2011), nearly one-third of human protein targets produced in a large scale screen of protein expression had no detectable expression levels.

Thus, there is a need for agents and methods for high-level production of recombinant proteins and protein domains that do not require RNA optimization for each individual target gene.

SUMMARY OF CERTAIN EMBODIMENTS OF THE INVENTION

This invention relates to a system for high-level production of recombinant proteins and protein domains that does not require RNA optimization for each individual target gene.

Certain embodiments of the invention provide a method of preparing an expression vector, wherein the expression vector comprises, in order of position: a first nucleic acid sequence encoding a 5' untranslated region of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag; and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame with the second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the method comprises specifically modifying the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

Certain embodiments of the invention provide an expression vector designed using the methods described herein.

Certain embodiments of the invention provide an expression vector comprising, in order of position: a first nucleic acid sequence encoding a 5' untranslated region of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag; and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame with the second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag has been specifically modified to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

Certain embodiments of the invention provide a host cell comprising an expression vector as described herein.

The details of one or more embodiments of the invention are set forth in the description below. Other features, objects, and advantages of the invention will be apparent from the description and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a set of diagrams showing sequences of Avi-tag and Nano-tag based Transcript-Optimized Expression Enhancement Technology (TOEET) expression vectors. The pNESG_Avi6HT Avi-tag sequence (top) (DNA, RNA and protein sequence), the His-tag sequences and the TEV Protease Recognition Site sequences are shown as indicated. Similarly, for pNESG_Nano6HT (bottom) the Nano-tag sequences, the His-tag sequences and TEV Protease Recognition Site sequences are shown as indicated. The T7 RNA transcript produced by each vector is shown under each vector with untranslated sequences indicated with brackets. The Multiple Cloning Site (MCS) is also shown after the tag sequences, including the positions and identity of restriction sites available for cloning.

FIG. 2 is a diagram showing the predicted mRNA secondary structure resulting from T7-RNA Polymerase based transcription off of the pNESG_Avi6HT T7 promoter. Numbering of the transcript from nucleotides 1 – 156 is indicated; negative numbers (in italics) show the estimated strength, in kcal/mole, of the predicted base-paired regions. The arrow indicates a predicted open structure (lack of base pairing) at the RBS/translation initiation region. RNA secondary structure predictions were done using GeneBee-NET (http://www.genebee.msu.su/services/rna2_reduced.html).

FIG. 3 is a set of photographs showing representative SDS-PAGE analysis of expression and solubility for two human protein domains cloned into each of the three vectors pET15_NESG, pNESG_Nano6HT and pNESG_Avi6HT. Left Panel shows the expression and solubility of HR7724C (HUGO ID: ZNF281) residues 291-374. Right Panel shows the expression and solubility of HR8241 (HUGO ID: NR4A21) residues 261-342. Total cell lysate (Tot) and the soluble portion (Sol) of the cell lysate are run in adjacent lanes for each of the two protein domains and the three expression vectors. An asterisk (*) indicates an overexpressed band of the correct size. Note the lack of protein expression in the case of pET15_NESG constructs.

FIG. 4. Wild-Type and TOEET-Optimized *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP). The sequences at the top corresponds to the first 30 residues of the wild-type PfR-MBP DNA sequence lacking the native secretion signal. The protein open reading frame (DNA sequence) is shown

above the corresponding protein sequence. Directly below is the T7 RNA polymerase mediated RNA transcript resulting from the cloning of the PfR-MBP into the pET15_NESG backbone. The Ribosome Binding Site (RBS) is underlined and highlighted in bold, the translation initiation codon is shown in bold-italics. The lower set of sequences correspond to TOEET-optimized PfR-MBP. Bold nucleotides with arrows indicate positions where silent mutations were introduced for codon optimization, predicted decrease in RNA secondary structure in the regions of the RBS and translation initiation codon, or both. The RNA transcript for the TOEET optimized sequence is also shown following the parameters outlined above. The silent mutations were introduced using primers incorporating the nucleotide changes and 5 successive rounds of PCR, negating the need for expensive total gene synthesis.

FIG. 5. The predicted mRNA secondary structure resulting from T7-RNA Polymerase based transcription off of the pET15_NESG vector backbone with *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP) without TOEET optimization. The arrows indicate significant secondary structure (base pairing) at both the Ribosome Binding Site (RBS) and the translation initiation site (Initiation Codon). RNA secondary structure predictions were performed using GeneBee-NET (http://www.genebee.msu.su/services/rna2_reduced.html).

FIG. 6. The predicted mRNA secondary structure resulting from T7-RNA Polymerase based transcription off of the pET15_NESG vector backbone with *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP) after TOEET optimization. The arrows indicates the Ribosome Binding Site (RBS) and the translation initiation site (Initiation Codon) and the prediction of significantly greater open structure (lack of base pairing) after TOEET optimization. RNA secondary structure predictions were done using GeneBee-NET (http://www.genebee.msu.su/services/rna2_reduced.html).

FIG. 7. Histogram plots comparing Expression scores (E ranging from 0 to 5) using the TOEET technology (E_TOEET) compared to expression scores for the same target protein using a pET vector lacking TOEET technology (E_pET). The data shown in Figure 7a is for 98 protein target genes cloned into the pNESG_Avi6HT TOEET vector compared with the exact same genes cloned into the pET15_NESG vector (lacking TOEET). The data shown in Figure 7b is for 94 protein target genes cloned into the pNESG_Nano6HT TOEET vector

compared with the exact same genes cloned into pET15_NESG vector (lacking TOEET). In these histogram plots, a value E_TOEET – E_pET = 0 indicates that the expression levels for both vectors were identical; values E_TOEET – E_pET > 0 indicate that the TOEET technology provided higher level expression, values E_TOEET – E_pET < 0 indicate that the TOEET technology provided lower level expression.

DETAILED DESCRIPTION

mRNA stem-loop structures often inhibit translation initiation and therefore reduce recombinant protein expression (Nomura et al., 1984). High level expression of proteins is affected by a lack of mRNA secondary structure near the translation start site (Kudla et al., 2009; Rocha et al., 1999). In addition, rare codons present within the first ten residues of a protein have deleterious effects on protein expression levels (Gonzalez de Valdivia and Isaksson, 2004). *E. coli*, like all organisms, prefers to use a subset of the possible codons. The codons that an organism utilizes only infrequently are termed "rare codons" of that organism.

Heterologous genes from other organisms, which generally have a different codon bias, often contain *E. coli* rare codons. Decreasing or minimizing mRNA secondary structure near the Ribosome Binding Site (RBS) and translation initiation site, and separately that a lack of rare codons near the start of translation, are important for high level *E. coli* protein expression (Gonzalez de Valdivia and Isaksson, 2004; Kudla et al., 2009). However, the DNA coding sequence of a target gene destined for heterologous expression in *E. coli* has evolved under different conditions and may intrinsically contain deleterious rare codons and mRNA secondary structure when cloned into an expression vector. Deleterious rare codons and mRNA secondary structure features are particularly problematic when expressing domains or specific segments of target proteins; e.g., gene segments coding for fragments other than the native N-terminal region of the protein have not evolved to provide for efficient translation initiation. Total gene synthesis, or the chemical synthesis of a protein coding region, may address these problems to some extent, since the DNA sequence can be optimized to reduce these issues (Quan et al., 2011). However, the costs of total gene synthesis are prohibitive for large sets of protein targets, and generally is

not suitable for large-scale screening or projects involving expression of many different proteins.

This invention is based, at least in part, on an unexpected discovery of a new methodology for achieving high-level production of recombinant proteins and protein domains. RNA sequence optimization is a well-known approach for improving protein expression. A feature of the system described herein is that RNA sequence optimization is required only in DNA comprising the vector backbone, including the DNA coding for the 5'-UTR and a common N-terminal polypeptide tag. Each target gene, coding for various target proteins, that is cloned into this vector backbone, need not be optimized individually. Hence, the optimized vector backbone can be used to enhance expression of many different target proteins without the need for target-protein-specific gene sequence optimization. Unlike certain previous methods, gene-by-gene RNA transcript sequence optimization is not required in certain embodiments of the methods described herein. The methodology includes, among others, jointly designing and optimizing sequences encoding 5' untranslated and 5' translated regions of the mRNA transcript produced by an expression vector so as to minimize RNA secondary structure and/or optimize codon usage in the mRNA transcript.

In one aspect, this invention addresses, among others, the problems associated with mRNA secondary structure and codon bias. Accordingly, the invention provides systems for high-level production of recombinant proteins and protein domains based on the Transcript-Optimized Expression Enhancement Technology (TOEET). As disclosed herein, TOEET is used to design expression vectors that produce mRNA transcripts with minimal RNA secondary structure and optimum codon usage in the nucleotide region around the Ribosomal Binding Site (RBS) and the translation initiation site, as well as minimal RNA secondary structure and optimal codon usage in a region of the transcript coding for an N-terminal polypeptide tag that is encoded directly downstream of the translation initiation site. Optimization can extend up to approximately 100 or more nucleotides on each of the 5' and 3' sides of the RBS. This generally will involve producing a protein with an N-terminal polypeptide tag, which is called an Expression Enhancement Tag (EET). This EET may be designed with other features that support protein production, such as solubility enhancing properties or affinity purification sequence motifs.

Solubility enhancing tags known from the literature include the maltose-binding protein, the B1 domain of protein G, and domain of myxococcus protein S, to name a few representative examples. Expression vectors designed with TOEET allow most genes of interest to be produced with enhanced expression.

An advantage of the TOEET strategy over target gene optimization by total gene synthesis is that unless the 5' end of the synthetic gene is optimized in the context of the untranslated vector sequences, detrimental mRNA secondary structure may form near or around the RBS / translation initiation site. More specifically, even if the 5' translated region of the target gene is optimized by gene synthesis or by specific mutations, enhanced expression may not be realized unless the 5'- translated and 5'- untranslated regions of the transcript are jointly optimized, as described herein. Furthermore, by using a sufficiently long N-terminal EET tag, translated from an optimized RNA sequence that is encoded by the vector itself, there is no need to optimize the sequence of the target gene, avoiding the need for gene-specific synthesis or modification. This feature allows the TOEET technology to be used for target protein expression enhancement in high throughput applications, including expression screening studies and projects involving expression of many different proteins, where gene–specific synthesis or modification would be costly or impractical. The roughly 30 amino-acid residue (or larger) EETs effectively shift any deleterious RNA features of the target gene transcript significantly downstream of the RBS / translation initiation site, so that any potential RNA secondary structure formation with the 5' end of the transcript is avoided, and any RNA secondary structure within the RNA coded for by the target gene itself will likely have little or no effect on expression. This TOEET strategy, which is independent of the target gene sequence, could be used more generally to enhance the expression levels of proteins produced with almost any expression vector or system.

Accordingly, certain embodiments of the invention provide a method of preparing an expression vector, wherein the expression vector comprises, in order of position: a first nucleic acid sequence encoding a 5' untranslated region (UTR) of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag (i.e., at the N-terminal end of the expressed target protein); and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame

with the second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the method comprises specifically modifying the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

As used herein, a vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. The vector can be capable of autonomous replication or integrate into a host DNA. Examples of the vector include a plasmid, cosmid, or viral vector. The vector of this invention includes a nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. A "regulatory sequence" includes promoters, enhancers, repressor binding sites, and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. For example, in certain embodiments of the invention, an expression vector described herein comprises a 5' upstream sequence encoding an operable promoter and associated regulatory sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like.

As used herein, the 5'UTR of the encoded messenger RNA is transcribed from a promoter and includes a ribosome binding site several nucleotides preceding the start codon.

As used herein, a "cloning site" enables a sequence, such as, e.g., a target protein coding sequence, to be inserted into an expression vector. For example, the cloning site may be a multiple cloning site (MCS), also known as a polylinker, which is a short nucleic acid sequence that contains many restriction sites. For example, Figure 1 shows a multiple cloning site, comprising a series of restriction enzyme recognition sites. In certain embodiments, the sequence is inserted in-frame, enabling expression of the inserted sequence. In certain embodiments, after the sequence, such as, e.g., the target protein coding sequence, has been inserted into the cloning site of the vector, a portion of the cloning site remains as flanking sequence on one or both sides of the inserted

sequence. In other embodiments, the cloning site no longer remains after the insertion of the sequence into the cloning site of the vector.

As described herein, the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag may be specifically modified to minimize RNA secondary structure both within and/or between these two regions of the mRNA. In certain embodiments, one feature of the method described herein is that RNA optimization is required only in DNA comprising the vector backbone, including the DNA coding for the 5'-UTR and a common N-terminal polypeptide tag, and each gene coding for various target proteins, that is cloned into this vector backbone, need not be optimized individually. Accordingly, nucleic acids within the specific sequence encoding the 5' untranslated region and the adjacent polypeptide tag are replaced with different nucleic acids to minimize RNA secondary structure of the expressed mRNA as described herein. In particular, in certain embodiments, the RNA secondary structure is minimized in the region surrounding the RBS and/or translation initiation site of the expressed mRNA. For example, nucleic acids are replaced to reduce base pairing with the RBS and/or translation initiation site of the expressed mRNA. In certain embodiments, the nucleic acid sequence directly surrounding the RBS site and/or the translation initiation site (e.g., the consensus sequences and sequences between these two sites) is minimally modified or not modified. For example, after modification the RBS site and the translation initiation site remain functionally active. In certain embodiments, nucleotides within the nucleic acid sequence encoding the polypeptide tag are modified in a manner that results in silent mutations.

Prediction of RNA secondary structure can be readily determined by one skilled in the art using techniques and tools known in the art. For example, a skilled artisan may use RNA structure prediction software, including CentroidFold (Hamada et al., 2009), CentroidHomfold (Hamada et al., 2009), CONTRAfold (Do et al., 2006), CyloFold (Bindewald et al.), KineFold (Xayaphoummine et al., 2005; Xayaphoummine et al., 2003), Mfold (Zuker and Stiegler, 1981), GeneBee-NET (Brodskii et al., 1995), (Pknots (Rivas and Eddy, 1999), PknotsRG (Reeder et al., 2007), RNA123 (www.rna123.com), RNAfold (Gruber et al., 2008), RNAshapes (Voss et al., 2006), RNAstructure (Mathews et al., 2004), Sfold (Ding et al., 2004), UNAFold (Markham and Zuker, 2008),

Crumple (Schroeder et al., 2011), and Sliding Windows & Assembly (Schroeder et al., 2011) among others.

As described herein, a target protein may refer to any of the following non-limiting embodiments: a full-length naturally occurring protein, a polypeptide sequence corresponding to a fragment or domain of a naturally occurring protein sequence, a mutant or modified form of a full-length protein or protein fragment, or a polypeptide sequence coding for a non-natural protein, such as proteins that have been engineered or designed by artificial methods.

Certain embodiments of the invention provide a method of preparing an expression vector, wherein the expression vector comprises, in order of position, a 5' upstream sequence encoding an operable promoter and associated regulatory signals, a sequence encoding the 5' untranslated region of the messenger RNA transcribed from the promoter including a ribosome binding site several nucleotides preceding the translation start codon, a sequence beginning with the start codon encoding a polypeptide tag, and a cloning site that enables "target protein" coding sequences to be inserted into the vector in-frame with the polypeptide tag thus allowing their expression as fusions to the polypeptide tag, wherein the method comprises specifically modifying the entire sequence encoding the 5' untranslated region of the messenger RNA through and including the sequence encoding the polypeptide tag sequence in order to minimize RNA secondary structure upstream of the target insertion site.

In certain embodiments, the method further comprises specifically modifying the second nucleic acid sequence to reduce the presence of rare codons (i.e. mRNA codons for which the corresponding tRNAs are in low abundance in the host cell). For example, rare codons are replaced with high frequency codons to increase expression of any target protein expressed by the vector. Codons that are considered rare are dependent on the selected host cell that is used for expression of the vector and are known to and/or can be readily determined by one skilled in the art. For example, rare codons may be identified using computer software programs known in the art, for example, the Rare Codon Calculator (RaCC) for *E. coli* (http://nihserver.mbi.ucla.edu/RACC/), http://www.jcat.de/, or http://genomes.urv.es/OPTIMIZER/.

In certain embodiments, the modified region of the nucleic acid sequence spans from the first 5' nucleotide in the expressed mRNA to the last nucleotide of the polypeptide tag.

In certain embodiments, nucleotides within about the last 20 nucleotides of the first nucleic acid sequence are modified (*i.e.*, from the nucleotide that directly precedes the encoded start codon to 20 nucleotides upstream). In certain embodiments, nucleotides within about the last, e.g., 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975 or 1,000 nucleotides of the first nucleic acid sequence are modified.

In certain embodiments, nucleotides within about the first 20 nucleotides of the second nucleic acid sequence are modified (*i.e.*, from the first nucleotide within the encoded start codon to 20 nucleotides downstream). In certain embodiments, nucleotides within about the first, e.g., 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975 or 1,000 nucleotides of the second nucleic acid sequence are modified.

In certain embodiments, the expression vector further comprises a target protein coding sequence inserted into the vector in-frame with the nucleic acid tag sequence to encode a fusion protein comprising the polypeptide tag and the target protein.

In certain embodiments, the target protein coding sequence is not modified to minimize RNA secondary structure.

In certain embodiments, the target protein coding sequence is not modified to reduce the presence of rare codons.

In certain embodiments, the target protein coding sequence is modified to minimize RNA secondary structure.

In certain embodiments, the target protein coding sequence is modified to reduce the presence of rare codons.

As used herein, the second nucleic acid sequence encodes at least one polypeptide tag. In certain embodiments, the second nucleic acid sequence encodes more than one polypeptide tag. As used herein, when the second

nucleic acid sequence encodes more than one polypeptide tag, the respective sequences that encode each polypeptide tag are joined in-frame to result in a fusion protein that comprises each polypeptide tag. In certain embodiments, the second nucleic acid sequence encodes, e.g., two, three, four, five, etc. polypeptide tags.

As used herein, the second nucleic acid sequence may encode any polypeptide tag appropriate to the particular chosen application or selected target protein (e.g., an affinity purification tag and/or a solubility enhancement tag). Polypeptide tags are known to those skilled in the art. For example, the encoded polypeptide tag may be an Avi-tag, Calmodulin-tag, FLAG-tag, HA-tag, His-tag, Myc-tag, S-tag, SBP-tag, Softag 1, Softag 3, V5 tag, Xpress tag, Isopeptag, Spy tag, BCCP, Glutathione-S-transferase-tag, Green fluorescent protein-tag, Maltose binding protein-tag, Nus-tag, Strep-tag, Thioredoxin-tag, TC tag, Ty tag, Nano-tag, Halo-tag, protein G B1 domain tag, a myxococcus protein S tag or Protein A tag.

Accordingly, in certain embodiments, the at least one encoded polypeptide tag is selected from an Avi-tag, Calmodulin-tag, FLAG-tag, HA-tag, His-tag, Myc-tag, S-tag, SBP-tag, Softag 1, Softag 3, V5 tag, Xpress tag, Isopeptag, Spy tag, BCCP, Glutathione-S-transferase-tag, Green fluorescent protein-tag, Maltose binding protein-tag, Nus-tag, Strep-tag, Thioredoxin-tag, TC tag, Ty tag, Nano-tag, Halo-tag, protein G B1 domain tag, a myxococcus protein S tag or Protein A tag.

In certain embodiments, the second nucleic acid sequence encodes at least one affinity purification tag.

In certain embodiments, the second nucleic acid sequence encodes more than one affinity purification tag.

In certain embodiments, the second nucleic acid sequence encodes two affinity purification tags.

In certain embodiments, the encoded affinity purification tag(s) is/are selected from a Streptavidin binding moiety, a maltose binding protein moiety, and a HIS tag.

In certain embodiments, the Streptavidin binding moiety is a Nano-tag or a biotinylated Avi-tag.

In certain embodiments, the second nucleic acid sequence encodes no affinity purification tags.

In certain embodiments, the second nucleic acid sequence encodes at least one solubility enhancement tag.

In certain embodiments, the second nucleic acid sequence encodes more than one solubility enhancement tag.

In certain embodiments, the second nucleic acid sequence encodes two solubility enhancement tags.

In certain embodiments, the encoded solubility enhancement tag(s) is/are selected from a maltose binding protein tag, a protein G B1 domain tag, and a myxococcus protein S tag.

In certain embodiments, the second nucleic acid sequence encodes no solubility enhancement tags.

In certain embodiments, the second nucleic acid sequence further encodes at least one protease recognition site. In certain embodiments, the second nucleic acid sequence encodes more than one protease recognition site.

As used herein, when the second nucleic acid sequence further encodes a protease recognition site(s), the sequence that encodes this/these site(s) is/are inserted in-frame with the sequence(s) that encode the at least one polypeptide tag to result in a fusion protein that comprises the polypeptide tag(s) and the protease recognition site(s). In certain embodiments, the encoded protease recognition site(s) is/are downstream of the encoded polypeptide tag(s). In certain embodiments, the encoded protease recognition site is/are between a series of encoded polypeptide tag(s).

In certain embodiments, the protease recognition site(s) is/are a Tobacco Etch Virus (TEV), Thrombin, Factor Xa and/or a human rhinovirus (HRV) 3C (e.g., PreScission Protease, GE Healthcare Life Sciences, Pittsburgh, PA) protease recognition site.

As described herein, the PreScission Protease is a genetically engineered protein consisting of human rhinovirus 3C protease. It is often produced as a fusion protein with a hexaHis or GST affinity purification tag. It specifically cleaves between the Gln and Gly residues of the recognition sequence of LeuGluValLeuPheGln/GlyPro.

In certain embodiments, the second nucleic acid sequence is at least about 21 nucleotides in length. In certain embodiments, the second nucleic acid sequence is at least about , e.g., 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, 132, 135, 138, 141, 144, 147, 150, 201, 252, 303, 354, 405, 456, 507, 558, 609, 660, 711, 762, 813, 864, 915, 966, or 1,017 nucleotides in length.

In certain embodiments, the target protein coding sequence encodes a transcription factor, a transcription factor domain, an epigenetic regulatory factor, or an epigenetic regulatory factor domain.

In certain embodiments, the target protein coding sequence encodes a polypeptide sequence described in Table 2. As described herein, the target protein coding sequence may also encode a polypeptide sequence that has substantial identity to or is a functional equivalent of a polypeptide sequence described in Table 2.

In certain embodiments, the target protein coding sequence encodes a protein antigen for producing an affinity capture reagent.

In certain embodiments, the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

In certain embodiments, the target protein coding sequence encodes a protein antigen for producing an antibody or Fab by phage display.

In certain embodiments, the expression of the target protein is about 1.5 fold greater than the expression of a target protein generated from an expression vector that was not modified as described herein. In certain embodiments, the expression of the target protein is, e.g., about 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18, 18.5, 19, 19.5, or 20, etc., fold greater than the expression of a target protein generated from an expression vector that was not modified as described herein.

As described herein, in certain embodiments, expression of a target protein from a vector that is not TOEET modified as described herein is undetectable, whereas expression of the same target protein from a vector that has been modified as described herein is detectable.

Certain embodiments of the invention provide an expression vector prepared using a method as described herein.

Certain embodiments of the invention provide a target protein expression vector (e.g. a target protein expression vector) comprising, in order of position: a first nucleic acid sequence encoding a 5' untranslated region of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag; and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame with the second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag has been specifically modified to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

In certain embodiments, the second nucleic acid sequence has been specifically modified to reduce the presence of rare codons.

In certain embodiments, the modified region of the nucleic acid sequence spans from the first 5' nucleotide in the expressed mRNA to the last nucleotide of the polypeptide tag.

In certain embodiments, nucleotides within about the last 20 nucleotides of the first nucleic acid sequence have been modified (*i.e.*, from the nucleotide that directly precedes the encoded start codon to 20 nucleotides upstream). In certain embodiments, nucleotides within about the last, e.g., 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975 or 1,000 nucleotides of the first nucleic acid sequence have been modified.

In certain embodiments, nucleotides within about the first 20 nucleotides of the second nucleic acid sequence have been modified (*i.e.*, from the first nucleotide within the encoded start codon to 20 nucleotides downstream). In certain embodiments, nucleotides within about the first, e.g., 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975 or 1,000 nucleotides of the second nucleic acid sequence have been modified.

In certain embodiments, an expression vector as described herein, further comprises a target protein coding sequence inserted into the vector in-frame with the nucleic acid tag sequence to encode a fusion protein comprising the polypeptide tag and the target protein.

In certain embodiments, the target protein coding sequence has not been modified to minimize RNA secondary structure.

In certain embodiments, the target protein coding sequence has not been modified to eliminate rare codons.

In certain embodiments, the target protein coding sequence has been modified to minimize RNA secondary structure.

In certain embodiments, the target protein coding sequence has been modified to eliminate rare codons.

In certain embodiments, the second nucleic acid sequence encodes at least one affinity purification tag.

In certain embodiments, the second nucleic acid sequence encodes more than one polypeptide tag. As used herein, when the second nucleic acid sequence encodes more than one polypeptide tag, the respective sequences that encode each polypeptide tag are joined in-frame to result in a fusion protein that comprises each polypeptide tag. In certain embodiments, the second nucleic acid sequence encodes, e.g., two, three, four, five, etc. polypeptide tags.

As used herein, the second nucleic acid sequence may encode any polypeptide tag appropriate to the particular chosen application or selected target protein (e.g., an affinity purification tag or a solubility enhancement tag). Polypeptide tags are known to those skilled in the art. For example, the encoded polypeptide tag may be an Avi-tag, Calmodulin-tag, FLAG-tag, HA-tag, His-tag, Myc-tag, S-tag, SBP-tag, Softag 1, Softag 3, V5 tag, Xpress tag, Isopeptag, Spy tag, BCCP, Glutathione-S-transferase-tag, Green fluorescent protein-tag, Maltose binding protein-tag, Nus-tag, Strep-tag, Thioredoxin-tag, TC tag, Ty tag, Nano-tag, Halo-tag, protein G B1 domain tag, a myxococcus protein S tag or Protein A tag.

Accordingly, in certain embodiments, the at least one encoded polypeptide tag is selected from an Avi-tag, Calmodulin-tag, FLAG-tag, HA-tag, His-tag, Myc-tag, S-tag, SBP-tag, Softag 1, Softag 3, V5 tag, Xpress tag, Isopeptag, Spy tag, BCCP, Glutathione-S-transferase-tag, Green fluorescent

protein-tag, Maltose binding protein-tag, Nus-tag, Strep-tag, Thioredoxin-tag, TC tag, Ty tag, Nano-tag, Halo-tag, protein G B1 domain tag, a myxococcus protein S tag or Protein A tag.

In certain embodiments, the second nucleic acid sequence encodes more than one affinity purification tag.

In certain embodiments, the second nucleic acid sequence encodes two affinity purification tags.

In certain embodiments, the encoded affinity purification tag(s) is/are selected from a Streptavidin binding moiety, a maltose binding protein moiety, and a HIS tag.

In certain embodiments the Streptavidin binding moiety is a Nano-tag or a biotinylated Avi-tag.

In certain embodiments, the second nucleic acid sequence encodes no affinity purification tags.

In certain embodiments, the second nucleic acid sequence encodes at least one solubility enhancement tag.

In certain embodiments, the second nucleic acid sequence encodes more than one solubility enhancement tag.

In certain embodiments, the second nucleic acid sequence encodes two solubility enhancement tags.

In certain embodiments, the encoded solubility enhancement tag(s) is/are selected from a maltose binding protein tag, a protein G B1 domain tag, and a myxococcus protein S tag.

In certain embodiments, the second nucleic acid sequence encodes at least one protease recognition site.

As used herein, when the second nucleic acid sequence further encodes a protease recognition site(s), the sequence that encodes this/these site(s) is/are inserted in-frame with the sequence(s) that encode the at least one polypeptide tag to result in a fusion protein that comprises the polypeptide tag(s) and the protease recognition site(s). In certain embodiments, the encoded protease recognition site(s) is/are downstream of the encoded polypeptide tag(s). In certain embodiments, the encoded protease recognition site is/are between a series of encoded polypeptide tag(s).

In certain embodiments, the protease recognition site(s) is/are a Tobacco Etch Virus (TEV), Thrombin, Factor Xa and/or a HRV 3C protease recognition site.

In certain embodiments, the second nucleic acid sequence is at least about 21 nucleotides in length. In certain embodiments, the second nucleic acid sequence is at least about , e.g., 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, 132, 135, 138, 141, 144, 147, 150, 201, 252, 303, 354, 405, 456, 507, 558, 609, 660, 711, 762, 813, 864, 915, 966, or 1,017 nucleotides in length.

In certain embodiments, the target protein coding sequence encodes a transcription factor, a transcription factor domain, an epigenetic regulatory factor, or an epigenetic regulatory factor domain.

In certain embodiments, the target protein coding sequence encodes a polypeptide sequence described in Table 2. As described herein, the target protein coding sequence may also encode a polypeptide sequence that has substantial identity to or is a functional equivalent of a polypeptide sequence described in Table 2.

In certain embodiments, the target protein coding sequence encodes a protein antigen for producing an affinity capture reagent.

In certain embodiments, the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

In certain embodiments, the target protein coding sequence encodes a protein antigen for producing an antibody or Fab by phage display.

In certain embodiments, the target protein is expressed at about a 1.5 fold higher level than a target protein generated from an expression vector that was not modified as described herein. In certain embodiments, the target protein is expressed at about, e.g., a 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, 11, 11.5, 12, 12.5, 13, 13.5, 14, 14.5, 15, 15.5, 16, 16.5, 17, 17.5, 18, 18.5, 19, 19.5, or 20, etc., higher level than a target protein generated from an expression vector that was not modified as described herein.

As described herein, in certain embodiments, expression of a target protein from a vector not modified as described herein is undetectable, whereas

expression of the same target protein from a vector that has been modified as described herein is detectable.

Certain embodiments of the invention provide a host cell comprising the expression vector as described herein. Host cells are used for the expression of vectors and are known in the art. For example, a host cell may be a bacterial cell, such as *E. coli.*

Certain embodiments of the invention provide a method for expressing a target protein in a host cell, comprising culturing the host cell as described herein for a period of time under conditions permitting expression of the target protein.

In certain embodiments, the target protein is a protein antigen for producing an affinity capture reagent.

In certain embodiments, the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

In certain embodiments, the target protein is a protein antigen for producing an antibody or Fab by phage display.

In one aspect, the invention features a method of designing an expression vector for expressing a recombinant protein in a host cell, e.g., bacterial cell (such as *E. coli.* cell). The method includes steps of: obtaining a first sequence encoding the recombinant protein; obtaining an expression vector containing an insertion site for the first sequence, wherein once inserted at the insertion site, the first sequence is joined in frame with a 5' sequence from the expression vector to form a first fusion sequence that encodes a RNA sequence, the RNA sequence having a Ribosomal Binding Site (RBS) and a translation initiation site; modifying the RNA sequence by (i) designing the RNA sequence so as to minimize RNA secondary structure in a region around the RBS site or translation initiation site, or (ii) optimizing codon usage in the RNA sequence based on codon usage of the host cell, to obtain a second fusion sequence; and cloning the second fusion sequence into the expression vector in such a way to replace the first fusion sequence.

In one embodiment, the designing step or optimizing step is carried out using Transcript-Optimized Expression Enhancement Technology (TOEET) as shown and described herein. In another, the designing step or optimizing step is carried out by introducing a third sequence encoding a N-terminal polypeptide

expression-enhancement tag (EET) directly downstream of the initiation site.

The expression-enhancement tag can be an affinity purification tag, such as one having the sequence of an Avi tag, a Nano-tag, or a 6xHis tag.

In a second aspect, the invention provides an expression vector that is designed using the method described above. In the expression vector, the second fusion sequence can have a sequence selected from the sequences shown in Figure 1. In one example, the expression vector is selected from the group consisting of pNESG_Avi6HT and pNESG_Nano6HT. The invention also provides a host cell having the expression vector.

In a third aspect, the invention features a method for increasing the expression and solubility of a recombinant protein in a host cell. The method includes obtaining the just described host cell; culturing the host cell in a culture for period of time; and recovering the recombinant protein from the host cell or the culture. To that end, the recombinant protein can be a protein antigen for producing an affinity capture reagent (such as an antibody, an antibody fragment, or an aptamer) or a protein antigen for producing antibody or Fab by phage display.

In a fourth aspect, the invention provides an immunogenic composition having the recombinant protein produced by the method described above. The composition can be administered to a subject in need thereof for generating an immune response in the subject.

In a fifth aspect, the invention provides a method of generating an antibody (either polyclonal or monoclonal) by, among others, administrating to a subject the immunogenic composition described above.

The invention also provides an isolated polypeptide, a nucleic acid encoding it, a high throughput method for identifying a soluble protein or protein domain, and a high throughput method for isolating a soluble protein or protein domain substantially as shown and described herein.

The term "nucleic acid" refers to deoxyribonucleotides (DNA, e.g., a cDNA or genomic DNA), ribonucleotides (RNA, e.g., an mRNA), or a DNA or RNA analog and polymers thereof, in either single- or double-stranded form, but preferably is double-stranded DNA, made of monomers (nucleotides) containing a sugar, phosphate and a base that is either a purine or pyrimidine. A DNA or RNA analog can be synthesized from nucleotide analogs. Unless specifically

limited, the term encompasses nucleic acids containing known analogs of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues.

The term "nucleotide sequence" refers to a polymer of DNA or RNA which can be single-stranded or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases capable of incorporation into DNA or RNA polymers. The terms "nucleic acid," "nucleic acid molecule," or "polynucleotide" are used interchangeably.

Certain embodiments of the invention encompass isolated or substantially purified nucleic acid compositions. An "isolated nucleic acid" is a nucleic acid the structure of which is not identical to that of any naturally occurring nucleic acid or to that of any fragment of a naturally occurring genomic nucleic acid. The term therefore covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic DNA molecule but is not flanked by both of the coding sequences that flank that part of the molecule in the genome of the organism in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic DNA of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any naturally occurring vector or genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment; and (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. Specifically excluded from this definition are nucleic acids present in mixtures of different (i) DNA molecules, (ii) transfected cells, or (iii) cell clones, e.g., as these occur in a DNA library such as a cDNA or genomic DNA library. The nucleic acid described above can be used to express a fusion protein of this invention. For this purpose, one can operatively link the nucleic acid to suitable regulatory sequences to generate an expression vector.

The following terms are used to describe the sequence relationships between two or more nucleotide sequences: (a) "reference sequence," (b) "comparison window," (c) "sequence identity," (d) "percentage of sequence identity," and (e) "substantial identity."

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Thus, the determination of percent identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (Myers and Miller, CABIOS, 4, 11 (1988)); the local homology algorithm of Smith *et al.* (Smith *et al.*, Adv. Appl. Math., 2, 482 (1981)); the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, JMB, 48, 443 (1970)); the search-for-similarity-method of Pearson and Lipman (Pearson and Lipman, Proc. Natl. Acad. Sci. USA, 85, 2444 (1988)); the algorithm of Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA, 87, 2264 (1990)), modified as in Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA 90, 5873 (1993)).

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, Calif.); the ALIGN

program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wis., USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (Higgins *et al.*, CABIOS, 5, 151 (1989)); Corpet *et al.* (Corpet *et al.*, Nucl. Acids Res., 16, 10881 (1988)); Huang *et al.* (Huang *et al.*, CABIOS, 8, 155 (1992)); and Pearson *et al.* (Pearson *et al.*, Meth. Mol. Biol., 24, 307 (1994)). The ALIGN program is based on the algorithm of Myers and Miller, *supra*. The BLAST programs of Altschul *et al.* (Altschul *et al.*, JMB, 215, 403 (1990)) are based on the algorithm of Karlin and Altschul *supra*.

Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a

reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, less than about 0.01, or even less than about 0.001.

To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (*e.g.*, BLASTN for nucleotide sequences, BLASTX for proteins) can be used. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. Alignment may also be performed manually by inspection.

For purposes of the present invention, comparison of nucleotide sequences for determination of percent sequence identity to the promoter sequences disclosed herein may be made using the BlastN program (version 1.4.7 or later) with its default parameters or any equivalent program. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the program.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences makes reference to a specified percentage of residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window, as measured by sequence comparison algorithms or by visual inspection. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences

that differ by such conservative substitutions are said to have "sequence similarity" or "similarity." Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, Calif.).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even at least 95%, 96%, 97%, 98%, or 99% sequence identity, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 70%, 80%, 90%, or even at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally,

stringent conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C, depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, or 94%, or even 95%, 96%, 97%, 98% or 99%, sequence identity to the reference sequence over a specified comparison window. In certain embodiments, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, JMB, 48, 443 (1970)). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Thus, certain embodiments of the invention provide nucleic acid molecules that are substantially identical to the nucleic acid molecules described herein.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

As noted above, another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.*, total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. The thermal melting point (Tm) is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl (1984); T_m 81.5°C + 16.6 (log M) + 0.41 (%GC) - 0.61 (% form) - 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. T_m is reduced by about 1°C for each 1% of mismatching; thus, T_m, hybridization, and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with >90% identity are sought, the T_m can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the T_m for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the T_m; moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the T_m; low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the T_m. Using the equation,

hybridization and wash compositions, and desired temperature, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a temperature of less than 45°C (aqueous solution) or 32°C (formamide solution), the SSC concentration is increased so that a higher temperature can be used. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the T_m for the specific sequence at a defined ionic strength and pH.

An example of highly stringent wash conditions is 0.15 M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2 x SSC wash at 65°C for 15 minutes. Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 1 x SSC at 45°C for 15 minutes. For short nucleotide sequences (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.5 M, less than about 0.01 to 1.0 M, Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C and at least about 60°C for long probes (*e.g.*, >50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2 x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent conditions for hybridization of complementary nucleic acids that have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formamide, *e.g.*, hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1 x SSC at 60 to 65°C. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1 x to 2 x SSC (20 x SSC = 3.0 M NaCl/0.3 M

trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5 x to 1 x SSC at 55 to 60°C.

In addition to the chemical optimization of stringency conditions, analytical models and algorithms can be applied to hybridization data-sets (*e.g.* microarray data) to improve stringency.

An expression vector as described herein can be introduced into host cells to produce a fusion protein of this invention. Also within the scope of this invention is a host cell that contains the above-described nucleic acid. Examples include E. coli cells, insect cells (e.g., using baculovirus expression vectors), yeast cells, plant cells, or mammalian cells. See e.g., Goeddel, (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. To produce a fusion protein of this invention, one can culture a host cell in a medium under conditions permitting expression of the protein encoded by a nucleic acid of this invention, and isolate the protein from the cultured cell or the medium of the cell. The presence of the fusion protein in an occlusion body allows one to prepare the protein from the host cell by simply separating the occlusion body from the host cell. Alternatively, the nucleic acid of this invention can be transcribed and translated in vitro, for example, using T7 promoter regulatory sequences and T7 polymerase.

The terms "peptide," "polypeptide," and "protein" are used herein interchangeably to describe the arrangement of amino acid residues in a polymer. A peptide, polypeptide, or protein can be composed of the standard 20 naturally occurring amino acid, in addition to rare amino acids and synthetic amino acid analogs. They can be any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation). The peptide, polypeptide, or protein "of this invention" includes recombinantly or synthetically produced fusion versions having the particular domains or portions that are soluble. The term also encompasses polypeptides that have an added amino-terminal methionine (useful for expression in prokaryotic cells).

A "recombinant" peptide, polypeptide, or protein refers to a peptide, polypeptide, or protein produced by recombinant DNA techniques; i.e., produced from cells transformed by an exogenous DNA construct encoding the

desired peptide. A "synthetic" peptide, polypeptide, or protein refers to a peptide, polypeptide, or protein prepared by chemical synthesis. The term "recombinant" when used with reference, e.g., to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified.

Within the scope of this invention are fusion proteins containing one or more of the afore-mentioned sequences and a heterologous sequence. A heterologous polypeptide, nucleic acid, or gene is one that originates from a foreign species, or, if from the same species, is substantially modified from its original form. Two fused domains or sequences are heterologous to each other if they are not adjacent to each other in a naturally occurring protein or nucleic acid.

An "isolated" peptide, polypeptide, or protein refers to a peptide, polypeptide, or protein that has been separated from other proteins, lipids, and nucleic acids with which it is naturally associated. The polypeptide/protein can constitute at least 10% (i.e., any percentage between 10% and 100%, e.g., 20%, 30%, 40%, 50%, 60%, 70 %, 80%, 85%, 90%, 95%, and 99%) by dry weight of the purified preparation. Purity can be measured by any appropriate standard method, for example, by column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. An isolated polypeptide/protein described in the invention can be purified from a natural source, produced by recombinant DNA techniques, or by chemical methods.

A functional equivalent of a peptide, polypeptide, or protein of this invention refers to a polypeptide derivative of the peptide, polypeptide, or protein, e.g., a protein having one or more point mutations, insertions, deletions, truncations, a fusion protein, or a combination thereof. It retains substantially the activity of the corresponding unmodified peptide/polypeptide/protein (e.g., the activity of transcription factor). The isolated polypeptide can contain a sequence of a protein as listed in Table 1 or 2 or a functional fragment thereof. In general, the functional equivalent is at least 75% (e.g., any number between 75% and 100%, inclusive, e.g., 70 %, 80%, 85%, 90%, 95%, and 99%) identical to the corresponding unmodified peptide/polypeptide/protein.

The amino acid composition of the above-mentioned peptide/polypeptide/protein may vary without disrupting their biological activity, e.g., a transcription factor activity, i.e., ability to bind to a DNA element and/or trigger or inhibit the respective cellular response. For example, it can contain one or more conservative amino acid substitutions. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), β-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a polypeptide is preferably replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of the sequences, such as by saturation mutagenesis, and the resultant mutants can be screened for the respective biological activities.

A polypeptide described in this invention can be obtained as a recombinant polypeptide. To prepare a recombinant polypeptide, a nucleic acid encoding it can be linked to another nucleic acid encoding a fusion partner, e.g., the tags disclosed herein, glutathione-s-transferase (GST), 6x-His epitope tag (or Hexa-His), 8x-His (or Octa-His) epitope tag, or M13 Gene 3 protein. The resultant fusion nucleic acid expresses in suitable host cells a fusion protein that can be isolated by methods known in the art. The isolated fusion protein can be further treated, e.g., by enzymatic digestion (e.g., TEV protease digestion), to remove the fusion partner and obtain the recombinant polypeptide of this invention.

The peptide/polypeptide/protein of this invention covers chemically modified versions. Examples of chemically modified peptide/protein include those subjected to conformational change, addition or deletion of a sugar chain, and those to which a compound such as polyethylene glycol has been bound. Once purified and tested by standard methods or according to the methods

described in the examples below, the peptide/polypeptide/protein can be included in a composition, e.g., a pharmaceutical composition or an immunogenic composition.

The term "immunogenic" refers to a capability of producing an immune response in a host animal against an antigen or antigens. This immune response forms the basis of the protective immunity elicited by a vaccine against a specific infectious organism. "Immune response" refers to a response elicited in an animal, which may refer to cellular immunity (CMI); humoral immunity or both. "Antigenic agent," "antigen," or "immunogen" means a substance that induces a specific immune response in a host animal. The antigen can be a protein described above, a vector encoding it, a cell having the vector or protein, or any combination thereof.

The term "animal" includes all vertebrate animals including humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages. In particular, the term "vertebrate animal" includes, but not limited to, humans, canines (e.g., dogs), felines (e.g., cats); equines (e.g., horses), bovines (e.g., cattle), porcine (e.g., pigs), as well as in avians. The term "avian" refers to any species or subspecies of the taxonomic class ava, such as, but not limited to, chickens (breeders, broilers and layers), turkeys, ducks, a goose, a quail, pheasants, parrots, finches, hawks, crows and ratites including ostrich, emu and cassowary.

The immunogenic composition can be used to generate antibodies against the peptide/polypeptide/protein of this invention. As used herein, "antibody" is used in the broadest sense and specifically covers monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), and antibody fragments so long as they exhibit the desired biological activity.

As used herein, "antibody fragments", may comprise a portion of an intact antibody, generally including the antigen binding or variable region of the intact antibody, the Fab region of the antibody, or the Fc region of an antibody which retains FcR binding capability. Examples of antibody fragments include linear antibodies; single-chain antibody molecules; and multispecific antibodies formed from antibody fragments. The antibody fragments preferably retain at least part of the hinge and optionally the CH1 region of an IgG heavy chain.

More preferably, the antibody fragments retain the entire constant region of an IgG heavy chain, and include an IgG light chain.

As used herein, Affinity Capture Reagents are cognate molecules capable or recognizing and binding to a protein antigen, including protein antigens produced by TOEET-optimized expression vectors. Affinity Capture reagents include (but are not limited to) monoclonal and polyclonal antibodies, Fab or Fab fragments generated by phage and related antigen display methods, RNA aptamers, and various protein binding scaffolds which can be used to generate antigen-recognizing molecules.

As used herein, the term "Fc fragment" or "Fc region" is used to define a C-terminal region of an immunoglobulin heavy chain. The "Fc region" may be a native sequence Fc region or a variant Fc region. Although the boundaries of the Fc region of an immunoglobulin heavy chain might vary, the human IgG heavy chain Fc region is usually defined to stretch from an amino acid residue at position Cys226, or from Pro230, to the carboxyl-terminus thereof.

A "native sequence Fc region" comprises an amino acid sequence identical to the amino acid sequence of an Fc region found in nature. A "variant Fc region" as appreciated by one of ordinary skill in the art comprises an amino acid sequence which differs from that of a native sequence Fc region by virtue of at least one "amino acid modification." Preferably, the variant Fc region has at least one amino acid substitution compared to a native sequence Fc region or to the Fc region of a parent polypeptide, e.g., from about one to about ten amino acid substitutions, and preferably from about one to about five amino acid substitutions in a native sequence Fc region or in the Fc region of the parent polypeptide. The variant Fc region herein will preferably possess at least about 80% homology with a native sequence Fc region and/or with an Fc region of a parent polypeptide, and more preferably at least about 90% homology therewith, more preferably at least about 95% homology therewith, even more preferably, at least about 99% homology therewith.

Within the scope of this invention is a composition that contains a suitable carrier and one or more of the agents described above. The composition can be a pharmaceutical composition that contains a pharmaceutically acceptable carrier. The term “pharmaceutical composition” refers to the combination of an active agent with a carrier, inert or active, making the composition especially

suitable for diagnostic or therapeutic use *in vivo* or *ex vivo*. A "pharmaceutically acceptable carrier," after administered to or upon a subject, does not cause undesirable physiological effects. The carrier in the pharmaceutical composition must be "acceptable" also in the sense that it is compatible with the active ingredient and can be capable of stabilizing it. One or more solubilizing agents can be utilized as pharmaceutical carriers for delivery of an active compound. Examples of a pharmaceutically acceptable carrier include, but are not limited to, biocompatible vehicles, adjuvants, additives, and diluents to achieve a composition usable as a dosage form. Examples of other carriers include colloidal silicon oxide, magnesium stearate, cellulose, and sodium lauryl sulfate.

As used herein, a "subject" refers to a human and a non-human animal. Examples of a non-human animal include all vertebrates, e.g., mammals, such as non-human mammals, non-human primates (particularly higher primates), dog, rodent (e.g., mouse or rat), guinea pig, cat, and rabbit, and non-mammals, such as birds, amphibians, reptiles, etc. In one embodiment, the subject is a human. In another embodiment, the subject is an experimental, non-human animal or animal suitable as a disease model.

The composition of this invention can include an adjuvant agent or adjuvant. As used herein, the term "adjuvant agent" or "adjuvant" means a substance added to an immunogenic composition or a vaccine to increase the immunogenic composition or the vaccine's immunogenicity. Examples of an adjuvant include a cholera toxin, *Escherichia coli* heat-labile enterotoxin, liposome, unmethylated DNA (CpG) or any other innate immune-stimulating complex. Various adjuvants that can be used to further increase the immunological response depend on the host species and include Freund's adjuvant (complete and incomplete), mineral gels such as aluminum hydroxide, surface-active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Useful human adjuvants include BCG (bacille Calmette-Guerin) and Corynebacterium parvum.

Pharmaceutical compositions comprising an adjuvant and an antigen may be manufactured by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Pharmaceutical compositions may be formulated in

conventional manner using one or more physiologically acceptable carriers, diluents, excipients or auxiliaries which facilitate processing of the antigens of the invention into preparations which can be used pharmaceutically. Proper formulation is dependent upon the route of administration chosen.

A pharmaceutical composition of this invention can be administered parenterally, orally, nasally, rectally, topically, or buccally. The term "parenteral" as used herein refers to subcutaneous, intracutaneous, intravenous, intramuscular, intraarticular, intraarterial, intrasynovial, intrasternal, intrathecal, intralesional, or intracranial injection, as well as any suitable infusion technique. For injection, immunogenic or vaccine preparations may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, phosphate buffered saline, or any other physiological saline buffer. The solution may contain formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the peptides, polypeptides, or proteins may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

Determination of an effective amount of the immunogenic or vaccine formulation for administration is well within the capabilities of those skilled in the art, especially in light of the detailed disclosure provided herein. An effective dose can be estimated initially from *in vitro* assays. For example, a dose can be formulated in animal models to achieve an induction of an immune response using techniques that are well known in the art. One having ordinary skill in the art could readily optimize administration to all animal species based on results described herein. Dosage amount and interval may be adjusted individually. For example, when used as a vaccine, the vaccine formulations of the invention may be administered in about 1 to 3 doses for a 1-36 week period. Preferably, 1 or 2 doses are administered, at intervals of about 3 weeks to about 4 months, and booster vaccinations may be given periodically thereafter. Alternative protocols may be appropriate for individual animals. A suitable dose is an amount of the vaccine formulation that, when administered as described above, is capable of raising an immune response in an immunized animal sufficient to protect the animal from an infection for at least 4 to 12 months. In general, the amount of the antigen present in a dose ranges from about 1 pg to about 100 mg per kg of host, typically from about 10 pg to about 1 mg, and

preferably from about 100 pg to about 1 pg. Suitable dose range will vary with the route of injection and the size of the patient, but will typically range from about 0.1 ml to about 5 ml.

This invention also provides methods for making antibodies against the above-described proteins. The antibodies can be either polyclonal or monoclonal.

Polyclonal antibodies against a protein of the invention can be obtained as follows. After verifying that a desired serum antibody level has been reached, blood is withdrawn from the mammal sensitized with the antigen. Serum is isolated from this blood using well-known methods. The serum containing the polyclonal antibody may be used as the polyclonal antibody, or according to needs, the polyclonal antibody-containing fraction may be further isolated from the serum. For instance, a fraction of antibodies that specifically recognize the protein of the invention may be prepared by using an affinity column to which the protein is coupled. Then, the fraction may be further purified by using a Protein A or Protein G column in order to prepare immunoglobulin G or immunoglobulin M.

To obtain monoclonal antibodies, after verifying that the desired serum antibody level has been reached in the mammal sensitized with the above-described antigen, immunocytes are taken from the mammal and used for cell fusion. For this purpose, splenocytes can be preferable immunocytes. As parent cells fused with the above immunocytes, mammalian myeloma cells are preferably used. More preferably, myeloma cells that have acquired the feature, which can be used to distinguish fusion cells by agents, are used as the parent cell.

The cell fusion between the above immunocytes and myeloma cells can be conducted according to known methods, for example, the method of Milstein et al. (Methods Enzymol., 73:3-46, 1981). The hybridoma obtained from cell fusion is selected by culturing the cells in a standard selective culture medium, for example, HAT culture medium (hypoxanthine, aminopterin, thymidine-containing culture medium). The culture in this HAT medium is continued for a period sufficient enough for cells (non-fusion cells) other than the objective hybridoma to perish, usually from a few days to a few weeks. Next, the usual

limiting dilution method is carried out, and the hybridoma producing the objective antibody is screened and cloned.

Other than the above method for obtaining hybridomas, by immunizing an animal other than humans with the antigen, a hybridoma producing the objective human antibodies having the activity to bind to proteins can be obtained by the method of sensitizing human lymphocytes, for example, human lymphocytes infected with the EB virus, with proteins, protein-expressing cells, or lysates thereof *in vitro*, fusing the sensitized lymphocytes with myeloma cells derived from human having a permanent cell division ability.

The obtained monoclonal antibodies can be purified by, for example, ammonium sulfate precipitation, protein A or protein G column, DEAE ion exchange chromatography, an affinity column to which the protein of the present invention is coupled, and so on. The antibody may be useful for the purification or detection of a protein of the invention. It may also be a candidate for an agonist or antagonist of the protein. Furthermore, it is possible to use it for the antibody treatment of diseases in which the protein is implicated. For *in vivo* administration (in such antibody treatment), human antibodies or humanized antibodies may be favorably used because of their reduced antigenicity.

For example, a human antibody against a protein can be obtained using hybridomas made by fusing myeloma cells with antibody-producing cells obtained by immunizing a transgenic animal comprising a repertoire of human antibody genes with an antigen such as a protein, protein-expressing cells, or a cell lysate thereof. Other than producing antibodies by using hybridoma, antibody–producing immunocytes, such as sensitized lymphocytes that are immortalized by oncogenes, may also be used.

Such monoclonal antibodies can also be obtained as recombinant antibodies produced by using the genetic engineering technique. Recombinant antibodies are produced by cloning the encoding DNA from immunocytes, such as hybridoma or antibody-producing sensitized lymphocytes, incorporating this into a suitable vector, and introducing this vector into a host to produce the antibody. The present invention encompasses such recombinant antibodies as well.

Moreover, the antibody of the present invention may be an antibody fragment or a modified-antibody, so long as it binds to a protein of the invention.

For example, Fab, F (ab')$_2$, Fv, or single chain Fv in which the H chain Fv and the L chain Fv are suitably linked by a linker (scFv, Huston et al., Proc. Natl. Acad. Sci. USA, 85:5879-5883, 1988) can be given as antibody fragments. Specifically, antibody fragments are produced by treating antibodies with enzymes, for example, papain, pepsin, and such, or by constructing a gene encoding an antibody fragment, introducing this into an expression vector, and expressing this vector in suitable host cells (for example, Co et al., J. Immunol., 152:2968-2976, 1994; Better et al., Methods Enzymol., 178:476-496, 1989; Pluckthun et al., Methods Enzymol., 178:497-515, 1989; Lamoyi, Methods Enzymol., 121:652-663, 1986; Rousseaux et al., Methods Enzymol., 121:663-669, 1986; Bird et al., Trends Biotechnol., 9:132-137, 1991).

As modified antibodies, antibodies bound to various molecules such as polyethylene glycol (PEG) can be used. The antibody of the present invention encompasses such modified antibodies as well. To obtain such a modified antibody, chemical modifications are done to the obtained antibody. These methods are already established in the field.

The antibody of the invention may be obtained as a chimeric antibody, comprising non-human antibody-derived variable region and human antibody-derived constant region, or as a humanized antibody comprising non-human antibody-derived complementarity determining region (CDR), human antibody-derived framework region (FR), and human antibody-derived constant region by using conventional methods.

Antibodies thus obtained can be purified to uniformity. The separation and purification methods used in the present invention for separating and purifying the antibody may be any method usually used for proteins. For instance, column chromatography, such as affinity chromatography, filter, ultrafiltration, salt precipitation, dialysis, SDS-polyacrylamide gel electrophoresis, isoelectric point electrophoresis, and so on, may be appropriately selected and combined to isolate and purify the antibodies (Antibodies: a laboratory manual. Ed Harlow and David Lane, Cold Spring Harbor Laboratory, 1988), but is not limited thereto. Antibody concentration of the above mentioned antibody can be assayed by measuring the absorbance, or by the enzyme-linked immunosorbent assay (ELISA), etc. Protein A or Protein

G column can be used for the affinity chromatography. Protein A column may be, for example, Hyper D, POROS, Sepharose F.F., and so on.

Other chromatography may also be used, such as ion exchange chromatography, hydrophobic chromatography, gel filtration, reverse phase chromatography, and adsorption chromatography (Strategies for Protein Purification and Characterization: A laboratory Course Manual. Ed. by Marshak D.R. et al., Cold Spring Harbor Laboratory Press, 1996). These may be performed on liquid chromatography such as HPLC or FPLC.

Examples of methods that assay the antigen-binding activity of the antibodies of the invention include, for example, measurement of absorbance, enzyme-linked immunosorbent assay (ELISA), enzyme immunoassay (EIA), radio immunoassay (RIA), or fluorescent antibody method. For example, when using ELISA, a protein of the invention is added to a plate coated with the antibodies of the invention, and next, the objective antibody sample, for example, culture supernatants of antibody-producing cells, or purified antibodies are added. Then, secondary antibody recognizing the antibody, which is labeled by alkaline phosphatase and such enzymes, is added, the plate is incubated and washed, and the absorbance is measured to evaluate the antigen-binding activity after adding an enzyme substrate such as *p*-nitrophenyl phosphate. As the protein, a protein fragment, for example, a fragment comprising a C-terminus, or a fragment comprising an N-terminus may be used. To evaluate the activity of the antibody of the invention, BIAcore may be used.

The following non-limiting examples set forth herein below illustrate certain aspects of the invention.

EXAMPLE 1

This example describes two specific EET tags designed utilizing TOEET. These EETs were engineered and subcloned into the pET15_NESG expression vector (Acton et al., 2011). They contain dual tandem protein purification tags and a protease cleavage site to facilitate purification of the resulting proteins. These include the 6X-His tag (Crowe et al., 1994), and one of two Streptavidin binding moieties, either the Avi-tag (Scholle et al., 2004) or the Nano-tag (Lamla and Erdmann, 2004). The Nano-tag binds directly to streptavidin (Lamla and Erdmann, 2004); the Avi-tag is a substrate for the

enzyme BirA which can be used to catalyze the covalent attachment of biotin to the Avi Tag (Scholle et al., 2004). These tandem tags allow for two separate affinity purification steps, (i) Ni-based immobilized metal affinity chromatography (IMAC) and (ii) high-affinity Streptavidin-based chromatography. This dual purification strategy allows preparation of highly purified proteins using high-throughput affinity purification methods. The Tobacco Etch Virus (TEV) protease recognition site (Kapust et al., 2002) engineered into these EETs allows removal of the affinity tags, if required, after expression and purification of the protein target.

Briefly, in designing the DNA sequences coding for these EETs, the coding sequence of one of the two Streptavidin binding moieties i.e., Avi-tag (SEQ ID NO:1 - MSGLNDIFEAQKIEWHE) or Nano-tag (SEQ ID NO:2 - MDVEAWLDERVPLVET) (Lamla and Erdmann, 2004; Scholle et al., 2004), a 6X-His tag (Crowe et al., 1994), and a TEV protease recognition site (Kapust et al., 2002) were fused in frame and optimized to have a high Codon Adaptation Index (Sharp and Li, 1987) (Fig. 1). The DNA sequence coding for the EET was optimized with TOEET, together with the 5'- untranslated region of the pET15-NESG expression vector, to generate the expression vectors pNESG_Avi6HT and pNESG_Nano6HT, shown in Fig. 1. These features functioned together to enhance translation initiation and protein expression levels.

Using these expression vectors (Fig. 1), protein expression resulted in T7 RNA Polymerase mediated transcription producing an mRNA transcript consisting of (i) vector sequence (pET15_NESG-5'- untranslated region), (ii) nucleotides coding for the EET, and (iii) nucleotides coding for the target protein sequence. Both the untranslated region of the vector upstream of the EET-coding region, and the RNA coding for the EET itself were optimized to avoid secondary structure formation within and between these regions of the mRNA transcript. In this particular implementation, the length of the optimized nucleotide sequence coding for the EET was about 90 nucleotides. Together with the 70 upstream 5'- untranslated nucleotides of the transcript driven by the T7 promoter of the vector, the 5'- region of the transcript was optimized as a unit of about 160 nucleotides. Longer optimized nucleotide sequences, and potentially somewhat shorter optimized nucleotide sequences may also be effective in creating TOEET-based expression-enhanced vectors.

The optimized regions of the pNESG_Avi6HT and pNESG_Nano6HT based TOEET vectors are shown in Fig. 1. The figure shows the DNA sequences, RNA sequences, and the translated protein tag (SEQ ID NO:3-MSGLNDIFEAQKIEWHEHHHHHHENLYFQSH and SEQ ID NO:4 - MDVEAWLDERVPLVETHHHHHHENLYFQSH, respectively) sequences of the expression vectors, along with the DNA sequence coding for the multiple cloning site (MCS), a series of restriction endonuclease sites used for cloning into the expression plasmids. Figure 2 shows, as an example, the predicted RNA secondary structure in transcripts generated from the pNESG_Avi6HT vector, highlighting the lack of predicted RNA secondary structure near the RBS / translation initiation site.

A third vector comprising the *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP) was also constructed and optimized using TOEET. The MBP from *Pyrococcus furiosus* is much more thermally stable than that of *E coli*, and is expected to provide a more robust solubilization enhancement tag and affinity purification tag. Proteins that are expressed but not soluble in cell extracts can be solubilized and used successfully as antigens using various methods of solublization, including urea and guanidine denaturtants (Agaton et al, 2003). The PfR MBP provides improved purification of target proteins under such partially denaturing conditions or other harsh conditions. The sequences shown at the top of Figure 4 correspond to the first 30 residues of the wild-type PfR-MBP DNA sequence lacking the native secretion signal. The protein open reading frame (DNA sequence) is shown above the corresponding protein sequence and directly below is the T7 RNA polymerase mediated RNA transcript resulting from the cloning of the PfR-MBP into the pET15_NESG backbone. The lower set of sequences shown in Figure 4 correspond to TOEET optimized PfR-MBP. Silent mutations were introduced for codon optimization or to decrease the predicted RNA secondary structure in the regions of the RBS and translation initiation codon, or both. The silent mutations were introduced using primers incorporating the nucleotide changes and 5 successive rounds of PCR, negating the need for expensive total gene synthesis.

The predicted mRNA secondary structure resulting from T7-RNA Polymerase based transcription off of the pET15_NESG vector backbone with *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP) without TOEET

optimization is shown in Figure 5. Significant secondary structure (base pairing) at both the Ribosome Binding Site (RBS) and the translation initiation site (Initiation Codon) is predicted. The predicted mRNA secondary structure resulting from T7-RNA Polymerase based transcription off of the pET15_NESG vector backbone with *Pyrococcus furiosus* (PfR) Maltose Binding Protein (MBP) after TOEET optimization is shown in Figure 6. As illustrated by Figure 6, significantly greater open structure (lack of base pairing) after TOEET optimization is predicted.

EXAMPLE 2

The results obtained from expression studies with the above-described new vectors demonstrated that the TOEET strategy is both extremely successful and robust. In this example, similar expression and solubility studies were carried out using a high throughput methodology for the identification and isolation of soluble proteins and protein domains.

As mentioned above, the isolation of soluble, well-folded proteins and protein domains is of great use and importance to the biotechnology industry and biological researchers as a whole. However, the production of such protein reagents remains extremely challenging, especially in the cost effective, commonly used bacterial expression systems. These *Escherichia coli* expression systems are often successful in the production of simple bacterial proteins but are far less amenable to the production of eukaryotic, mulitdomain proteins or protein complexes, often resulting in no or low levels of expression and/or solubility (greatly complicating or thwarting their production as a protein reagent). There are a variety of reasons that contribute to the lower success rate of these proteins in bacterial expression systems including the fact that eukaryotic proteins are frequently multidomain in nature, this often results in misfolding when expressed using simple prokaryotic expression systems (Netzer and Hartl, 1997). Another major reason for the higher attrition rate relates to the increased levels of disordered regions in human and other eukaryotic proteins in comparison to simpler organisms (Lui et al., 2002). These disordered regions likely cause aggregation and misfolding in E. coli expression systems leading to proteins or domains with low expression and/or solubility, again, greatly interfering with their production.

To circumvent these issues, the NESG Construct Optimization Software and High ThroughPut (HTP) Molecular Cloning and Expression Screening Platform and Automated Purification Pipeline methods were developed for assaying multiple alternative constructs to identify soluble proteins or domains (Methods in Enzymology, Vol. 493, Burlington: Academic Press, 20~11, pp. 21-60.). Briefly, the NESG Construct Optimization Software used reports from the from the DisMeta Server (http://www-nmr.cabm.rutgers.edu/bioinformatics/disorder), a metaserver that generated a consensus analysis of eight sequence-based disorder predictors to identify protein regions that are likely to be disordered. In addition, secondary structure, transmembrane and signal peptides among others were also predicted. This data along with multiple sequence alignments of homologous proteins were used to predict possible structural domain boundaries. Based on this information, the NESG Construct Optimization software generated nested sets of alternative constructs, for full-length proteins, multidomain constructs, and single domain constructs. Primers for cloning were then designed using the software Primer Primer (Everett, J.K.; Acton, T.B.; Montelione, G.T.J. Struct. Funct. Genomics 2004, 5: 13-21. Primer Prim'r: A web based server for automated primer design.). Thus for a single targeted region, multiple open reading frames were generally designed varying the N and/or C-terminal sequences. These alternative constructs often possessed significantly better expression, solubility and biophysical behavior than their full-length parent sequences, increasing the possibility of successfully producing a protein reagent.

Although the NESG Construct Optimization Software identified protein subsequences that were more likely to produce soluble well-behaved samples, several variants of each were assayed to identify constructs amenable to protein sample production. Therefore the high-throughput NESG Molecular Cloning and Expression Screening Platform was developed utilizing 96-well parallel cloning / E. coli expression and Qiagen BioRobotS000-based liquid handling. Briefly, protein target sequences (constructs) were PCR amplified from Reverse Transcriptase (RT) generated cDNA pools or genomic DNA, gel purified and extracted in 96-well format (robotic liquid handling) and subcloned into pET_NESG, a series of T7 based (Novagen) bacterial expression vectors generated at Rutgers, using InFusion (Clonetech) Ligation Independent Cloning

(LIC). The RT generated cDNA pools were derived from normal and disease tissue (tumor cells and cell lines) allowing for the isolation of wild-type and polymorphic proteins. Correct clones (containing the desired protein open reading frame) were identified using plate based-PCR assays. An automated DNA Miniprep Protocol isolated the nascent expression vectors and a 96-well transformation protocol was used to introduce the plasmids into the BI21(DE3) pMgK E. *coli* expression strain. Following overnight growth, a single representative colony from each well (96) was transferred to LB in a 96-well S-Block and incubated for 6 hours. Automated liquid handling was then utilized to produce a 500 microliter overnight subculture of each of the 96 constructs in a single 96-well S-block. An aliquot of each well was then subcultured into the corresponding well of one of four 24-well blocks containing 2 ml of fresh media and incubated at 37 °C until mid-log phase growth. Protein expression is induced with IPTG (Isopropyl113-D-1-thiogalactopyranoside) and incubated overnight at 17 °C. The cells were harvested using automated liquid handling and sonicated in 96-well format. The expression and solubility of each construct was visualized by SDS-PAGE analysis and constructs suitable for protein production were identified.

The soluble expression constructs were then fermented in large volume using parallel fermentation system, consisting of 2.5-L baffled Ultra YieldTM Fernbach flasks, low-cost platform shakers, controlled temperature rooms and specialized MJ9 media (Jansson et al. 1996). This generally produced 10 - 100 mg of protein per liter of culture. The resulting proteins were then purified using high-throughput AKTAxpress-based parallel protein purification system. This consisted of a two-step automated Ni-affinity purification (pET_NESG imparts a 6X-His tag) followed by gel filtration chromatography. The purified proteins were then analyzed for quality including molecular weight validation by MALDI-TOF mass spectrometry, homogeneity analysis by SDS-PAGE, aggregation screening by analytical gel filtration with static light scattering, and finally concentration determination was performed.

Together the NESG Construct Optimization Software, Molecular Cloning and Expression Screening Platform and Automated Purification Pipeline allow for identification and isolation of large numbers of soluble well-behaved protein reagents in a time efficient and cost effective manner. Without this

technology, many of the proteins would prove elusive in regard to production as a protein reagent.

In this process, target protein expression constructs were designed using proprietary bioinformatics methods, cloning was done using robotic methods and protocols, and Expression (E, ranging from 0 to 5) and Solubility (S, ranging from 0 to 5) screening were performed in a high throughput fashion and assessed using SDS-PAGE analysis. The read out (ES score = E score x S score, ranging from 0 to 25) provided a measure of the usability of a particular target construct and expression vector system combination for large-scale protein sample production. In general, constructs providing ES scores ≥ 9 in this high throughout expression and solubility assay provided milligram-per-liter (or tens-of-milligram per liter) quantities of protein samples in medium scale (0.5 – 3 L) shake flask fermentations.

As a demonstration of the TOEET technology, a set of approximately 96 human transcription factor genes and epigenetic regulatory factor genes were cloned into the pET15_NESG vector (Acton et al., 2011) lacking a TOEET sequence, and into both the pNESG_Avi6HT and pNESG_Nano6HT vectors. These expression vectors were constructed, and the expression and solubility of target proteins assessed, using the technology outlined above. The results of this study are summarized in Table 1.

It was found that, using the pET15_NESG vector, only 20 of 99 constructs provided expression and solubility levels that can support scale-up protein sample production (ES score ≥ 9; highlighted in grey shade in Table 1). In contrast, using the pNESG_Nano6HT or pNESG_Avi6HT on this same set of target genes provided a significant increase in the number of highly-expressed and soluble targets suitable for scale-up production. As shown in Table 1, 42 of 98 tested, and 34 of 94 tested protein targets exhibited an ES score ≥ 9 (highlighted in grey shade in Table 1) in the pNESG_Avi6HT and pNESG_Nano6HT vectors, respectively. Several SDS-PAGE gels illustrating these expression and solubility enhancements are shown in Figure 3. Not only were more of these 99 human protein target genes expressed using TOEET, but both expression levels and solubility were generally increased. For example, while about half of the 99 protein targets had expression value E = 0 (i.e. no detectable expression) in the pET15_NESG vector (lacking TOEET), 95 of the

99 protein targets had expression values E ≥ 2 in either the pNESG_Nano6HT and pNESG_Avi6HT vectors (Table 1); many have E values E = 5 (the maximum level typically observed) in the expression vectors using TOEET.

Construct designs for a larger set of more than 2,000 human transcription factor proteins and domains are listed in Table 2. A large number of the proteins listed in Table 2 have been cloned into vectors optimized by TOEET, such as the pNESG_Nano6HT and pNESG_Avi6HT vectors, and exhibit high levels expression and solubility. Analysis of these data indicates that both the pNESG_Nano6HT vector and pNESG_Avi6HT vectors produced greater expression and solubility levels than a standard pET15_NESG vector that has not been optimized using the TOETT technology described in this disclosure.

Overall, TOEET allows for the production of a significantly greater number of human proteins and protein domains. The higher ES values obtained using TOETT also allow for simpler production and purification of the target proteins, since high ES scores mean that the cell extract has a larger amount of the target protein relative to background proteins.

The pNESG_Avi6HT also allows for the production of protein samples that can be readily biotinylated in the EET tag sequence. The pNESG_Nano6HT tag also provides a means for simple production of a streptavidin-binding protein (Scholle et al., 2004). Such biotinylated or Nano-tagged protein samples can be used for a variety of processes, including phage display antibody production, as well as for screening and discovering protein-protein and protein - nucleic acid interactions.

EXAMPLE 3

In certain applications, proteins that are expressed but not soluble in cell extracts can be solubilized and used successfully as antigens using various methods of solubilization, including urea and guanidine denaturants (Agaton et al. 2003). Accordingly, the ability to express a protein target, even it is not soluble in the high throughput Expression-Solubility screen described above [NESG High ThroughPut (HTP) Molecular Cloning and Expression Screening Platform methods] is critical, since if the protein cannot be expressed at all it is not possible to generate a suitable antigen. Accordingly, a particularly important value of the TOEET technology is enhancement of protein expression (E),

regardless of the resulting solubility. To illustrate this point, histogram plots are presented in Figures 7a and 7b comparing Expression scores (E ranging from 0 to 5) using the TOEET technology (E_TOEET) compared to expression scores for the same target protein using a pET vector lacking TOEET technology (E_pET). The data shown in Figure 7a is for 98 protein target genes cloned into the pNESG_Avi6HT TOEET vector compared with the exact same genes cloned into the pET15_NESG vector (lacking TOEET). The data shown in Figure 7b is for 94 protein target genes cloned pNESG_Nano6HT TOEET vectors compared with the exact same genes cloned into pET15_NESG vector (lacking TOEET). In these histogram plots, a value E_TOEET – E_pET = 0 indicates that the expression levels for both vectors were identical; values E_TOEET – E_pET > 0 indicate that the TOEET technology provided higher level expression, values E_TOEET – E_pET < 0 indicate that the TOEET technology provided lower level expression. For both target sets, the vast majority of genes exhibit much higher expression in the pNESG_Avi6HT TOEET and pNESG_Nano6HT TOEET vectors compared with the pET15_NESG vector (lacking TOEET). In many cases, E_TOEET – E_pET is 4 or 5, indicating that the expression in the non-TOEET vector was 0 or 1, which is too low to be useful for antigen production. Thus the TOEET vectors often provide high level expression of proteins which cannot be expressed at all, or those with are otherwise expressed as such marginal levels as to be useless for antigen production.

EXAMPLE 4

A representative method for practicing certain embodiments of the invention is described below.

The first step in the method is to identify the residues of the chosen tag/protein and the corresponding DNA sequences to be modified, for example, the 1st 30 residues of the tag/protein. Low usage codons are identified and are changed to optimal codons either manually or using servers, for example, such as http://www.jcat.de/ or http://genomes.urv.es/OPTIMIZER/, among others (Step 2). The transcription start site of vector and the resulting 5' untranslated region is then identified (Step 3). The 5' UTR RNA sequence is fused *in silico* with the optimized RNA sequence encoding the tag/protein (e.g., the first 30 residues of the tag/protein) (Step 4). Various RNA secondary structure

prediction methods may then be used to analyze the fused sequence, such as, for example: http://www.genebee.msu.su/services/rna2_reduced.html, http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi (Maximum Free Energy with partition function) or http://www.ncrna.org/centroidfold/ (Centroid Estimators-Statistical Decision Theory) (Step 5). The RBS and Initiation codon (IC) are then identified in the secondary structure prediction and the RNA positions in the first, e.g., 30 residues of the tag/protein that pair to the RBS/IC regions are determined (Step 6). Subsequently, alternative high frequency codons for the given residues base pairing with the RBS/IC are substituted and secondary structure is recalculated (Step 7). Steps 5 through 7 may be repeated until the secondary structure in RBS/IC is minimized and there is general agreement with the between the prediction servers (e.g., multiple predication servers may be used, such as the three servers listed above). This information is then used to design and produce the TOEET-optimized expression vector. Target proteins may then be cloned and expressed into the resulting expression system using the NESG Construct Optimization Software and High ThroughPut (HTP) Molecular Cloning and Expression Screening Platform and Automated Purification Pipeline methods, as outlined above.

Table 1. Expression Results

		pET15 NESG			pNESG Avi6HT			pNESG Nano6HT		
	Target	E	S	ES	E	S	ES	E	S	ES
1	HR3023-1-506	4	4	16	5	1	5	4	5	20
2	HR4485B-103-188	3	2	6	4	1	4	3	4	12
3	HR4538D	0	0	0	5	5	25	4	3	12
4	HR4622-24-281	5	5	25	5	5	25	4	5	20
5	HR4635D-55-161	3	4	12	5	2	10	4	0	0
6	HR4670B-55-202	4	4	16	5	5	25	3	5	15
7	HR4753B-182-262	3	3	9	2	2	4	0	0	0
8	HR5518A-127-207	4	4	16	3	5	15	4	5	20
9	HR6383B-119-195	4	4	16	4	2	8	3	1	3
10	HR6489A	2	0	0	5	3	15	4	4	16
11	HR6490A	0	0	0	4	4	16	3	3	9
12	HR6490A-30-371	0	0	0	5	4	20	3	3	9
13	HR6832	0	0	0	5	0	0			
14	HR6832-1-194	3	0	0	5	0	0			
15	HR6832A-33-194	3	0	0	5	0	0			
16	HR6832A-72-207	5	0	0	5	0	0			
17	HR6956-16-584	0	0	0	0	0	0	0	0	0
18	HR6956A-115-584	0	0	0				0	0	0
19	HR6956B-170-584	5	0	0	4	4	16	0	0	0
20	HR6956B-181-584	3	0	0	4	0	0	0	0	0
21	HR6956B-218-584	0	0	0	5	5	25	0	0	0
22	HR6956C-115-183	4	5	20	5	1	5	5	2	10
23	HR6956C-115-200	0	0	0	4	0	0	3	3	9
24	HR7049A-49-474	0	0	0	3	0	0	3	0	0
25	HR7049B-117-474	0	0	0	5	0	0	5	0	0
26	HR7049C-49-143	2	5	10	5	5	25	4	3	12
27	HR7049C-49-159	4	5	20	4	5	20	5	3	15
28	HR7049C-58-159	0	0	0	5	3	15	4	3	12
29	HR7097A-77-146	0	0	0	0	0	0	5	1	5
30	HR7097A-77-168	0	0	0	5	3	15	5	2	10
31	HR7097B-179-423	4	4	16	4	2	8	5	3	15
32	HR7097C-193-423	4	0	0	5	2	10	5	0	0
33	HR7130A-202-461	4	0	0	5	2	10	4	1	4
34	HR7130B-104-206	0	0	0	5	3	15	5	3	15
35	HR7133A-97-168	0	0	0	4	3	12	2	0	0
36	HR7133A-97-174	0	0	0	2	5	10	0	0	0
37	HR7133A-97-187	3	0	0	2	5	10	2	0	0
38	HR7133B-182-468	5	0	0	5	0	0	5	0	0
39	HR7133C-192-468	5	0	0	5	0	0	5	0	0
40	HR7224B-363-626	0	0	0	3	0	0	0	0	0
41	HR7224B-396-626	0	0	0	1	0	0	0	0	0
42	HR7224B-411-626	0	0	0	5	0	0	0	0	0

43	HR7224C-291-374	0	0	0	5	5	25	5	5	25
44	HR7224C-294-356	0	0	0	2	0	0	5	0	0
45	HR7370A-209-460	0	0	0	5	0	0	5	0	0
46	HR7372A-210-470	5	0	0	5	0	0	5	0	0
47	HR7372A-237-470	3	0	0	5	0	0	0	0	0
48	HR7372A-245-470	3	0	0	5	0	0	3	0	0
49	HR7378A-18-385	0	0	0	0	0	0	5	0	0
50	HR7378B-65-385	4	0	0	5	0	0	2	0	0
51	HR7378B-76-385	0	0	0	5	0	0	2	0	0
52	HR7378C-18-109	0	0	0	4	2	8	5	0	0
53	HR7378C-18-83	0	0	0	4	2	8	4	0	0
54	HR7378C-18-96	0	0	0	5	2	10	5	2	10
55	HR7469A-9-91	0	0	0	3	0	0	0	0	0
56	HR7469A-9-95	0	0	0	0	0	0	0	0	0
57	HR7469B-103-328	5	0	0	5	0	0	5	0	0
58	HR7475A-6-105	0	0	0	5	3	15	5	4	20
59	HR7475A-6-120	0	0	0	4	2	8	5	2	10
60	HR7475B-6-77	4	4	16	3	1	3	5	2	10
61	HR7475B-6-82	5	5	25	3	0	0	5	3	15
62	HR7475C-103-352	5	0	0	5	2	10	4	0	0
63	HR7515A-85-154	4	4	16	3	5	15	5	2	10
64	HR7515A-85-166	3	4	12	3	4	12	3	5	15
65	HR7515A-97-166	3	5	15	2	0	0	4	2	8
66	HR7515B-178-423	4	0	0	4	1	4	4	0	0
67	HR7515C-183-417	4	0	0	5	1	5	5	0	0
68	HR7515D-84-169	4	3	12	4	5	20	3	3	9
69	HR7522A-142-391	0	0	0	5	2	10	5	0	0
70	HR7522B-142-378	4	0	0	5	2	10	4	0	0
71	HR7522C-148-377	4	0	0	5	2	10	5	0	0
72	HR7522D-58-135	0	0	0	4	2	8	3	3	9
73	HR7564B-16-125	0	0	0	5	3	15	5	2	10
74	HR7636B-113-394	0	0	0	5	1	5	5	0	0
75	HR7785A-601-673	0	0	0	5	2	10	3	1	3
76	HR7785A-601-685	0	0	0	3	3	9	4	1	4
77	HR7785B-712-984	0	0	0	5	0	0	3	0	0
78	HR7870A-37-107	0	0	0	5	3	15	4	2	8
79	HR7870A-37-130	0	0	0	3	5	15	4	3	12
80	HR7870B-130-434	5	0	0	5	1	5	5	0	0
81	HR7870C-142-434	5	0	0	5	0	0	5	0	0
82	HR7906B-114-410	5	5	25	5	0	0	5	2	10
83	HR7906C-50-126	0	0	0	3	1	3	5	0	0
84	HR7993A-220-461	5	0	0	5	1	5	5	0	0
85	HR7993B-10-111	4	3	12	5	5	25	4	3	12
86	HR8028A-67-146	0	0	0	5	0	0	1	0	0
87	HR8028A-67-156	0	0	0	4	0	0	1	0	0
88	HR8028A-73-146	0	0	0	4	3	12	1	2	2

89	HR8028A-73-156	0	0	0	3	1	3	4	0	0
90	HR8028B-163-441	4	0	0	5	0	0	4	0	0
91	HR8234A-206-287	2	2	4	5	1	5			
92	HR8241A-261-342	0	0	0	5	5	25	3	5	15
93	HR8241A-264-328	0	0	0	2	0	0	4	0	0
94	HR8241B-328-598	0	0	0	5	0	0	0	0	0
95	HR8278A-123-216	0	0	0	4	4	16	3	0	0
96	HR8341A-100-171	0	0	0	3	1	3	3	2	6
97	HR8341A-100-181	0	0	0	5	2	10	2	5	10
98	HR8341A-100-200	0	0	0	4	2	8	4	1	4
99	HR8341B-381-579	5	4	20	5	3	15	5	2	10

E=Expression ; E=0-5 (no to high expression)
S=Solubility ; S= 0 - 5 (no to high solubility)
ES = E * S = (0 - 25) ES ≥ 9 usability (highlighted with grey fill)
ES ≥ 9 (typically results in ≥5 milligrams of protein per one liter of E. coli Fermentation)

Table 2. Human transcription factor protein and domain constructs designed using the NESG Construct Optimization Software for production using TOEET technologies. Each line in the table describes a unique protein construct for RT-PCR cloning, defined by the NESG Vector ID, the HUGO protein identifier, the Uniprot protein identifier, the first 15 amino acid residues in the targeted construct, the last 15 amino acid residues in the target construct, and the length of the targeted gene. The actual length of the targeted gene obtained by RT-PCR may be shorter or longer than indicated in the table due RNA spicing variations.

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7152A-140-202-TEV	ADAR	P55265	PVHYNGPSKAGYVDF	YSHGLPRCSPYKKLT	63
HR7675A-754-849-NHT	ADNP	Q9H2P0	LDPKGHEDDSYEARK	KHEMDFDAEWLFENH	96
HR7633A-1032-1131-NHT	ADNP2	Q6IQ32	KDEALQILALDPKKY	ELKNVKHRLNFEYEP	100
HR4425-1-595-15	AHR	P35869	MNSSSANITYASRKR	ILTYVQDSLSKSPFI	595
HR4425B-277-391-14	AHR	P35869	MILEIRTKNFIFRTK	DYIIVTQRPLTDEEG	116
HR4425B-277-406-14	AHR	P35869	MILEIRTKNFIFRTK	TEHLRKRNTKLPFMF	131
HR4425B-282-386-14	AHR	P35869	MTKNFIFRTKHKLDF	KNGRPDYIIVTQRPL	106
HR4425B-282-403-14	AHR	P35869	MTKNFIFRTKHKLDF	EEGTEHLRKRNTKLP	123
HR4425C-102-179-15	AHR	P35869	MRAANFREGLNLQEG	EDRAEFQRQLHWALN	79
HR4425C-108-178-14	AHR	P35869	MEGLNLQEGEFLLQA	TEDRAEFQRQLHWAL	72
HR4425C-97-184-15	AHR	P35869	MGQDNCRAANFREGL	FQRQLHWALNPSQCT	89
HR4425D-318-386-14	AHR	P35869	MRGSGYQFIHAADML	KNGRPDYIIVTQRPL	70
HR6398A-1-104-15	AIRE	O43918	MATDAALRRLLRLHR	YGRLQPILDSFPKDV	104
HR6398A-1-91-15	AIRE	O43918	MATDAALRRLLRLHR	FWRVLFKDYNLERYG	91
HR6398A-1-96-15	AIRE	O43918	MATDAALRRLLRLHR	FKDYNLERYGRLQPI	96
HR4766B-14-107-14	AKAP8	O43823	MGPANTQGAYGTGVA	IAKINQRLDMMSKEG	95
HR4766B-19-107-14	AKAP8	O43823	MQGAYGTGVASWQGY	IAKINQRLDMMSKEG	90

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8040A-384-551-Av6HT	AKAP8L	Q9ULX6	VERIQFVCSLCKYRT	KKLERYLKGENPFTD	168
HR6457-14	ALX1	Q15699	MEFLSEKFALKSPPS	RMKAKEHTANISWAM	326
HR7916A-159-235-Av6HT	ALX3	O95076	TFSTFQLEELEKVFQ	RNPFTAAYDISVLPR	77
HR4490C-209-280-NHT	ALX4	Q9H161	SNKGKKRRNRTTFTS	RAKWRKRERFGQMQQ	72
HR6941A-510-703-NHT	ANAPC2	Q9UJX6	GSKDLFINEYRSLLA	VALLRRRMSVWLQQG	194
HR6941A-511-695-Av6HT	ANAPC2	Q9UJX6	SKDLFINEYRSLLAD	LSKAVKMPVALLRRR	185
HR6941B-732-822-Av6HT	ANAPC2	Q9UJX6	SDDESDSGMASQADQ	LVYSAGVYRLPKNCS	91
HR6941B-765-822-Av6HT	ANAPC2	Q9UJX6	LESLSLDRIYNMLRM	LVYSAGVYRLPKNCS	58
HR6941C-498-713-Av6HT	ANAPC2	Q9UJX6	SSDIISLLVSIYGSK	WLQQGVLREEPPGTF	216
HR6941C-510-713-Av6HT	ANAPC2	Q9UJX6	GSKDLFINEYRSLLA	WLQQGVLREEPPGTF	204
HR8423A-486-593-Av6HT	ANKZF1	Q9H8Y5	AKAPGQPELWNALLA	STRNEFRRFMEKNPD	108
HR5083-14	APEX2	Q9UBZ4	MLRVVSWNINGIRRP	DPSSRCNFFLWSRPS	518
HR5083A-1-319-14	APEX2	Q9UBZ4	MLRVVSWNINGIRRP	CPVGAVLSVSSVPAK	319
HR5083A-1-323-14	APEX2	Q9UBZ4	MLRVVSWNINGIRRP	AVLSVSSVPAKQCPP	323
HR5083A-1-352-14	APEX2	Q9UBZ4	MLRVVSWNINGIRRP	KILRFLVPLEQSPVL	352
HR5083A-1-357-14	APEX2	Q9UBZ4	MLRVVSWNINGIRRP	LVPLEQSPVLEQSTL	357
HR8294A-15-116-TEV	APTX	Q7Z2E3	RVCWLVRQDSRHQRI	HMVNELYPYIVEFEE	100
HR7650B-267-331-TEV	ARHGAP35	Q9NRY4	SQQIATAKDKYEWLV	AKKLFLQHIHRLKHE	65
HR7542A-507-616-NHT	ARID2	Q68CP9	QHVAPPPGIVEIDSE	RAIPLPIQMYYQQQP	110
HR4394C-14	ARID3A	Q99856	MPDHGDWTYEEQFKQ	ELQAAIDSNRREGRR	135
HR4394C-15	ARID3A	Q99856	MPDHGDWTYEEQFKQ	ELQAAIDSNRREGRR	135
HR4394C-218-351-Av6HT	ARID3A	Q99856	MPDHGDWTYEEQFKQ	ELQAAIDSNRREGRR	135

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4394C-218-351-TEV	ARID3A	Q99856	PDHGDWTYEEQFKQL	ELQAAIDSNRREGRR	134
HR8410A-318-424-TEV	ARID5B	Q14865	RADEQAFLVALYKYM	KGEEDKPLPPIKPRK	107
HR7845A-354-470-TEV	ARNT	P27540	SNVCQPTEFISRHNI	YIICTNTNVKNSSQE	116
HR7821A-334-439-TEV	ARNT2	Q9HBZ2	PTEFLSRHNSDGIIT	SDEIEYIICTNTNVK	106
HR7274A-178-295-NHT	ARNTL2	Q8WYA1	QDNELRHLILKTAEG	SFFCRIKSCKISVKE	118
HR6915A-334-389-TEV	ARX	Q96QS3	TFTSYQLEELERAFQ	WFQNRRAKWRKREKA	56
HR4461B-112-194-14	ASCL1	P50553	MLPQQQPAAVARRNE	VSAAFQAGVLSPTIS	84
HR4461B-112-210-14	ASCL1	P50553	MLPQQQPAAVARRNE	NYSNDLNSMAGSPVS	100
HR4461B-132-189-14	ASCL1	P50553	MKLVNLGFATLREHV	DEHDAVSAAFQAGVL	59
HR4461B-132-206-14	ASCL1	P50553	MKLVNLGFATLREHV	TISPNYSNDLNSMAG	76
HR4461B-146-206-14	ASCL1	P50553	MPNGAANKKMSKVET	TISPNYSNDLNSMAG	62
HR4510B-64-138-14	ASCL2	Q99929	MKLVNLGFQALRQHV	AVRPSAPRGPPGTTP	76
HR7137A-2665-2824-TEV	ASH1L	Q9NR48	YLMRDSRRTPDGHPV	PKKLTPKKDFSPHYV	160
HR3149-106-270-14	ATF1	P18846	SGQYIAIAPNGALQL	IEELKTLKDLYSNKS	165
HR3149-14	ATF1	P18846	MEDSHKSTTSETAPQ	EELKTLKDLYSNKSV	271
HR3149-15	ATF1	P18846	MEDSHKSTTSETAPQ	EELKTLKDLYSNKSV	271
HR3149-21	ATF1	P18846	MEDSHKSTTSETAPQ	EELKTLKDLYSNKSV	271
HR3149-87-270-14	ATF1	P18846	GVSAAVTSMSVPTPI	IEELKTLKDLYSNKS	184
HR3149-96-270-14	ATF1	P18846	SVPTPIYQTSSGQYI	IEELKTLKDLYSNKS	175
HR4498B-354-414-TEV	ATF2	P15336	KRRKFLERNRAAASR	LLRNEVAQLKQLLLA	61
HR4572-21-181-14	ATF3	P18847	MPCLSPPGSLVFEDF	RNLFIQQIKEGTLQS	162
HR4572B-103-170-14	ATF3	P18847	MCRNKKKEKTECLQK	RAQNGRTPEDERNLF	69

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4572B-103-181-14	ATF3	P18847	MCRNKKKEKTECLQK	RNLFIQQIKEGTLQS	80
HR4572B-77-181-14	ATF3	P18847	MTKAEVAPEEDERKK	RNLFIQQIKEGTLQS	106
HR6914A-280-341-Av6HT	ATF4	P18848	MKKLKKMEQNKTAAT	LAKEIQYLKDLIEEV	63
HR6914A-280-341-TEV	ATF4	P18848	KKLKKMEQNKTAATR	LAKEIQYLKDLIEEV	62
HR4531-39-469-14	ATF7	P17544	MPARTDSVIIADQTP	SAAEAVATSVLTQMA	432
HR8374A-151-218-Av6HT	ATOH1	Q92858	KQVNGVQKQRRLAAN	AQIYINALSELLQTP	68
HR7270A-225-288-NHT	ATOH8	Q96SQ7	KALQQTRRLLANARE	IACNYILSLARLADL	64
HR7350A-7-128-TEV	BACH1	O14867	SVFAYESSVHSTNVL	SVHNIEESCFQFLKF	122
HR8413A-12-132-Av6HT	BACH2	Q9BYV9	YVYESTVHCTNILLG	MHNLEDSCFSFLQTQ	121
HR7112A-169-265-NHT	BARHL1	Q9BZE3	DSPPVRLKKPRKART	SALQRMFPSPYFYPQ	97
HR7390A-223-314-NHT	BARHL2	Q9NY43	ESPPVRAKKPRKART	EAGNYSALQRMFPSP	92
HR7183A-133-199-TEV	BARX1	Q9HBU1	GEPGTKAKKGRRSRT	QVKTWYQNRRMKWKK	67
HR7561-1-174-Av6HT	BARX2	Q9UMQ3	HCHAELRLSSPGQLK	TPDRLDLAQSLGLTQ	173
HR7561-1-187-Av6HT	BARX2	Q9UMQ3	HCHAELRLSSPGQLK	TQLQVKTWYQNRRMK	186
HR7561-1-196-Av6HT	BARX2	Q9UMQ3	HCHAELRLSSPGQLK	QNRRMKWKKMVLKGG	195
HR7561A-118-196-Av6HT	BARX2	Q9UMQ3	SSESETEQPTPRQKK	QNRRMKWKKMVLKGG	79
HR6459-34-125-14	BATF	Q16520	EKNRIAAQKSRQRQT	HAFHQPHVSSPRFQP	92
HR6459-34-125-15	BATF	Q16520	EKNRIAAQKSRQRQT	HAFHQPHVSSPRFQP	92
HR6459A-19-125-14	BATF	Q16520	GKQDSSDDVRRVQRR	HAFHQPHVSSPRFQP	107
HR6459A-34-118-14	BATF	Q16520	EKNRIAAQKSRQRQT	PEVVYSAHAFHQPHV	85
HR7115A-1107-1202-NHT	BAZ1A	Q9NRL2	RSYKTVLDRWRESLL	GDWFCPECRPKQRSR	96
HR7115B-420-468-Av6HT	BAZ1A	Q9NRL2	LPPEIFGDALMVLEF	LEVLEEALVGNDSEG	49

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7115B-420-486-Av6HT	BAZ1A	Q9NRL2	LPPEIFGDALMVLEF	ELLFFFLTAIFQAIA	67
HR7115C-1408-1534-Av6HT	BAZ1A	Q9NRL2	CRKRQSPEPSPVTLG	TRLQAFFHIQAQKLG	127
HR7115D-1420-1534-Av6HT	BAZ1A	Q9NRL2	TLGRRSSGRQGGVHE	TRLQAFFHIQAQKLG	115
HR7115D-1432-1534-Av6HT	BAZ1A	Q9NRL2	VHELSAFEQLVVELV	TRLQAFFHIQAQKLG	103
HR7115E-1-122-Av6HT	BAZ1A	Q9NRL2	PLLHRKPFVRQKPPA	IFAYVKDRYFVEETV	121
HR7115E-1-129-Av6HT	BAZ1A	Q9NRL2	PLLHRKPFVRQKPPA	RYFVEETVEVIRNNG	128
HR7115E-1-142-Av6HT	BAZ1A	Q9NRL2	PLLHRKPFVRQKPPA	NGARLQCRILEVLPP	141
HR7115E-22-122-Av6HT	BAZ1A	Q9NRL2	EEVFYCKVTNEIFRH	IFAYVKDRYFVEETV	101
HR7115E-22-129-Av6HT	BAZ1A	Q9NRL2	EEVFYCKVTNEIFRH	RYFVEETVEVIRNNG	108
HR7115E-22-142-Av6HT	BAZ1A	Q9NRL2	EEVFYCKVTNEIFRH	NGARLQCRILEVLPP	121
HR7190A-1634-1742-NHT	BAZ2A	Q9UIF9	SYEITPRIRVWRQTL	VEGEFTQKPGFPKRG	109
HR8090A-2062-2166-TEV	BAZ2B	Q9UIF8	DSKDLALCSMILTEM	NMRKYFEKKWTDTFK	105
HR7285A-80-154-TEV	BBX	Q8WY36	ARRPMNAFLLFCKRH	FMKANPGYKWCPTTN	75
HR4436B-523-602-14	BCL6	P41182	MCDCRFSEEASLKRH	NLKTHTRIHSGEKPY	81
HR4436B-523-606-14	BCL6	P41182	MCDCRFSEEASLKRH	HTRIHSGEKPYKCET	85
HR4436B-528-601-14	BCL6	P41182	MSEEASLKRHTLQTH	ANLKTHTRIHSGEKP	75
HR4436B-540-602-14	BCL6	P41182	MTHSDKPYKCDRCQA	NLKTHTRIHSGEKPY	64
HR4436B-542-598-14	BCL6	P41182	MSDKPYKCDRCQASF	NRPANLKTHTRIHSG	58
HR4436C-5-129-TEV	BCL6	P41182	ADSCIQFTRHASDVL	EHVVDTCRKFIKASE	125
HR7156A-284-387-Av6HT	BDP1	A6H8Y1	ERGSTTTYSSFRKNY	KVLAEEEKRKQKSVK	104
HR8401A-71-161-Av6HT	BHLHA15	Q7RTS1	DSSIQRRLESNERER	PKLYQHYQQQQQVAG	91
HR7639A-64-125-Av6HT	BHLHA9	Q7RTU4	KARRMAANVRERKRI	IHRIAALSLVLRASP	62

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8288A-236-314-Av6HT	BHLHE22	Q8NFJ8	KSKEQKALRLNINAR	LEEMRRLVAYLNQGQ	79
HR7576A-47-183-NHT	BHLHE40	O14503	EDSKETYKLPHRLIE	SQLVTHLHRVVSELL	137
HR7576B-142-174-Av6HT	BHLHE40	O14503	FCSGFQTCAREVLQY	HENTRDLKSSQLVTH	33
HR7576B-142-181-Av6HT	BHLHE40	O14503	FCSGFQTCAREVLQY	KSSQLVTHLHRVVSE	40
HR7518A-44-116-NHT	BHLHE41	Q9C0J9	TYKLPHRLIEKKRRD	LTEQQHQKIIALQNG	73
HR3082 1-125 pET15TEV_NESG (in progress)	BLOC1S1	P78537	MLSRLLKEHQAKQNE	ALEYVYKGQLQSAPS	125
HR3082-1-119-14	BLOC1S1	P78537	MLSRLLKEHQAKQNE	MRTIATALEYVYKGQ	119
HR3082-14	BLOC1S1	P78537	MLSRLLKEHQAKQNE	ALEYVYKGQLQSAPS	125
HR3082-MBP3	BLOC1S1	P78537	MLSRLLKEHQAKQNE	LEYVYKGQLQSAPS*	126
HR3082A-32-125-14	BLOC1S1	P78537	TCLTEALVDHLNVGV	ALEYVYKGQLQSAPS	94
HR3082B-43-125-15	BLOC1S1	P78537	MNVGVAQAYMNQRKL	ALEYVYKGQLQSAPS	84
HR7816A-294-396-NHT	BMP2	P12643	SSCKRHPLYVDFSDV	LKNYQDMVVEGCGCR	103
HR7816A-294-396-TEV	BMP2	P12643	SSCKRHPLYVDFSDV	LKNYQDMVVEGCGCR	103
HR7409A-9-128-TEV	BOLA1	Q9Y3E2	GLVSMAGRVCLCQGS	WRENSQLDTSPPCLG	120
HR8185-1-86-TEV	BOLA2B	Q9H3K6	ASAKSLDRWKARLLE	EYLREKLQRDLEAEH	85
HR7562-8-107-Av6HT	BOLA3	Q53S33	AAAPLLRGIRGLPLH	KEMHGLRIFTSVPKR	100
HR7886A-6-308-TEV	BPNT1	O95861	TVLMRLVASAYSIAQ	YASRVPESIKNALVP	303
HR7955A-134-243-TEV	BRD9	Q9H8M2	KDKIVANEYKSVTEF	EPEGNACSLTDSTAE	110
HR6995B-633-746-TEV	BRPF1	P55201	FLILLRKTLEQLQEK	GAVLRQARRQAEKMG	114
HR8142A-104-176-Av6HT	BSX	Q3C1V8	PGKHCRRRKARTVFS	RMKHKKQLRKSQDEP	73
HR1875-14	C12orf28	Q96LU7	MAFCALTIVALYILS	IFFTDYFFYFYRRCA	275
HR7476A-824-884-NHT	C14orf43	Q6PJG2	TYHYTGSDQWKMAER	FYYTYKKQVKIGRNG	61

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7019A-867-954-TEV	CAMTA1	Q9Y6Y1	SGRVFMVTDYSPEWS	NNQIISNSVVFEYKA	88
HR7019B-108-183-TEV	CAMTA1	Q9Y6Y1	ILYNRKKVKYRKDGY	LQNPDIVLVHYLNVP	76
HR7019B-69-183-TEV	CAMTA1	Q9Y6Y1	KERHRWNTNEEIAAY	LQNPDIVLVHYLNVP	115
HR7019B-73-183-TEV	CAMTA1	Q9Y6Y1	RWNTNEEIAAYLITF	LQNPDIVLVHYLNVP	111
HR7019C-1029-1162-Av6HT	CAMTA1	Q9Y6Y1	ALGSCFESRVVVVCE	LGIARSRGHVKLAEC	134
HR7019C-1029-1168-Av6HT	CAMTA1	Q9Y6Y1	ALGSCFESRVVVVCE	RGHVKLAECLEHLQR	140
HR7019C-1058-1162-Av6HT	CAMTA1	Q9Y6Y1	IHSKTFRGMTLLHLA	LGIARSRGHVKLAEC	105
HR7019C-1058-1168-Av6HT	CAMTA1	Q9Y6Y1	IHSKTFRGMTLLHLA	RGHVKLAECLEHLQR	111
HR7019D-1486-1624-Av6HT	CAMTA1	Q9Y6Y1	KPNLPSAADWSEFLS	CGKRRQARRTAVIVQ	139
HR7019D-1486-1660-Av6HT	CAMTA1	Q9Y6Y1	KPNLPSAADWSEFLS	FLRRCRHSPLVDHRL	175
HR7019D-1501-1673-Av6HT	CAMTA1	Q9Y6Y1	ASTSEKVENEFAQLT	RLYKRSERIEKGQGT	173
HR7019D-1513-1624-Av6HT	CAMTA1	Q9Y6Y1	QLTLSDHEQRELYEA	CGKRRQARRTAVIVQ	112
HR7019D-1513-1660-Av6HT	CAMTA1	Q9Y6Y1	QLTLSDHEQRELYEA	FLRRCRHSPLVDHRL	148
HR7295A-60-130-Av6HT	CARHSP1	Q9Y2V2	GPVYKGVCKCFGRSK	PKNEKLQAVEVVITH	71
HR8150A-1916-1982-Av6HT	CASP8AP2	Q9UKL3	KNVIKKKGEIIILWT	RFQQLMKLFEKSKCR	67
HR7269A-2-135-Av6HT	CBFB	Q13951	MPRVVPDQRSKFENE	GMGCLEFDEERAQQE	135
HR7269A-2-135-TEV	CBFB	Q13951	PRVVPDQRSKFENEE	GMGCLEFDEERAQQE	134
HR7615A-104-190-NHT	CBLL1	Q75N03	TPVHFCDKCGLPIKI	YLSQRDLQAHINHRH	87
HR6520A-9-62-Av6HT	CBX2	Q14781	MEQVFAAECILSKRL	NILDPRLLLAFQKKE	55
HR6520A-9-62-TEV	CBX2	Q14781	EQVFAAECILSKRLR	NILDPRLLLAFQKKE	54
HR8494A-624-717-Av6HT	CCDC79	Q8NA31	IVEAEDRYKSELRKS	QQGRKAVDLAHKYHK	94
HR8086A-57-112-TEV	CDC5L	Q99459	SIKKTEWSREEEEKL	EHYEFLLDKAAQRDN	56

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7252A-160-214-Av6HT	CDX1	P47902	VYTDHQRLELEKEFH	IWFQNRRAKERKVNK	55
HR7064A-185-251-Av6HT	CDX2	Q99626	TKDKYRVVYTDHQRL	RRAKERKNKKKLQQ	67
HR7957A-172-246-Av6HT	CDX4	O14627	TKEKYRVVYTDHQRL	IKKKISQFENSGGSV	75
HR7823A-281-340-TEV	CEBPA	P49715	NSNEYRVRRERNNIA	KRVEQLSRELDTLRG	60
HR4764B-273-336-TEV	CEBPB	P17676	EYKIRRERNNIAVRK	RELSTLRNLFKQLPE	64
HR7557A-190-272-15	CEBPE	Q15744	MAGPLHKGKKAVNKD	DTLRNLFRQIPEAAN	84
HR7557A-195-268-15	CEBPE	Q15744	MKGKKAVNKDSLEYR	TQELDTLRNLFRQIP	75
HR7557A-195-281-15	CEBPE	Q15744	MKGKKAVNKDSLEYR	IPEAANLIKGVGGCS	88
HR7557A-203-281-15	CEBPE	Q15744	MDSLEYRLRRERNNI	IPEAANLIKGVGGCS	80
HR6439-59-150-Av6HT	CEBPG	P53567	DRNSDEYRQRRERNN	ISTENTTADGDNAGQ	92
HR8022A-431-525-Av6HT	CENPT	Q96BT3	EPAEPLLVRHPPRPR	KPEDLELLMRRQGLV	95
HR7210A-268-373-Av6HT	CHD1	O14646	MEEEFETIERFMDCR	TKRWLKNASPEDVEY	107
HR7210A-268-373-TEV	CHD1	O14646	EEEFETIERFMDCRI	TKRWLKNASPEDVEY	106
HR7330A-260-394-NHT	CHD2	O14647	SETIEKVLDSRLGKK	VERVIAVKTSKSTLG	135
HR7397A-371-431-TEV	CHD6	Q8TD26	NPDYVEVDRILEVAH	DVDPAKVKEFESLQV	61
HR7397B-679-941-Av6HT	CHD6	Q8TD26	LRRLKDDVEKNLAPK	LDKAVLQDINRKGGT	262
HR7397B-679-968-Av6HT	CHD6	Q8TD26	LRRLKDDVEKNLAPK	DLLRKGAYGALMDEE	289
HR7397B-679-974-Av6HT	CHD6	Q8TD26	LRRLKDDVEKNLAPK	AYGALMDEEDEGSKF	295
HR7397B-679-997-Av6HT	CHD6	Q8TD26	LRRLKDDVEKNLAPK	LQRRTHTITIQSEGK	318
HR8217A-2631-2715-TEV	CHD7	Q9P2D1	RNPNKLDINTLTGEE	DRLLTGPVVRGEGAS	85
HR8217B-2561-2614-TEV	CHD7	Q9P2D1	GQLDPDTRIPVINLE	TYTVDMPSYVPKNAD	54
HR7629A-1-98-NHT	CHRAC1	Q9NRG0	ADVVVGKDKGGEQRL	SETFQFLADILPKKI	97

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7688A-98-186-NHT	CLOCK	O15516	QDWKPTFLSNEEFTQ	THLLESDSLTPEYLK	89
HR7654A-1987-2362-TEV	CNOT1	A5YKK6	QLPYHRIFIMLLLEL	EIEKLFQSVAQCCMG	376
HR7654B-1987-2369-TEV	CNOT1	A5YKK6	QLPYHRIFIMLLLEL	SVAQCCMGQKQAQQV	383
HR7654B-1987-2376-TEV	CNOT1	A5YKK6	QLPYHRIFIMLLLEL	GQKQAQQVMEGTGAS	390
HR2981-28-443-15	COPS2	P61201	MPNVDLENQYYNSKA	NQLNSLNQAVVSKLA	417
HR2981-301-418-14	COPS2	P61201	PYKNDPEILAMTNLV	QVNQLLELDHQKRGG	118
HR2981-45-443-15	COPS2	P61201	MDDPKAALSSFQKVL	NQLNSLNQAVVSKLA	400
HR2981A-306-415-14	COPS2	P61201	PEILAMTNLVSAYQN	RIDQVNQLLELDHQK	110
HR2981B-339-418-14	COPS2	P61201	DDPFIREHIEELLRN	QVNQLLELDHQKRGG	80
HR2981C-45-163-15	COPS2	P61201	MDDPKAALSSFQKVL	FKTNTKLGKLYLERE	120
HR2981C-45-184-15	COPS2	P61201	MDDPKAALSSFQKVL	KILRQLHQSCQTDDG	141
HR2981C-45-210-15	COPS2	P61201	MDDPKAALSSFQKVL	EIYALEIQMYTAQKN	167
HR3016-1-411-14	COPS3	Q9UNS2	MASALEQFVNSVRQL	ITVNPQFVQKSMGSQ	411
HR3016-14	COPS3	Q9UNS2	MASALEQFVNSVRQL	GSQEDDSGNKPSSYS	423
HR3016A-49-114-15	COPS3	Q9UNS2	LDVQEHSLGVLAVLF	FAGLCHQLTNALVER	66
HR3016B-88-154-15	COPS3	Q9UNS2	CNGEHIRYATDTFAG	SIHADLCQLCLLAKC	67
HR3016C-270-368-15	COPS3	Q9UNS2	NNPSELRNLVNKHSE	KDGMVSFHDNPEKYN	99
HR3016C-270-396-15	COPS3	Q9UNS2	NNPSELRNLVNKHSE	KCIELDERLKAMDQE	127
HR3016C-270-411-15	COPS3	Q9UNS2	NNPSELRNLVNKHSE	ITVNPQFVQKSMGSQ	142
HR3016D-352-413-15	COPS3	Q9UNS2	NQKDGMVSFHDNPEK	VNPQFVQKSMGSQED	62
HR3016D-358-409-15	COPS3	Q9UNS2	VSFHDNPEKYNNPAM	QEITVNPQFVQKSMG	52
HR3105-1-292-15	COPS4	Q9BT78	MAAAVRQDLAQLMNS	LQEFAAMLMPHQKAT	292

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR3105-1-297-15	COPS4	Q9BT78	MAAAVRQDLAQLMNS	AMLMPHQKATTADGS	297
HR3105-15	COPS4	Q9BT78	MAAAVRQDLAQLMNS	APEWTAQAMEAQMAQ	406
HR6309A-15	CPSF4	O95639	MSGEKTVVCKHWLRG	NKECPFLHIDPESKI	62
HR7458A-62-130-NHT	CPSF4L	A6NMK7	GEKMVVCKHWLRGLC	KPAFKSQDCPWYDQG	69
HR3140-21	CREB1	P16220	MTMESGAENQQSGDA	EELKALKDLYCHKSD	341
HR6927-139-461-Av6HT	CREB3L3	Q68CI9	PVIQVPEASVTIDLE	TGSGRAGLEAAGDEL	323
HR7960-1-298-Av6HT	CREB3L4	Q8TEY5	DLGIPDLLDAWLEPP	IAQTSNKAAQTSTCV	297
HR6873A-34-89-Av6HT	CRX	O43186	SAPRKQRRERTTFTR	KINLPESRVQVWFKN	56
HR6873A-45-103-Av6HT	CRX	O43186	TFTRSQLEELEALFA	NRRAKCRQQRQQQKQ	59
HR8272A-84-160-NHT	CSDA	P16989	KKVLATKVLGTVKWF	VEGEKGAEAANVTGP	77
HR7792B-673-744-TEV	CSDE1	O75534	LRRATVECVKDQFGF	CSACNVWRVCEGPKA	72
HR7173A-399-462-TEV	CTCF	P49711	RTHSGEKPYECYICH	RKSDLGVHLRKQHSY	64
HR7173B-515-592-Av6HT	CTCF	P49711	MRTHTGEKPYACSHC	AGPDGVEGENGGETK	79
HR7173B-515-592-TEV	CTCF	P49711	RTHTGEKPYACSHCD	AGPDGVEGENGGETK	78
HR7558A-411-776-TEV	CUL1	Q13616	AQSSSKSPELLARYC	LERVDGEKDTYSYLA	365
HR7558B-15-410-Av6HT	CUL1	Q13616	IGLDQIWDDLRAGIQ	DKACGRFINNNAVTK	396
HR3327B-643-745-14	CUL2	Q13617	NFSSKRTKFKITTSM	IERSQASADEYSYVA	103
HR3327B-655-745-14	CUL2	Q13617	TSMQKDTPQEMEQTR	IERSQASADEYSYVA	91
HR3327C-1-408-15	CUL2	Q13617	MSLKPRVVDFDETWN	YCDNLLKKSAKGMTE	408
HR3327D-4-745-TEV	CUL2	Q13617	KPRVVDFDETWNKLL	IERSQASADEYSYVA	742
HR3437D-677-768-TEV	CUL3	Q13618	VAAKQGESDPERKET	LARTPEDRKVYTYVA	92
HR3342C-672-759-TEV	CUL4A	Q13619	IQMKETVEEQVSTTE	MERDKDNPNQYHYVA	88

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7263A-808-895-TEV	CUL4B	Q13620	IQMKETVEEQASTTE	MERDKENPNQYNYIA	88
HR3340C-7-395-Av6HT	CUL5	Q93034	LKNKGSLQFEDKWDF	TIFKLELPLKQKGVG	389
HR8510A-825-917-TEV	CUX2	O14529	PRGDEAPVPPEDEAA	RQVKEKLAKNGICQR	93
HR7807A-15-94-Av6HT	CXXC1	Q9P0U4	EDSKSENGENAPIYC	LEIRYRHKKSRERDG	80
HR7690A-184-282-TEV	DACH1	Q9UI36	TPQNNECKMVDLRGA	LISRKDFETLYNDCT	99
HR6867A-61-162-TEV	DACH2	Q96NX9	GNTNTNECRMVDMHG	TRKDFETLFTDCTNA	102
HR7176A-249-325-Av6HT	DBP	Q10586	VPEEQKDEKYWSRRY	YRAVLSRYQAQHGAL	77
HR7176A-254-325-Av6HT	DBP	Q10586	KDEKYWSRRYKNNEA	YRAVLSRYQAQHGAL	72
HR7911A-178-244-Av6HT	DBX2	Q6ZNG2	DSNSKARRGILRRAV	VKIWFQNRRMKWRNS	67
HR7702A-201-279-NHT	DEAF1	O75398	SELPVRCRNISGTLY	CLIQDGILNPHAASC	79
HR7922A-11-108-TEV	DEPDC1A	Q5TB30	YRATKLWNEVTTSFR	SENVDDNNQLFRFPA	98
HR7073A-153-209-TEV	DLX2	Q07687	RKPRTIYSSFQLAAL	QVKIWFQNRRSKFKK	57
HR8208A-130-186-TEV	DLX3	O60479	RKPRTIYSSYQLAAL	QVKIWFQNRRSKFKK	57
HR7595A-138-194-TEV	DLX5	P56178	RKPRTIYSSFQLAAL	QVKIWFQNKRSKIKK	57
HR8524A-46-106-TEV	DLX6	P56179	GKKIRKPRTIYSSLQ	QVKIWFQNKRSKFKK	61
HR4696-44-404-15	DMAP1	Q9NPF5	MTLTFKRPEGMHREV	MLRHRHEALARAGVL	362
HR4696-49-403-15	DMAP1	Q9NPF5	MRPEGMHREVYALLY	QMLRHRHEALARAGV	356
HR4696B-208-404-15	DMAP1	Q9NPF5	MVPGTDLKIPVFDAG	MLRHRHEALARAGVL	198
HR4696B-213-404-15	DMAP1	Q9NPF5	MLKIPVFDAGHERRR	MLRHRHEALARAGVL	193
HR4696B-236-403-15	DMAP1	Q9NPF5	MRTPEQVAEEEYLLQ	QMLRHRHEALARAGV	169
HR7582A-79-146-Av6HT	DMBX1	Q8NFW5	TAQQLEALEKTFQKT	SLQKEQLQKQKEAEG	68
HR7142-1-340-TEV	DMC1	Q14565	KEDQVVAEEPGFQDE	ATFAITAGGIGDAKE	339

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8371A-114-182-Av6HT	DMRT2	Q9Y5R5	PRKLSRTPKCARCRN	LRRQQATEDKKGLSG	69
HR7805A-20-88-NHT	DMRT3	Q9NQL9	RAPLQRTPKCARCRN	LRRQQANESLESLIP	69
HR6947A-318-361-NHT	DMRTA1	Q5VZB9	SLPTVSSRPRDPLDI	GILRFCKGDVVQAIE	44
HR7753A-1-53-NHT	DMRTB1	Q96MA1	ADKMVRTPKCSRCRN	KCYLISERQKIMAAQ	52
HR7387A-44-84-Av6HT	DMRTC2	Q8IXT2	RCRNHGVTAHLKGHK	KCVLILERRRVMAAQ	41
HR8011A-205-276-15	DMTF1	Q9Y222	MSTEPGDIVTQGVSW	RIAELDVADENDINW	73
HR8011A-205-293-15	DMTF1	Q9Y222	MSTEPGDIVTQGVSW	LAEGWSSVRSPQWLR	90
HR8011B-255-339-15	DMTF1	Q9Y222	MDEINLILRIAELDV	QKNNPTLLENKSGSG	86
HR8011B-255-356-15	DMTF1	Q9Y222	MDEINLILRIAELDV	NSNTNSSVQHVQIRV	103
HR8011B-268-339-15	DMTF1	Q9Y222	MVADENDINWDLLAE	QKNNPTLLENKSGSG	73
HR8011B-268-356-15	DMTF1	Q9Y222	MVADENDINWDLLAE	NSNTNSSVQHVQIRV	90
HR6887A-327-385-TEV	DNAJC1	Q96KC8	APEWTEEDLSQLTRS	AKQLKDSVTCSPGMV	59
HR7581A-1-76-NHT	DNAJC21	Q5F1R6	KCHYEALGVRRDASE	RAWYDNHREALLKGG	75
HR8109A-314-391-Av6HT	DPF2	Q92785	AAVKTYRWQCIECKC	LLKEKASIYQNQNSS	77
HR8202A-15-83-Av6HT	DPRX	A6NFQ7	HSHRKRTMFTKKQLE	AKLKKAKCKHIHQKQ	69
HR7601-1-176-Av6HT	DR1	Q01658	MASSSGNDDDLTIPR	NQAGSSQDEEDDDDI	176
HR7601-1-176-TEV	DR1	Q01658	ASSSGNDDDLTIPRA	NQAGSSQDEEDDDDI	175
HR6975A-1-77-TEV	DRAP1	Q14919	PSKKKKYNARFPPAR	KTMTTSHLKQCIELE	76
HR7517A-25-174-NHT	DUSP12	Q9UNI6	GQMLEVQPGLYFGGA	WQLKLYQAMGYEVDT	150
HR7523A-86-164-NHT	DUXA	A6NLW8	SQGQDQPGVEFQSRE	QNRRSRLLLQRKREP	79
HR4713B-251-345-TEV	DVL1	O14640	TVTLNMERHHFLGIS	ISLTVAKCWDPTPRS	95
HR5191A-15	DVL1L1	P54792	MTVTLNMERHHFLGI	ISLTVAKAWDPTPRS	96

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4606C-408-551-14	DVL2	O14641	MLPDGCEGRGLSVHT	APLPGATPWPLLPTF	145
HR4606C-412-526-14	DVL2	O14641	MCEGRGLSVHTDMAS	CESYLVNLSLNDNDG	116
HR4606C-417-519-14	DVL2	O14641	MLSVHTDMASVTKAM	FGDLSGGCESYLVNL	104
HR4606C-417-551-14	DVL2	O14641	MLSVHTDMASVTKAM	APLPGATPWPLLPTF	136
HR4606D-260-358-TEV	DVL2	O14641	TMSLNIITVTLNMEK	PGPIVLTVAKCWDPS	99
HR5528A-14	DVL2	O14641	MTITSGSSLPDGCEG	SEQCYYVFGDLSGGC	113
HR5528A-15	DVL2	O14641	MTITSGSSLPDGCEG	SEQCYYVFGDLSGGC	113
HR7051A-248-338-TEV	DVL3	Q92997	ITVTLNMEKYNFLGI	HKPGPITLTVAKCWD	91
HR7051B-397-504-15	DVL3	Q92997	MDTERLDDFHLSIHS	CYYIFGDLCGNMANL	109
HR7051B-397-511-15	DVL3	Q92997	MDTERLDDFHLSIHS	LCGNMANLSLHDHDG	116
HR7051B-397-530-15	DVL3	Q92997	MDTERLDDFHLSIHS	SDQDTLAPLPHPGAA	135
HR7051B-403-504-15	DVL3	Q92997	MDFHLSIHSDMAAIV	CYYIFGDLCGNMANL	103
HR7051B-403-530-15	DVL3	Q92997	MDFHLSIHSDMAAIV	SDQDTLAPLPHPGAA	129
HR7051C-1-79-TEV	DVL3	Q92997	GETKIYHLDGQETP	AKLPCFNGRVVSWLV	78
HR4672B-14	E2F1	Q01094	PGEKSRYETSLNLTT	QGPIDVFLCPEETVG	183
HR4672C-116-196-14	E2F1	Q01094	MGKGVKSPGEKSRYE	KKSKNHIQWLGSHTT	82
HR4672C-121-192-14	E2F1	Q01094	MSPGEKSRYETSLNL	QLIAKKSKNHIQWLG	73
HR4672C-122-196-14	E2F1	Q01094	MPGEKSRYETSLNLT	KKSKNHIQWLGSHTT	76
HR4672C-127-192-14	E2F1	Q01094	MRYETSLNLTTKRFL	QLIAKKSKNHIQWLG	67
HR4672C-147-192-14	E2F1	Q01094	MADGVVDLNWAAEVL	QLIAKKSKNHIQWLG	47
HR4672D-192-301-TEV	E2F1	Q01094	GSHTTVGVGGRLEGL	KSKQGPIDVFLCPEE	110
HR6383-65-437-14	E2F2	Q14209	ATPHGPEGQVVRCLP	SDLFDSYDLGDLLIN	373

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6383-70-437-14	E2F2	Q14209	PEGQVVRCLPAGRLP	SDLFDSYDLGDLLIN	368
HR6383A-195-308-15	E2F2	Q14209	RGMFEDPTRPGKQQQ	TQGPIEVYLCPEEVQ	114
HR6383A-198-296-15	E2F2	Q14209	FEDPTRPGKQQQLGQ	RTEDNLQIYLKSTQG	99
HR6383B-114-200-15	E2F2	Q14209	MGLPSPKTPKSPGEK	AKNNIQWVGRGMFED	88
HR6383B-114-204-15	E2F2	Q14209	MGLPSPKTPKSPGEK	IQWVGRGMFEDPTRP	92
HR6383B-119-195-15	E2F2	Q14209	MKTPKSPGEKTRYDT	LIRKKAKNNIQWVGR	78
HR6383B-119-195-Av6HT	E2F2	Q14209	KTPKSPGEKTRYDTS	LIRKKAKNNIQWVGR	77
HR6383B-119-195-TEV	E2F2	Q14209	KTPKSPGEKTRYDTS	LIRKKAKNNIQWVGR	77
HR6383C-126-200-15	E2F2	Q14209	MEKTRYDTSLGLLTK	AKNNIQWVGRGMFED	76
HR6383C-131-195-15	E2F2	Q14209	MDTSLGLLTKKFIYL	LIRKKAKNNIQWVGR	66
HR6383C-131-202-15	E2F2	Q14209	MDTSLGLLTKKFIYL	NNIQWVGRGMFEDPT	73
HR4418C-14	E2F3	O00716	KTRYDTSLGLLTKKF	QGPIEVYLCPEETET	182
HR4418D-14	E2F3	O00716	NNVQWMGCSLSEDGG	LCPEETETHSPMKTN	128
HR4470C-84-203-14	E2F4	Q16254	MVGPGCNTREIADKL	PIEVLLVNKEAWSSP	121
HR4470C-89-200-14	E2F4	Q16254	MNTREIADKLIELKA	VSGPIEVLLVNKEAW	113
HR4470D-11-86-TEV	E2F4	Q16254	PPGTPSRHEKSLGLL	EKKSKNSIQWKGVGP	76
HR4678B-113-232-14	E2F5	Q15329	MQWKGVGAGCNTKEV	KSHSGPIHVLLINKE	121
HR4678B-119-232-14	E2F5	Q15329	MAGCNTKEVIDRLRY	KSHSGPIHVLLINKE	115
HR4622-1-237-15	E2F6	O75461	MSQQRPARKLPSLLL	HIRSTNGPIDVYLCE	237
HR4622-1-242-15	E2F6	O75461	MSQQRPARKLPSLLL	NGPIDVYLCEVEQGQ	242
HR4622-19-242-15	E2F6	O75461	MEETVRRRCRDPINV	NGPIDVYLCEVEQGQ	225
HR4622-19-281-15	E2F6	O75461	MEETVRRRCRDPINV	EENPQQSEELLEVSN	264

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4622-24-237-15	E2F6	O75461	MRRCRDPINVEGLLP	HIRSTNGPIDVYLCE	215
HR4622-24-281-15	E2F6	O75461	MRRCRDPINVEGLLP	EENPQQSEELLEVSN	259
HR4622-24-281-Av6HT	E2F6	O75461	RRCRDPINVEGLLPS	EENPQQSEELLEVSN	258
HR4622-24-281-TEV	E2F6	O75461	RRCRDPINVEGLLPS	EENPQQSEELLEVSN	258
HR4622B-128-247-15	E2F6	O75461	GSDLSNFGAVPQQKK	VYLCEVEQGQTSNKR	120
HR4622B-133-243-15	E2F6	O75461	NFGAVPQQKKLQEEL	GPIDVYLCEVEQGQT	111
HR4622C-127-242-15	E2F6	O75461	IGSDLSNFGAVPQQK	NGPIDVYLCEVEQGQ	116
HR4622C-132-237-15	E2F6	O75461	SNFGAVPQQKKLQEE	HIRSTNGPIDVYLCE	106
HR4622D-54-137-15	E2F6	O75461	RKALKVKRPRFDVSL	HIRWIGSDLSNFGAV	84
HR4622D-54-180-15	E2F6	O75461	RKALKVKRPRFDVSL	QQLFELTDDKENERL	127
HR4622D-54-242-15	E2F6	O75461	RKALKVKRPRFDVSL	NGPIDVYLCEVEQGQ	189
HR4622D-58-132-15	E2F6	O75461	KVKRPRFDVSLVYLT	KKSKNHIRWIGSDLS	75
HR4622D-58-175-15	E2F6	O75461	KVKRPRFDVSLVYLT	IKDCAQQLFELTDDK	118
HR4622D-58-237-15	E2F6	O75461	KVKRPRFDVSLVYLT	HIRSTNGPIDVYLCE	180
HR8499A-141-251-Av6HT	E2F7	Q96AV8	SRKQKSLGLLCQKFL	YLQQKELDLIDYKFG	111
HR7611A-112-223-NHT	E2F8	A0AVK6	SRKEKSLGLLCHKFL	IKKKEYEQEPDFIKS	112
HR8342-1-508-Av6HT	E4F1	Q66K89	EGAMAVRVTAAHTAE	GDCGKLYKTIAHVRG	507
HR8342-1-600-Av6HT	E4F1	Q66K89	EGAMAVRVTAAHTAE	EHGTLNRHLRTKGGC	599
HR8342A-522-586-15	E4F1	Q66K89	MPKCGKRYKTKNAQQ	EKPFKCYKCGRGFAE	66
HR8342A-523-600-15	E4F1	Q66K89	MKCGKRYKTKNAQQV	EHGTLNRHLRTKGGC	79
HR8342A-527-581-15	E4F1	Q66K89	MRYKTKNAQQVHFRT	RHHTGEKPFKCYKCG	56
HR8342A-527-600-15	E4F1	Q66K89	MRYKTKNAQQVHFRT	EHGTLNRHLRTKGGC	75

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8342B-51-219-15	E4F1	Q66K89	MEEDEDDVHRCGRCQ	SILKAHMVTHSSRKD	170
HR8342B-51-231-15	E4F1	Q66K89	MEEDEDDVHRCGRCQ	RKDHECKLCGASFRT	182
HR8342B-51-249-15	E4F1	Q66K89	MEEDEDDVHRCGRCQ	LIRHHRRHTDERPYK	200
HR8342B-56-214-15	E4F1	Q66K89	MDVHRCGRCQAEFTA	TFKTGSILKAHMVTH	160
HR8342B-56-226-15	E4F1	Q66K89	MDVHRCGRCQAEFTA	VTHSSRKDHECKLCG	172
HR8342B-56-244-15	E4F1	Q66K89	MDVHRCGRCQAEFTA	RTKGSLIRHHRRHTD	190
HR3014A-10-250-TEV	EBF1	Q9UH73	RSGSSMKEEPLGSGM	NNSKHGRRARRLDPS	241
HR7745A-10-250-TEV	EBF3	Q9H4W6	RGGTTMKEEPLGSGM	NNSKHGRRARRLDPS	241
HR6883A-10-251-TEV	EBF4	Q9BQW3	NLKEEPLLPAGLGSV	HGRRARRLDPSEAAT	242
HR7307A-71-148-Av6HT	EDF1	O60869	MDRVTLEVGKVIQQG	GKDIGKPIEKGPRAK	79
HR7307A-71-148-TEV	EDF1	O60869	DRVTLEVGKVIQQGR	GKDIGKPIEKGPRAK	78
HR7944A-1347-1411-TEV	EEA1	Q15075	RKWAEDNEVQNCMAC	KPVRVCDACFNDLQG	65
HR4555D-366-418-Av6HT	EGR1	P18146	MKPFQCRICMRNFSR	KFARSDERKRHTKIH	54
HR4555D-366-418-TEV	EGR1	P18146	KPFQCRICMRNFSRS	KFARSDERKRHTKIH	53
HR8206A-368-420-TEV	EGR2	P11161	KPFQCRICMRNFSRS	KFARSDERKRHTKIH	53
HR8198A-273-328-TEV	EGR3	Q06889	RPHACPAEGCDRRFS	FSRSDHLTTHIRTHT	56
HR8048A-204-299-TEV	EHF	Q9NZC4	PRGTHLWEFIRDILL	VYKFGKNARGWRENE	96
HR7395A-770-879-NHT	EIF3C	Q99613	PEADKVRTMLVRKIQ	SLVENNERVFDHKQG	110
HR2095-14	EIF3K	Q9UBQ5	MAMFEQMRANVGKLL	KIDFDSVSSIMASSQ	218
HR564-14	EIF3K	Q9UBQ5	MAMFEQMRANVGKLL	KIDFDSVSSIMASSQ	218
HR6332A-198-348-15	ELF1	P32519	KKNKDGKGNTIYLWE	SPGVKGGATTVLKPG	151
HR6332A-198-353-15	ELF1	P32519	KKNKDGKGNTIYLWE	GGATTVLKPGNSKAA	156

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6332A-203-348-15	ELF1	P32519	GKGNTIYLWEFLLAL	SPGVKGGATTVLKPG	146
HR6332B-152-304-15	ELF1	P32519	ETQQVQEKYADSPGA	KEMPKDLIYINDEDP	153
HR6332B-157-299-15	ELF1	P32519	QEKYADSPGASSPEQ	LVYQFKEMPKDLIYI	143
HR6332B-157-304-15	ELF1	P32519	QEKYADSPGASSPEQ	KEMPKDLIYINDEDP	148
HR6332B-198-304-15	ELF1	P32519	KKNKDGKGNTIYLWE	KEMPKDLIYINDEDP	107
HR6332C-203-353-15	ELF1	P32519	GKGNTIYLWEFLLAL	GGATTVLKPGNSKAA	151
HR7067A-150-308-15	ELF2	Q15723	MLWEFLLDLLQDKNT	GVARVVNITSPGHDA	160
HR7067A-157-303-15	ELF2	Q15723	MLLQDKNTCPRYIKW	SRAEKGVARVVNITS	148
HR7067A-200-308-15	ELF2	Q15723	MNYETMGRALRYYYQ	GVARVVNITSPGHDA	110
HR7867A-45-132-TEV	ELF3	P78545	SNPQMSLEGTEKASW	GDQLHAQLRDLTSSS	88
HR7867B-269-371-TEV	ELF3	P78545	APRGTHLWEFIRDIL	NSSGWKEEEVLQSRN	103
HR8186A-1-104-TEV	ELF4	Q99607	AITLQPSDLIFEFAS	HTMSTAEVLLNMESP	103
HR8186A-1-87-TEV	ELF4	Q99607	AITLQPSDLIFEFAS	QILEGSFLLTDDNEA	86
HR8186A-1-94-TEV	ELF4	Q99607	AITLQPSDLIFEFAS	LLTDDNEATSHTMST	93
HR7396A-166-265-TEV	ELF5	Q9UKW6	SRTSLQSSHLWEFVR	YKFGKNAHGWQEDKL	100
HR7616A-1-93-TEV	ELK3	P41970	ESAITLWQFLLQLLL	KFVYKFVSFPEILKM	92
HR4449C-1-93-TEV	ELK4	P28324	DSAITLWQFLLQLLQ	KFVYKFVSYPEILNM	92
HR8153A-249-313-Av6HT	EN2	P19622	TAFTAEQLQRLKAEF	IKKATGNKNTLAVHL	65
HR7174A-264-457-Av6HT	EOMES	O95936	GFRAHVYLCNRPLWL	LKIDHNPFAKGFRDN	194
HR4540F-1221-1288-14	EP300	Q09472	MQPQTTINKEQFSKR	GCLKKSARTRKENKF	69
HR4540F-1226-1281-14	EP300	Q09472	MINKEQFSKRKNDTL	PAGFVCDGCLKKSAR	57
HR4540F-1236-1281-14	EP300	Q09472	MNDTLDPELFVECTE	PAGFVCDGCLKKSAR	47

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4540G-323-423-TEV	EP300	Q09472	GSGAHTADPEKRKLI	HDCPVCLPLKNAGDK	100
HR4540H-1045-1161-Av6HT	EP300	Q09472	KKKIFKPEELRQALM	EVFEQEIDPVMQSLG	117
HR4540I-1726-1817-Av6HT	EP300	Q09472	SPGDSRRLSIQRCIQ	VPFCLNIKQKLRQQQ	92
HR4540J-1135-1205-15	EP300	Q09472	MTSRVYKYCSKLSEV	YYSYQNRYHFCEKCF	72
HR4540J-1135-1220-15	EP300	Q09472	MTSRVYKYCSKLSEV	NEIQGESVSLGDDPS	87
HR4540I-1165-1205-15	EP300	Q09472	MGRKLEFSPQTLCCY	YYSYQNRYHFCEKCF	42
HR4540J-1165-1220-15	EP300	Q09472	MGRKLEFSPQTLCCY	NEIQGESVSLGDDPS	57
HR7040A-1285-1379-Av6HT	EP400	Q96L91	HVLKCRLSNRQKALY	RDFWKEADLSMFDLI	95
HR8188A-239-350-TEV	EPAS1	Q99814	LDSKTFLSRHSMDMK	CIMCVNYVLSEIEKN	112
HR6944A-8-123-Av6HT	ERF	P50548	GFAFPDWAYKPESSP	NKLVLVNYPFIDVGL	116
HR6944A-8-160-NHT	ERF	P50548	GFAFPDWAYKPESSP	PSTPSEVLSPTEDPR	153
HR6944B-24-123-Av6HT	ERF	P50548	SRQIQLWHFILELLR	NKLVLVNYPFIDVGL	100
HR6944B-24-160-Av6HT	ERF	P50548	SRQIQLWHFILELLR	PSTPSEVLSPTEDPR	137
HR4801B-180-254-TEV	ESR1	P03372	KETRYCAVCNDYASG	CRLRKCYEVGMMKGG	75
HR4685B-144-218-TEV	ESR2	Q92731	RDAHFCAVCSDYASG	CRLRKCYEVGMVKCG	75
HR7097A-77-146-15	ESRRA	P11474	MRLCLVCGDVASGYH	QACRFTKCLRVGMLK	71
HR7097A-77-146-Av6HT	ESRRA	P11474	RLCLVCGDVASGYHY	QACRFTKCLRVGMLK	70
HR7097A-77-146-TEV	ESRRA	P11474	RLCLVCGDVASGYHY	QACRFTKCLRVGMLK	70
HR7097A-77-168-15	ESRRA	P11474	MRLCLVCGDVASGYH	VRGGRQKYKRRPEVD	93
HR7097A-77-168-Av6HT	ESRRA	P11474	RLCLVCGDVASGYHY	VRGGRQKYKRRPEVD	92
HR7097A-77-168-TEV	ESRRA	P11474	RLCLVCGDVASGYHY	VRGGRQKYKRRPEVD	92
HR7097B-179-423-15	ESRRA	P11474	MGPLAVAGGPRKTAA	PMHKLFLEMLEAMMD	246

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7097B-179-423-Av6HT	ESRRA	P11474	GPLAVAGGPRKTAAP	PMHKLFLEMLEAMMD	245
HR7097B-179-423-TEV	ESRRA	P11474	GPLAVAGGPRKTAAP	PMHKLFLEMLEAMMD	245
HR7097C-193-423-15	ESRRA	P11474	MPVNALVSHLLVVEP	PMHKLFLEMLEAMMD	232
HR7097C-193-423-Av6HT	ESRRA	P11474	PVNALVSHLLVVEPE	PMHKLFLEMLEAMMD	231
HR7097C-193-423-TEV	ESRRA	P11474	PVNALVSHLLVVEPE	PMHKLFLEMLEAMMD	231
HR8438A-101-433-15	ESRRB	O95718	MRLCLVCGDIASGYH	VPMHKLFLEMLEAKA	334
HR8438A-78-435-15	ESRRB	O95718	MDCASGIMEDSAIKC	MHKLFLEMLEAKAWA	359
HR8438B-169-433-15	ESRRB	O95718	MLKEGVRLDRVRGGR	VPMHKLFLEMLEAKA	266
HR8438B-182-433-15	ESRRB	O95718	MRQKYKRRLDSESSP	VPMHKLFLEMLEAKA	253
HR8438B-203-433-15	ESRRB	O95718	MPPAKKPLTKIVSYL	VPMHKLFLEMLEAKA	232
HR7566D-122-219-Av6HT	ESRRG	P62508	MSMPKRLCLVCGDIA	GGRQKYKRRIDAENS	99
HR7566D-122-219-TEV	ESRRG	P62508	SMPKRLCLVCGDIAS	GGRQKYKRRIDAENS	98
HR6900A-130-214-NHT	ESX1	Q8N693	AEGPQPPERKRRRRT	VLMLRNTATADLAHP	85
HR8013A-320-415-Av6HT	ETS1	P14921	VIPAAALAGYTGSGP	IIHKTAGKRYVYRFV	96
HR8013A-320-415-TEV	ETS1	P14921	VIPAAALAGYTGSGP	IIHKTAGKRYVYRFV	96
HR5529-1-329-15	ETS2	P15036	MNDFGIKNMDQVAPV	EDDCSQSLCLNKPTM	329
HR5529A-14	ETS2	P15036	MHDSANCELPLLTPC	EHLEQMIKENQEKTE	116
HR5529A-15	ETS2	P15036	MHDSANCELPLLTPC	EHLEQMIKENQEKTE	116
HR8505A-240-333-Av6HT	ETV2	O00321	IQLWQFLLELLHDGA	FGGRVPSLAYPDCAG	94
HR7364A-15-174-NHT	ETV3	P41162	GGYQFPDWAYKTESS	PTNDVQPGRFSASSL	160
HR6967A-1-136-NHT	ETV3L	Q6ZN32	HCSCLAEGIPANPGN	SKLIVVNYPLWEVRA	135
HR5533A-14	ETV4	P43268	MREGPPYQRRGALQL	QRPALKAEFDRPVSE	122

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR5533A-15	ETV4	P43268	MREGPPYQRRGALQL	QRPALKAEFDRPVSE	122
HR8084A-311-445-Av6HT	ETV4	P43268	CVVPEKFEGDIKQEG	AFPDNQRPALKAEFD	135
HR7423A-333-470-NHT	ETV5	P41161	LYFDDTCVVPERLEG	SMAFPDNQRPFLKAE	138
HR6884A-338-443-TEV	ETV6	P41212	CRLLWDYVYQLLSDS	GRTDRLEHLESQELD	106
HR6884B-47-129-TEV	ETV6	P41212	SIRLPAHLRLQPIYW	ELLQHILKQRKPRIL	83
HR7437A-8-133-NHT	ETV7	Q9Y603	ISPISPVAAMPPLGT	ALVCGPFFGGIFRLK	126
HR7509A-183-242-TEV	EVX1	P49640	RRYRTAFTREQIARL	KVWFQNRRMKDKRQR	59
HR7284A-188-247-TEV	EVX2	Q03828	VRRYRTAFTREQIAR	KVWFQNRRMKDKRQR	60
HR7802A-349-453-TEV	EWSR1	Q01844	PPVDPDEDSDNSAIY	LKVSLARKKPPMNSM	105
HR7511A-1-99-TEV	EXOC2	Q96KP1	SRSRQPPLVTGISPN	TSTVSFKLLKPEKIG	98
HR6516A-463-729-NHT	EZH1	Q92800	KTCKQVFQFAVKESL	RAIQAGEELFFDYRY	267
HR6323-214-746-14	EZH2	Q15910	PPRKFPSDKIFEAIS	DALKYVGIEREMEIP	533
HR7273-1-589-TEV	FARSB	Q9NSD9	PTVSVKRDLLFQALG	TMPCSSLEINVGPFL	588
HR8271C-2054-2125-NHT	FBN1	P35555	QDLRMSYCYAKFEGG	CPYGSGIIVGPDDSA	72
HR6868A-99-166-NHT	FERD3L	Q96RJ6	TYAQRQAANIRERKR	FMTELLESCEKKESG	68
HR6882A-43-139-TEV	FEV	Q99581	GSGQIQLWQFLLELL	RFDFQGLAQACQPPP	97
HR6968A-258-310-NHT	FEZF1	A0PJY2	KVFTCEVCGKVFNAH	GFRQASTLCRHKIIH	53
HR7661A-275-327-NHT	FEZF2	Q8TBJ5	KNFTCEVCGKVFNAH	GFRQASTLCRHKIIH	53
HR3605C-806-930-14	FGD1	P98174	MRRRSILEKQASVAA	LGRAGRGDTFCPGPT	126
HR3605C-811-925-14	FGD1	P98174	MLEKQASVAAENSVI	RWMAVLGRAGRGDTF	116
HR8434A-55-150-Av6HT	FIGLA	Q6QHK4	SSTENLQLVLERRRV	SYSNNSSESHTSSAR	96
HR8078A-77-129-Av6HT	FIZ1	Q96SL8	RPYRCSACPKGFRDS	RFSSRSSLGRHLKRQ	53

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4739B-114-198-Av6HT	FLI1	Q01543	MPPNMTTNERRVIVP	TEVLLSHLSYLRESS	86
HR4739B-114-198-TEV	FLI1	Q01543	PPNMTTNERRVIVPA	TEVLLSHLSYLRESS	85
HR6395-41-355-15	FOS	P01100	MGSPVNAQDFCTDLA	FVFTYPEADSFPSCA	316
HR6395-41-361-15	FOS	P01100	MGSPVNAQDFCTDLA	EADSFPSCAAAHRKG	322
HR6395-46-350-15	FOS	P01100	MAQDFCTDLAVSSAN	AYTSSFVFTYPEADS	306
HR6395-46-361-15	FOS	P01100	MAQDFCTDLAVSSAN	EADSFPSCAAAHRKG	317
HR3160-41-293-15	FOSL2	P15408	MPGSGSAFIPTINAI	NLVFTYPSVLEQESP	254
HR3160-44-288-15	FOSL2	P15408	MGSAFIPTINAITTS	TPGTSNLVFTYPSVL	246
HR7662A-167-264-NHT	FOXA1	P55317	PHAKPPYSYISLITM	SGNMFENGCYLRRQK	98
HR7840A-114-211-NHT	FOXA3	P55318	AHAKPPYSYISLITM	SGNMFENGCYLRRQK	98
HR7656A-11-100-NHT	FOXB1	Q99853	DQKPPYSYISLTAMA	FWALHPSCGDMFENG	90
HR7565A-15-100-NHT	FOXB2	Q5VYV0	PYSYISLTAMAIQHS	FWALHPDCGDMFENG	86
HR8399A-76-168-TEV	FOXC1	Q12948	VKPPYSYIALITMAI	LDPDSYNMFENGSFL	92
HR6945A-70-162-TEV	FOXC2	Q99958	LVKPPYSYIALITMA	LDPDSYNMFENGSFL	93
HR8366A-126-222-NHT	FOXD2	O60548	VKPPYSYIALITMAI	ADMFDNGSFLRRRKR	97
HR8366A-126-222-TEV	FOXD2	O60548	VKPPYSYIALITMAI	ADMFDNGSFLRRRKR	97
HR7150A-140-236-Av6HT	FOXD3	Q9UJU5	VKPPYSYIALITMAI	EDMFDNGSFLRRRKR	97
HR7150A-140-236-TEV	FOXD3	Q9UJU5	VKPPYSYIALITMAI	EDMFDNGSFLRRRKR	97
HR7841A-104-204-TEV	FOXD4	Q12950	KPPSSYIALITMAIL	NGSFLRRRKRFQRHQ	101
HR6889A-107-207-TEV	FOXD4L1	Q9NU39	KPPYSYIALITMAIL	NGSFLRRRKRFKRHQ	101
HR8496-108-208-Av6HT	FOXD4L3	Q6VB84	KPPYSYIALITMAIL	NGSFLRRRKRFKRHQ	101
HR7982-108-208-Av6HT	FOXD4L4	Q6VB85	KPPYSYIALITMAIL	NGSFLRRRKRFKRHQ	101

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7029A-108-208-TEV	FOXD4L5	Q5VV16	KPPYSYIALITMAIL	NGSFLRRRKRFKRHQ	101
HR7874-108-208-Av6HT	FOXD4L6	Q3SYB3	KPPYSYIALITMAIL	NGSFLRRRKRFKRHQ	101
HR5544A-14	FOXE1	O00358	MAGAGVPGEATGRGA	FLRRRKRFKRSDLST	131
HR5544A-15	FOXE1	O00358	MAGAGVPGEATGRGA	FLRRRKRFKRSDLST	131
HR6991A-51-146-NHT	FOXE1	O00358	RGKPPYSYIALIAMA	NAEDMFESGSFLRRR	96
HR7179A-69-165-NHT	FOXE3	Q13461	RGKPPYSYIALIAMA	AADMFDNGSFLRRRK	97
HR8233A-48-138-15	FOXF1	Q12946	MKPPYSYIALIVMAI	IDPASEFMFEEGSFR	92
HR8233A-48-138-Av6HT	FOXF1	Q12946	KPPYSYIALIVMAIQ	IDPASEFMFEEGSFR	91
HR7975A-101-190-Av6HT	FOXF2	Q12947	PPYSYIALIVMAIQS	IDPASEFMFEEGSFR	90
HR4505B-182-298-14	FOXG1	P55316	PPFSYNALIMMAIRQ	LAFKRGARLTSTGLT	117
HR4505B-183-294-14	FOXG1	P55316	PFSYNALIMMAIRQS	SRAKLAFKRGARLTS	112
HR4505C-183-276-Av6HT	FOXG1	P55316	PFSYNALIMMAIRQS	DDVFIGGTTGKLRRR	94
HR5526A-14	FOXH1	O75593	MYLRHDKPPYTYLAM	RLQNTALCRRWQNGG	109
HR5526A-15	FOXH1	O75593	MYLRHDKPPYTYLAM	RLQNTALCRRWQNGG	109
HR8014A-138-239-Av6HT	FOXI3	A8MTJ6	EDLMKMVRPPYSYSA	CEKMFDNGNFRRKRK	102
HR6903A-121-211-NHT	FOXJ1	O92949	KPPYSYATLICMAMQ	IDPQYAERLLSGAFK	91
HR8000A-64-153-Av6HT	FOXJ2	Q9P0K8	DGKPRYSYATLITYA	YWTIDTCPDISRKRR	90
HR7453A-82-173-NHT	FOXJ3	Q9UPW0	SYASLITFAINSSPK	KEDVLPTRPKKRARS	92
HR7148A-303-403-TEV	FOXK1	P85037	ESKPPFSYAQLIVQA	LVEQAFRKRRQRGVS	101
HR8426A-256-353-Av6HT	FOXK2	Q01167	MDSKPPYSYAQLIVQ	ESKLIEQAFRKRRPR	99
HR8426A-256-353-TEV	FOXK2	Q01167	DSKPPYSYAQLIVQA	ESKLIEQAFRKRRPR	98
HR8426B-34-133-Av6HT	FOXK2	Q01167	GWAVARLEGREFEYL	NGVFVDGVFQRRGAP	100

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8426B-34-153-Av6HT	FOXK2	Q01167	GWAVARLEGREFEYL	RVCTFRFPSTNIKIT	120
HR8426B-58-139-TEV	FOXK2	Q01167	RNSSQGSVDVSMGHS	GVFQRRGAPPLQLPR	82
HR8426B-58-143-TEV	FOXK2	Q01167	RNSSQGSVDVSMGHS	RRGAPPLQLPRVCTF	86
HR8426B-63-133-TEV	FOXK2	Q01167	GSVDVSMGHSSFISR	NGVFVDGVFQRRGAP	71
HR8426B-63-139-TEV	FOXK2	Q01167	GSVDVSMGHSSFISR	GVFQRRGAPPLQLPR	77
HR8426B-63-153-Av6HT	FOXK2	Q01167	GSVDVSMGHSSFISR	RVCTFRFPSTNIKIT	91
HR8426B-70-133-Av6HT	FOXK2	Q01167	GHSSFISRRHLEIFT	NGVFVDGVFQRRGAP	64
HR8426B-70-153-Av6HT	FOXK2	Q01167	GHSSFISRRHLEIFT	RVCTFRFPSTNIKIT	84
HR7608A-10-139-Av6HT	FOXL1	Q12952	PALAASPMLYLYGPE	LDPRCLDMFENGNYR	130
HR7608A-43-139-15	FOXL1	Q12952	MRAETPQKPPYSYIA	LDPRCLDMFENGNYR	98
HR7608A-48-111-15	FOXL1	Q12952	MQKPPYSYIALIAMA	IRHNLSLNDCFVKVP	65
HR7608A-48-139-15	FOXL1	Q12952	MQKPPYSYIALIAMA	LDPRCLDMFENGNYR	93
HR7608B-10-134-Av6HT	FOXL1	Q12952	PALAASPMLYLYGPE	GSYWTLDPRCLDMFE	125
HR7608B-50-164-Av6HT	FOXL1	Q12952	PPYSYIALIAMAIQD	GAPEAKRPRAETHQR	115
HR7161A-56-143-NHT	FOXL2	P58012	PYSYVALIAMAIRES	TLDPACEDMFEKGNY	88
HR6909A-222-360-Av6HT	FOXM1	Q08050	MPSRPSASWQNSVSE	NPELRRNMTIKTELP	140
HR6909A-222-360-TEV	FOXM1	Q08050	PSRPSASWQNSVSER	NPELRRNMTIKTELP	139
HR7300A-268-368-NHT	FOXN1	O15353	LFPKPIYSYSILIFM	DKMQEELQKWKRKDP	101
HR7988A-110-208-Av6HT	FOXN2	P32314	TSKPPYSFSLLIYMA	KPNLIQALKKQPFSS	99
HR6979A-111-207-NHT	FOXN3	O00409	PNCKPPYSFSCLIFM	PEYRQNLIQALKKTP	97
HR7465A-193-285-NHT	FOXN4	Q96NZ1	KPIYSYSCLIAMALK	NLARIDKMEEEMHKW	93
HR4552B-151-249-TEV	FOXO1	Q12778	KSSSSRRNAWGNLSY	SWWMLNPEGGKSGKS	99

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR5548A-14	FOXO1	Q12778	MPPAAAGPLAGQPRK	SKFAKSRSRAAKKKA	139
HR5548A-15	FOXO1	Q12778	MPPAAAGPLAGQPRK	SKFAKSRSRAAKKKA	139
HR5549A-14	FOXO3	O43524	MLPPPQPGAAGGSGQ	NKYTKSRGRAAKKKA	141
HR5549A-15	FOXO3	O43524	MLPPPQPGAAGGSGQ	NKYTKSRGRAAKKKA	141
HR4610C-102-197-TEV	FOXO4	P98177	GNQSYAELISQAIES	EGGKSGKAPRRRAAS	96
HR7590A-462-548-TEV	FOXP1	Q9H334	AEVRPPFTYASLIRQ	WTVDEVEFQKRRPQK	87
HR7934A-501-587-TEV	FOXP2	O15409	DVRPPFTYATLIRQA	TVDEVEYQKRRSQKI	87
HR6897A-464-550-TEV	FOXP4	Q8IVH2	ADVRPPFTYASLIRQ	WTVDEREYQKRRPPK	87
HR8323A-118-217-NHT	FOXQ1	Q9C009	PKPPYSYIALIAMAI	TFADGVFRRRRKRLS	100
HR8323A-118-217-TEV	FOXQ1	Q9C009	PKPPYSYIALIAMAI	TFADGVFRRRRKRLS	100
HR7058A-170-271-NHT	FOXR1	Q6PIV2	LWSRPPLNYFHLIAL	GHRRFAEEARALAST	102
HR8252A-189-311-Av6HT	FOXR2	Q6PIQ5	SWQRPPLNCSHLIAL	ECMSQPELLTSLFDL	123
HR7804A-20-110-NHT	FOXS1	O43638	PYSYIALIAMAIQSS	PDCHDMFEHGSFLRR	91
HR8359A-35-121-TEV	GABPA	Q06546	AECVSQAIDINEPIG	KLNILEIVKPADTVE	87
HR7128A-251-311-TEV	GATA1	P15976	SKRAGTQCTNCQTTT	MRKDGIQTRNRKASG	61
HR4414D-340-402-TEV	GATA2	P23769	SAARRAGTCCANCQT	MKKEGIQTRNRKMSN	63
HR7641A-308-370-TEV	GATA3	P23771	LSAARRAGTSCANCQ	TMKKEGIQTRNRKMS	63
HR4783B-262-321-TEV	GATA4	P43694	SASRRVGLSCANCQT	PLAMRKEGIQTRKRK	60
HR8231A-242-324-15	GBX1	Q14549	MTGAEEGAPVTAGVT	QNRRAKWKRIKAGNV	84
HR8231A-242-324-Av6HT	GBX1	Q14549	TGAEEGAPVTAGVTA	QNRRAKWKRIKAGNV	83
HR6959A-1-173-TEV	GCM1	Q9NP62	EPDDFDSEDKEILSW	TKLEAEARRAMKKVN	172
HR8430A-6-165-Av6HT	GCM2	O75603	VQEAVGVCSYGMQLS	FQAKGVHDHPRPESK	160

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4429D-233-303-14	GFI1	Q99684	MKGAGVKVESELLCT	CGKTFGHAVSLEQHK	72
HR4429D-233-315-14	GFI1	Q99684	MKGAGVKVESELLCT	QHKAVHSQERSFDCK	84
HR4429D-238-298-14	GFI1	Q99684	MKVESELLCTRLLLG	FACEMCGKTFGHAVS	62
HR4429D-238-310-14	GFI1	Q99684	MKVESELLCTRLLLG	AVSLEQHKAVHSQER	74
HR4429E-311-392-TEV	GFI1	Q99684	SFDCKICGKSFKRSS	SQSSNLITHSRKHTG	82
HR7937A-234-388-TEV	GLI1	P08151	YVCKLPGCTKRYTDP	RLDQLHQLRPIGTRG	155
HR7924A-436-590-TEV	GLI2	P10070	ETNCHWEDCTKEYDT	TDPSSLRKHVKTVHG	155
HR7118A-479-633-TEV	GLI3	P10071	ETNCHWEGCAREFDT	TDPSSLRKHVKTVHG	155
HR7155A-189-350-NHT	GLIS1	Q8NBF1	RVVAGRQACRWVDCC	PSSLRKHVKAHSAKE	162
HR7416A-116-298-Av6HT	GLIS2	Q9BZE0	DFQPLRYLDGVPSSF	TRTHYVDKPYYCKMP	183
HR7416A-116-318-Av6HT	GLIS2	Q9BZE0	DFQPLRYLDGVPSSF	YTDPSSLRKHIKAHG	203
HR7416A-150-318-Av6HT	GLIS2	Q9BZE0	LTPPKDKCLSPDLPL	YTDPSSLRKHIKAHG	169
HR7416A-163-298-Av6HT	GLIS2	Q9BZE0	PLPKQLVCRWAKCNQ	TRTHYVDKPYYCKMP	136
HR7416A-163-318-Av6HT	GLIS2	Q9BZE0	PLPKQLVCRWAKCNQ	YTDPSSLRKHIKAHG	156
HR7200A-261-553-TEV	GLYR1	Q49A26	GSITPTDKKIGFLGL	QSDNDMSAVYRAYIH	293
HR7763A-87-182-TEV	GMEB1	Q9Y692	ANEDMEIAYPITCGE	YQHDKVCSNTCRSTK	96
HR7418A-64-203-NHT	GMEB2	Q9UKD1	AFTASSQLKEAVLVK	LSSPTSAEYIPLTPA	140
HR7418A-87-176-Av6HT	GMEB2	Q9UKD1	EAEIVYPITCGDSRA	LDFYQHDKVCSNTCR	90
HR7418A-87-203-Av6HT	GMEB2	Q9UKD1	EAEIVYPITCGDSRA	LSSPTSAEYIPLTPA	117
HR7418B-64-179-Av6HT	GMEB2	Q9UKD1	AFTASSQLKEAVLVK	YQHDKVCSNTCRSTK	116
HR7418B-83-179-Av6HT	GMEB2	Q9UKD1	GENLEAEIVYPITCG	YQHDKVCSNTCRSTK	97
HR7418B-83-203-Av6HT	GMEB2	Q9UKD1	GENLEAEIVYPITCG	LSSPTSAEYIPLTPA	121

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8221A-2125-2211-NHT	GON4L	Q3T8J9	PEGEQQPKAAEATVC	RELMQLFHTACEASS	87
HR7528A-714-848-NHT	GPR155	Q7Z3F1	DKHLIILPFKRRLEF	LQKSPEQSPPAINAN	135
HR7997A-174-442-Av6HT	GRHL1	Q9NZI5	VYHPEPTERVVVFDR	KIRDEERKQSKRKVS	269
HR7758A-219-444-Av6HT	GRHL2	Q6ISB3	SFKDAATEKFRSASV	RKQNRKKGKGQASQT	226
HR7267A-161-223-TEV	GSC	P56915	RRHRTIFTDEQLEAL	KNRRAKWRRQKRSSS	63
HR8103A-123-177-Av6HT	GSC2	O15499	QRRTRRHRTIFSEEQ	IRLREERVEVWFKNR	55
HR7705A-139-207-NHT	GSX1	Q9H4S2	SSSNQLPSSKRMRTA	IWFQNRRVKHKKEGK	69
HR8308A-66-146-TEV	GTF2E2	P29084	ALSGSSGYKFGVLAK	VIDGKYAFKPKYNVR	81
HR8128A-449-517-Av6HT	GTF2F1	P35269	SGDVQVTEDAVRRYL	ERKMINDKMHFSLKE	69
HR8128A-449-517-TEV	GTF2F1	P35269	SGDVQVTEDAVRRYL	ERKMINDKMHFSLKE	69
HR7967A-175-243-TEV	GTF2F2	P13984	RARADKQHVLDMLFS	HKNTWELKPEYRHYQ	69
HR7205-1-238-15	GTF2H2C	Q6P1K8	MDEEPERTKRWEGGY	DESHYKELLTHHLSP	238
HR7205-1-327-15	GTF2H2C	Q6P1K8	MDEEPERTKRWEGGY	VSAPHLARSYHHLFP	327
HR7205-1-332-15	GTF2H2C	Q6P1K8	MDEEPERTKRWEGGY	LARSYHHLFPLDAFQ	332
HR7205-10-327-15	GTF2H2C	Q6P1K8	MRWEGGYERTWEILK	VSAPHLARSYHHLFP	319
HR7205-10-395-15	GTF2H2C	Q6P1K8	MRWEGGYERTWEILK	CCPGCIHKIPAPSGV	387
HR7205-15	GTF2H2C	Q6P1K8	MDEEPERTKRWEGGY	CPGCIHKIPAPSGV*	396
HR7205A-328-386-TEV	GTF2H2C	Q6P1K8	LDAFQEIPLEEYNGE	DVFVHDSLHCCPGGI	59
HR7205B-10-216-15	GTF2H2C	Q6P1K8	MRWEGGYERTWEILK	SAEVRVCTVLARETG	208
HR7205B-48-220-15	GTF2H2C	Q6P1K8	MEHHGQVRLGMMRHL	RVCTVLARETGGTYH	174
HR7205B-48-238-15	GTF2H2C	Q6P1K8	MEHHGQVRLGMMRHL	DESHYKELLTHHLSP	192
HR7205B-53-216-15	GTF2H2C	Q6P1K8	MVRLGMMRHLYVVVD	SAEVRVCTVLARETG	165

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7205B-53-236-15	GTF2H2C	Q6P1K8	MVRLGMMRHLYVVVD	ILDESHYKELLTHHL	185
HR7205C-1-216-TEV	GTF2H2C	Q6P1K8	DEEPERTKRWEGGYE	SAEVRVCTVLARETG	215
HR7205C-1-255-TEV	GTF2H2C	Q6P1K8	DEEPERTKRWEGGYE	ASSSSECSLIRMGFP	254
HR7205C-10-236-TEV	GTF2H2C	Q6P1K8	RWEGGYERTWEILKE	ILDESHYKELLTHHL	227
HR7205C-10-255-TEV	GTF2H2C	Q6P1K8	RWEGGYERTWEILKE	ASSSSECSLIRMGFP	246
HR7820A-107-194-TEV	GTF2IRD2	Q86UP8	LRKAVEDYFCFCYGK	NRPFLGPESQLGGPG	88
HR7355A-318-415-TEV	GTF2IRD2B	Q6EKJ0	NEKERLSSIEKIKQL	KFTVIRPLPGLELSN	98
HR7357A-130-188-NHT	GTF3A	Q92664	KQYESFEDCKKTFK	ASPSKLKRHAKAHEG	59
HR7579A-1-143-NHT	GZF1	Q9H116	ESGAVLLESKSSPFN	KKQMLESVLLELQNF	142
HR7057A-44-107-Av6HT	H1FX	Q92522	QPGKYSQLVVETIRR	IKALVQNDTLLQVKG	64
HR7057A-44-123-Av6HT	H1FX	Q92522	QPGKYSQLVVETIRR	GANGSFKLNRKKLEG	80
HR7057A-61-123-Av6HT	H1FX	Q92522	ERNGSSLAKIYTEAK	GANGSFKLNRKKLEG	63
HR4599B-103-162-14	HAND2	P61296	MTANRKERRRTQSIN	IAYLMDLLAKDDQNG	61
HR4798B-422-503-TEV	HBP1	O60381	SSGTVSATSPNKCKR	ALAEEQKRLNPDCWK	82
HR7788A-435-507-TEV	HDX	Q7Z353	KYRLMGIEVPPPRGG	SSQEEPNEVVPNDAR	73
HR7299A-109-148-Av6HT	HES1	Q14469	KYRAGFSECMNEVTR	EVRTRLLGHLANCMT	40
HR7299A-109-153-Av6HT	HES1	Q14469	KYRAGFSECMNEVTR	LLGHLANCMTQINAM	45
HR7306A-7-75-NHT	HES2	Q9Y543	AGDAAELRKSLKPLL	EMTVRFLQELPASSW	69
HR8387-1-122-Av6HT	HES3	Q5TGS1	EKKRRARINVSLEQL	GLGQEAPALFRPCTP	121
HR6986-1-108-TEV	HES5	Q5TA89	APSTVAVELLSPKEK	WCLQEAVQFLTLHAA	107
HR6986-1-122-TEV	HES5	Q5TA89	APSTVAVELLSPKEK	ASDTQMKLLYHFQRP	121
HR6986-49-166-TEV	HES5	Q5TA89	RHQPNSKLEKADILE	AAAAHQPACGLWRPW	118

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6986A-11-108-Av6HT	HES5	Q5TA89	LSPKEKNRLRKPVVE	WCLQEAVQFLTLHAA	98
HR6986A-11-80-Av6HT	HES5	Q5TA89	LSPKEKNRLRKPVVE	VSYLKHSKAFVAAAG	70
HR6986A-21-108-Av6HT	HES5	Q5TA89	KPVVEKMRRDRINSS	WCLQEAVQFLTLHAA	88
HR6986A-25-108-Av6HT	HES5	Q5TA89	EKMRRDRINSSIEQL	WCLQEAVQFLTLHAA	84
HR6872A-108-174-TEV	HESX1	Q9UBX0	GRRPRTAFTQNQIEV	RAKLKRSHRESQFLM	67
HR7863A-111-167-TEV	HEY1	Q9Y5J3	AGGKGYFDAHALAMD	PLRVRLVSHLNNYAS	57
HR7070A-110-166-TEV	HEY2	Q9UBP5	GYFDAHALAMDFMSI	RLVSHLSTCATQREA	57
HR7572A-104-158-15	HEYL	Q9NQ87	MTGFFDARALAVDFR	PVRIRLLSHLNSYAA	55
HR7572A-77-163-15	HEYL	Q9NQ87	MQGSSKLEKAEVLQM	LLSHLNSYAAEMEPS	88
HR7572A-82-158-15	HEYL	Q9NQ87	MLEKAEVLQMTVDHL	PVRIRLLSHLNSYAA	78
HR7851A-138-194-Av6HT	HHEX	Q03014	MKGGQVRFSNDQTIE	QVKTWFQNRRAKWRR	58
HR7851A-138-194-TEV	HHEX	Q03014	KGGQVRFSNDQTIEL	QVKTWFQNRRAKWRR	57
HR7402A-1-153-NHT	HIC1	Q14526	TFPEADILLKSGECA	PDLVALCKKRLKRHG	152
HR7195A-20-139-NHT	HIC2	Q96JB3	GPDMELPSHSKQLLL	YLQLPELAALCRRKL	120
HR3603B-775-826-TEV	HIF1A	Q16665	PSDLACRLLGQSMDE	LLQGEELLRALDQVN	52
HR7384A-39-112-TEV	HIST1H1A	Q02539	AGPSVSELIVQAASS	QTKGTGASGSFKLNK	74
HR7165A-40-112-TEV	HIST1H1B	P16401	GPPVSELITKAVAAS	QTKGTGASGSFKLNK	73
HR7583A-36-109-TEV	HIST1H1C	P16403	SGPPVSELITKAVAA	QTKGTGASGSFKLNK	74
HR7583A-37-110-TEV	HIST1H1C	P16403	GPPVSELITKAVAAS	TKGTGASGSFKLNKK	74
HR8248A-2087-2143-TEV	HIVEP1	P15822	KYICEECGIRCKKPS	GNLTKHMKSKAHSKK	57
HR7166A-1798-1854-TEV	HIVEP2	P31629	KYICEECGIRCKKPS	GNLTKHMKSKAHMKK	57
HR7042A-1753-1809-TEV	HIVEP3	Q5T1R4	KYVCEECGIRCKKPS	GNLTKHMKSKAHSKK	57

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7786A-523-577-NHT	HKR1	P10072	KPFVCAECGRGFNDK	RQKPNLFRHKRAHSG	55
HR7711A-33-224-TEV	HLA-DQB1	P01920	RDSPEDFVFQFKGMC	PSLQSPITVEWRAQS	192
HR7053A-30-228-TEV	HLA-DRB1	P01911	GDTRPRFLWQPKREC	TVEWRARSESAQSKM	199
HR8520A-30-221-TEV	HLA-DRB1	P04229	GDTRPRFLWQLKFEC	PSVTSPLTVEWRARS	192
HR7721A-30-219-TEV	HLA-DRB3	P79483	GDTRPRFLELRKSEC	EHPSVTSALTVEWRA	190
HR7380A-30-221-TEV	HLA-DRB5	Q30154	GDTRPRFLQQDKYEC	PSVTSPLTVEWRAQS	192
HR7352A-219-295-TEV	HLF	Q16534	IPDDLKDDKYWARRR	CKNILAKYEARHGPL	77
HR8006A-269-334-Av6HT	HLX	Q14774	PQTYKRKRSWSRAVF	VKVWFQNRRMKWRHS	66
HR7519A-268-352-TEV	HMBOX1	Q6NT76	RGSRFTWRKECLAVM	KRRANIEAAILESHG	85
HR1506-15	HMG20A	Q9NP66	MENLMTSSTLPPLFA	SSNAAEGNEQRHEDE	79
HR1506-15.2wt	HMG20A	Q9NP66	MENLMTSSTLPPLFA	SSNAAEGNEQRHEDE	79
HR7093A-68-149-TEV	HMG20B	Q9P0W2	NGPKAPVTGYVRFLN	RAYQQSEAYKMCTEK	82
HR7828-30	HMGB1	P09429	MGKGDPKKPRGKMSS	EDEEDEDEEEDDDDE	215
HR7828A-8-78-Av6HT	HMGB1	P09429	KPRGKMSSYAFFVQT	AKADKARYEREMKTY	71
HR7828A-8-78-TEV	HMGB1	P09429	KPRGKMSSYAFFVQT	AKADKARYEREMKTY	71
HR7828B-30	HMGB1	P09429	KKKFKDPNAPKRPPS	LKEKYEKDIAAYRAK	80
HR8516A-8-78-TEV	HMGB1P1	B2RPK0	KPRGKMSSYAFFVQT	AKADKTHYERQMKTY	71
HR8015A-1-77-Av6HT	HMGB2	P26583	MGKGDPNKPRGKMSS	MAKSDKARYDREMKN	77
HR8015A-1-77-TEV	HMGB2	P26583	GKGDPNKPRGKMSSY	MAKSDKARYDREMKN	76
HR8319A-1-79-TEV	HMGB3	O15347	AKGDPKKPKGKMSAY	KADKVRYDREMKDYG	78
HR8540A-11-186-Av6HT	HMGB4	Q8WW32	ANVSSYVHFLLNYRN	MSARNRCRGKRVRQS	176
HR7956-381-466-Av6HT	HMGXB4	Q9UGU5	LHTDGHSEKKKKKEE	DKLIWKQKAQYLQHK	86

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7411A-1-264-TEV	HMOX2	P30519	SAEVETSEGVDESEK	EDGFPVHDGKGDMRK	263
HR7000A-188-261-NHT	HMX1	Q9NR08	AAGETRGGVGVGGGR	VKIWFQNRRNKWKRQ	74
HR8029A-131-216-Av6HT	HMX2	A2RU54	PGSERPRDGGAERQA	NKWKRQLSAELEAAN	86
HR6871A-218-296-NHT	HMX3	A6NHT5	SPEKKPACRKKKTRT	WKRQLAAELEAANLS	79
HR8251A-233-325-NHT	HNF1B	P35680	RNRFKWGPASQQILY	QKLAMDAYSSNQTHS	93
HR8251A-233-325-TEV	HNF1B	P35680	RNRFKWGPASQQILY	QKLAMDAYSSNQTHS	93
HR7522A-142-391-15	HNF4A	P41235	MSSYEDSSLPSINAL	GSPSDAPHAHHPLHP	251
HR7522A-142-391-Av6HT	HNF4A	P41235	SSYEDSSLPSINALL	GSPSDAPHAHHPLHP	250
HR7522A-142-391-TEV	HNF4A	P41235	SSYEDSSLPSINALL	GSPSDAPHAHHPLHP	250
HR7522B-142-378-15	HNF4A	P41235	MSSYEDSSLPSINAL	AKIDNLLQEMLLGGS	238
HR7522B-142-378-Av6HT	HNF4A	P41235	SSYEDSSLPSINALL	AKIDNLLQEMLLGGS	237
HR7522B-142-378-TEV	HNF4A	P41235	SSYEDSSLPSINALL	AKIDNLLQEMLLGGS	237
HR7522C-148-377-15	HNF4A	P41235	MSLPSINALLQAEVL	MAKIDNLLQEMLLGG	231
HR7522C-148-377-Av6HT	HNF4A	P41235	SLPSINALLQAEVLS	MAKIDNLLQEMLLGG	230
HR7522C-148-377-TEV	HNF4A	P41235	SLPSINALLQAEVLS	MAKIDNLLQEMLLGG	230
HR7522D-58-135-15	HNF4A	P41235	MALCAICGDRATGKH	FRAGMKKEAVQNERD	79
HR7522D-58-135-Av6HT	HNF4A	P41235	ALCAICGDRATGKHY	FRAGMKKEAVQNERD	78
HR7522D-58-135-TEV	HNF4A	P41235	ALCAICGDRATGKHY	FRAGMKKEAVQNERD	78
HR7469A-9-77-15	HNF4G	Q14541	MVLDPTYTTLEFETM	ASSCDGCKGFFRRSI	70
HR7469A-9-91-15	HNF4G	Q14541	MVLDPTYTTLEFETM	IRKSHVYSCRFSRQC	84
HR7469A-9-91-Av6HT	HNF4G	Q14541	VLDPTYTTLEFETMQ	IRKSHVYSCRFSRQC	83
HR7469A-9-91-TEV	HNF4G	Q14541	VLDPTYTTLEFETMQ	IRKSHVYSCRFSRQC	83

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7469A-9-95-15	HNF4G	Q14541	MVLDPTYTTLEFETM	HVYSCRFSRQCVVDK	88
HR7469A-9-95-Av6HT	HNF4G	Q14541	VLDPTYTTLEFETMQ	HVYSCRFSRQCVVDK	87
HR7469A-9-95-TEV	HNF4G	Q14541	VLDPTYTTLEFETMQ	HVYSCRFSRQCVVDK	87
HR7469B-103-328-15	HNF4G	Q14541	MYCRLRKCFRAGMKK	RQYDSRGRFGELLLL	227
HR7469B-103-328-Av6HT	HNF4G	Q14541	YCRLRKCFRAGMKKE	RQYDSRGRFGELLLL	226
HR7469B-103-328-TEV	HNF4G	Q14541	YCRLRKCFRAGMKKE	RQYDSRGRFGELLLL	226
HR8063A-429-485-TEV	HOMEZ	Q8IX15	SFQDPAIPTPPPSTR	AAHQQLRETDIPQLS	57
HR6881-1-73-TEV	HOPX	Q9BPY8	SAETASGPTEDQVEI	RRSEGLPSECRSVTD	72
HR7310A-197-291-TEV	HOXA1	P49639	ETSSPAQTFDWMKVK	FQNRRMKQKKREKEG	95
HR4742B-299-369-14	HOXA10	P31260	MKDSLGNSKGENAAN	SVHLTDRQVKIWFQN	72
HR4742B-299-393-14	HOXA10	P31260	MKDSLGNSKGENAAN	RENRIRELTANFNFS	96
HR4742C-309-369-14	HOXA10	P31260	MNAANWLTAKSGRKK	SVHLTDRQVKIWFQN	62
HR4742C-314-393-14	HOXA10	P31260	MLTAKSGRKKRCPYT	RENRIRELTANFNFS	81
HR4742C-320-393-14	HOXA10	P31260	MRKKRCPYTKHQTLE	RENRIRELTANFNFS	75
HR8427A-342-397-Av6HT	HOXA10	P31260	PYTKHQTLELEKEFL	WFQNRRMKLKKMNRE	56
HR8104A-227-302-Av6HT	HOXA11	P31270	GHTEDKAGGSSGQRT	WFQNRRMKEKKINRD	76
HR8104A-227-313-Av6HT	HOXA11	P31270	GHTEDKAGGSSGQRT	INRDRLQYYSANPLL	87
HR8104B-227-296-Av6HT	HOXA11	P31270	GHTEDKAGGSSGQRT	DRQVKIWFQNRRMKE	70
HR7749A-317-379-TEV	HOXA13	P31271	SSYRRGRKKRVPYTK	QVTIWFQNRRVKEKK	63
HR7478A-131-205-NHT	HOXA2	O43364	ESLEADGSGGGSRR	FQNRRMKHKRQTQCK	75
HR7187A-190-266-Av6HT	HOXA3	O43365	SSKRARTAYTSAQLV	GKGMLTSSGGQSPSR	77
HR7193A-222-275-TEV	HOXA4	Q00056	YTRQQVLELEKEFHF	IWFQNRRMKWKKDHK	54

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7149A-201-257-TEV	HOXA5	P20719	AYTRYQTLELEKEFH	FQNRRMKWKKDNKLK	57
HR7149A-202-258-TEV	HOXA5	P20719	YTRYQTLELEKEFHF	QNRRMKWKKDNKLKS	57
HR7925A-156-215-TEV	HOXA6	P31267	RRGRQTYTRYQTLEL	IWFQNRRMKWKKENK	60
HR4674B-194-270-TEV	HOXA9	P31269	NNPAANWLHARSTRK	NRRMKMKKINKDRAK	77
HR8367A-171-266-TEV	HOXB1	P14653	EPNTPTARTFDWMKV	QNRRMKQKKREREEG	96
HR8236A-216-273-TEV	HOXB13	Q92826	GRKKRIPYSKGQLRE	QITIWFQNRRVKEKK	58
HR7791A-149-216-Av6HT	HOXB2	P14652	AYTNTQLLELEKEFH	TQHREPPDGEPACPG	68
HR8135A-187-244-Av6HT	HOXB3	P14651	ASKRARTAYTSAQLV	RQIKIWFQNRRMKYK	58
HR8135A-196-252-Av6HT	HOXB3	P14651	TSAQLVELEKEFHFN	NRRMKYKKDQKAKGL	57
HR8135B-179-239-Av6HT	HOXB3	P14651	DKSPPGSAASKRART	LNLSERQIKIWFQNR	61
HR8135B-179-244-Av6HT	HOXB3	P14651	DKSPPGSAASKRART	RQIKIWFQNRRMKYK	66
HR8335A-169-222-Av6HT	HOXB4	P17483	MYTRQQVLELEKEFH	IWFQNRRMKWKKDHK	55
HR8335A-169-222-NHT	HOXB4	P17483	YTRQQVLELEKEFHY	IWFQNRRMKWKKDHK	54
HR8335A-169-222-TEV	HOXB4	P17483	YTRQQVLELEKEFHY	IWFQNRRMKWKKDHK	54
HR8261-201-257-Av6HT	HOXB5	P09067	YTRYQTLELEKEFHF	QNRRMKWKKDNKLKS	57
HR7319A-147-206-Av6HT	HOXB6	P17509	MRRGRQTYTRYQTLE	IWFQNRRMKWKKESK	61
HR7319A-147-206-TEV	HOXB6	P17509	RRGRQTYTRYQTLEL	IWFQNRRMKWKKESK	60
HR8504A-143-202-TEV	HOXB7	P09629	TYTRYQTLELEKEFH	RRMKWKKENKTAGPG	60
HR7846A-146-205-TEV	HOXB8	P17481	RRRGRQTYSRYQTLE	KIWFQNRRMKWKKEN	60
HR7230A-174-249-TEV	HOXB9	P17482	NPSANWLHARSSRKK	NRRMKMKKMNKEQGK	76
HR4478B-250-312-14	HOXC10	Q9NYD6	MKEEIKAENTTGNWL	RLEISKTINLTDRQV	64
HR4478B-255-342-14	HOXC10	Q9NYD6	MAENTTGNWLTAKSG	RENRIRELTSNFNFT	89

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4478B-263-312-14	HOXC10	Q9NYD6	MLTAKSGRKKRCPYT	RLEISKTINLTDRQV	51
HR4478B-268-342-14	HOXC10	Q9NYD6	MGRKKRCPYTKHQTL	RENRIRELTSNFNFT	76
HR4478C-247-342-14	HOXC10	Q9NYD6	MNEAKEEIKAENTTG	RENRIRELTSNFNFT	97
HR7286A-240-304-Av6HT	HOXC11	O43248	SKFQIRELEREFFFN	LSRDRLQYFSGNPLL	65
HR7847A-205-271-NHT	HOXC12	P31275	APWYPINSRSRKKRK	QVKIWFQNRRMKKKR	67
HR7251A-255-316-Av6HT	HOXC13	P31276	SSYRRGRKKRVPYTK	RQVTIWFQNRRVKEK	62
HR8257A-163-216-NHT	HOXC4	P09017	YTRQQVLELEKEFHY	IWFQNRRMKWKKDHR	54
HR8257A-163-216-TEV	HOXC4	P09017	YTRQQVLELEKEFHY	IWFQNRRMKWKKDHR	54
HR7011A-156-219-TEV	HOXC5	Q00444	KRSRTSYTRYQTLEL	NRRMKWKKDSKMKSK	64
HR7839A-148-201-TEV	HOXC6	P09630	YSRYQTLELEKEFHF	IWFQNRRMKWKKESN	54
HR6394A-149-208-TEV	HOXC8	P31273	RRSGRQTYSRYQTLE	KIWFQNRRMKWKKEN	60
HR7283A-180-255-TEV	HOXC9	P31274	SNPVANWIHARSTRK	QNRRMKMKKMNKEKT	76
HR8256A-200-280-Av6HT	HOXD1	Q9GZZ0	AAFSTFEWMKVKRNA	LHLNDTQVKIWFQNR	81
HR8148A-269-340-15	HOXD10	P28358	MKRCPYTKHQTLELE	RENRIRELTANLTFS	73
HR8148A-274-327-Av6HT	HOXD10	P28358	TKHQTLELEKEFLFN	WFQNRRMKLKKMSRE	54
HR8148A-274-340-15	HOXD10	P28358	MTKHQTLELEKEFLF	RENRIRELTANLTFS	68
HR8017A-257-326-Av6HT	HOXD11	P31277	SSSAVAPQRSRKKRC	IWFQNRRMKEKKLNR	70
HR7443A-276-333-TEV	HOXD13	P35453	GRKKRVPYTKLQLKE	QVTIWFQNRRVKDKK	58
HR7220A-181-257-Av6HT	HOXD3	P31249	GESCEDKSPPGPASK	QNRRMKYKKDQKAKG	77
HR7700A-161-220-TEV	HOXD4	P09016	YTRQQVLELEKEFHF	RMKWKKDHKLPNTKG	60
HR7832A-197-256-TEV	HOXD8	P13378	RRRGRQTYSRFQTLE	KIWFQNRRMKWKKEN	60
HR6999A-263-337-TEV	HOXD9	P28356	SQPQQQQLDPNNPAA	NLTERQVKIWFQNRR	75

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7031A-153-237-TEV	HP1BP3	Q5SSJ5	ASSPRPKMDAILTEA	GASGSFVVVQKSRKT	84
HR7031B-249-335-15	HP1BP3	Q5SSJ5	MSAVDPEPQVKLEDV	GASGTFQLKKSGEKP	88
HR7031B-254-330-15	HP1BP3	Q5SSJ5	MEPQVKLEDVLPLAF	QITGKGASGTFQLKK	78
HR7031B-262-330-Av6HT	HP1BP3	Q5SSJ5	VLPLAFTRLCEPKEA	QITGKGASGTFQLKK	69
HR7031C-332-407-15	HP1BP3	Q5SSJ5	MGEKPLLGGSLMEYA	KNGWMEQISGKGFSG	77
HR7031C-332-418-15	HP1BP3	Q5SSJ5	MGEKPLLGGSLMEYA	GFSGTFQLCFPYYPS	88
HR7031C-337-403-15	HP1BP3	Q5SSJ5	MLGGSLMEYAILSAI	QKCEKNGWMEQISGK	68
HR7031C-337-413-15	HP1BP3	Q5SSJ5	MLGGSLMEYAILSAI	QISGKGFSGTFQLCF	78
HR3023-1-506-15	HSF1	Q00613	MDLPVGPGAAGPSNV	FELGEGSYFSEGDGF	506
HR3023-1-506-Av6HT	HSF1	Q00613	DLPVGPGAAGPSNVP	FELGEGSYFSEGDGF	505
HR3023-1-506-TEV	HSF1	Q00613	DLPVGPGAAGPSNVP	FELGEGSYFSEGDGF	505
HR3023A-14	HSF1	Q00613	MDLPVGPGAAGPSNV	PERDDTEFQHPCFLR	106
HR3023A-15	HSF1	Q00613	MDLPVGPGAAGPSNV	PERDDTEFQHPCFLR	106
HR3023C-1-123-15	HSF1	Q00613	MDLPVGPGAAGPSNV	EQLLENIKRKVTSVS	123
HR3023C-10-123-15	HSF1	Q00613	MAGPSNVPAFLTKLW	EQLLENIKRKVTSVS	115
HR3023C-15-118-15	HSF1	Q00613	MVPAFLTKLWTLVSD	FLRGQEQLLENIKRK	105
HR3023C-7-118-15	HSF1	Q00613	MPGAAGPSNVPAFLT	FLRGQEQLLENIKRK	113
HR8180A-12-124-15	HSF4	Q9ULV5	MPGPSPVPAFLGKLW	REQLLERVRRKVPAL	114
HR8180A-12-97-15	HSF4	Q9ULV5	MPGPSPVPAFLGKLW	VVSIEQGGLLRPERD	87
HR8180A-17-119-15	HSF4	Q9ULV5	MVPAFLGKLWALVGD	SFVRGREQLLERVRR	104
HR8180A-17-93-15	HSF4	Q9ULV5	MVPAFLGKLWALVGD	GFRKVVSIEQGGLLR	78
HR8170A-9-94-Av6HT	HSF5	Q4G112	INPNNFPAKLWRLVN	FIRQLNLYGFRKVVL	86

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7245A-97-218-NHT	HSFX1	Q9UBD0	LPFPQKLWRLVSSNQ	LLVRMKRRVGVKSAP	122
HR3123-1-116-15	ID1	P41134	MKVASGSTATAAAGP	IRDLQLELNSESEVG	116
HR3123-1-121-15	ID1	P41134	MKVASGSTATAAAGP	LELNSESEVGTPGGR	121
HR3123-15	ID1	P41134	MKVASGSTATAAAGP	AEAACVPADDRILCR	155
HR3123-21	ID1	P41134	MKVASGSTATAAAGP	AEAACVPADDRILCR	155
HR3123A-14	ID1	P41134	ALKAGKTASGAGEVV	RDLQLELNSESEVGT	100
HR3123B-14	ID1	P41134	ALKAGKTASGAGEVV	AEAACVPADDRILCR	138
HR3123C-14	ID1	P41134	KTASGAGEVVRCLSE	RDLQLELNSESEVGT	95
HR3123D-14	ID1	P41134	KTASGAGEVVRCLSE	AEAACVPADDRILCR	133
HR3123E-14	ID1	P41134	AGEVVRCLSEQSVAI	RDLQLELNSESEVGT	90
HR3123F-14	ID1	P41134	AGEVVRCLSEQSVAI	AEAACVPADDRILCR	128
HR3123G-54-145-15	ID1	P41134	MPALLDEQQVNVLLY	TLNGEISALTAEAAC	93
HR3123G-59-139-15	ID1	P41134	MEQQVNVLLYDMNGC	VRAPLSTLNGEISAL	82
HR2921-14	ID2	Q02363	MKAFSPVRSVRKNSL	FPSELMSNDSKALCG	134
HR2921-15	ID2	Q02363	MKAFSPVRSVRKNSL	FPSELMSNDSKALCG	134
HR2921-17-85-14	ID2	Q02363	DHSLGISRSKTPVDD	YILDLQIALDSHPTI	69
HR2921-21	ID2	Q02363	MKAFSPVRSVRKNSL	FPSELMSNDSKALCG	134
HR2921-22-134-15	ID2	Q02363	MISRSKTPVDDPMSL	FPSELMSNDSKALCG	114
HR2921-22-85-14	ID2	Q02363	ISRSKTPVDDPMSLL	YILDLQIALDSHPTI	64
HR2921-27-124-15	ID2	Q02363	MTPVDDPMSLLYNMN	ISILSLQASEFPSEL	99
HR2921-27-134-15	ID2	Q02363	MTPVDDPMSLLYNMN	FPSELMSNDSKALCG	109
HR2921-27-85-14	ID2	Q02363	TPVDDPMSLLYNMND	YILDLQIALDSHPTI	59

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR2921-40-134-15	ID2	Q02363	MNDCYSKLKELVPSI	FPSELMSNDSKALCG	96
HR3111-14	ID3	Q712G9	MKALSPVRGCYEAVC	APELVISNDKRSFCH	119
HR3111-15	ID3	Q712G9	MKALSPVRGCYEAVC	APELVISNDKRSFCH	119
HR3111-21	ID3	Q712G9	MKALSPVRGCYEAVC	APELVISNDKRSFCH	119
HR3111A-27-83-15	ID3	Q712G9	MGRGKGPAAEEPLSL	ILQRVIDYILDLQVV	58
HR3111A-32-83-15	ID3	Q712G9	MPAAEEPLSLLDDMN	ILQRVIDYILDLQVV	53
HR4584C-58-112-14	ID4	P47928	DEPALCLQCDMNDCY	IDYILDLQLALETHP	55
HR4626B	IFI16	Q16666	QVTPRRNVLQKRPVI	ISEMHSFIQIKKKTN	202
HR3005-100-519-15	IKZF1	Q13422	MGSSALSGVGGIRLP	FSSHITRGEHRFHMS	421
HR3005-108-519-15	IKZF1	Q13422	MGGIRLPNGKLKCDI	FSSHITRGEHRFHMS	413
HR3005-93-519-15	IKZF1	Q13422	MNGSHRDQGSSALSG	FSSHITRGEHRFHMS	428
HR3005A-IDT-14	IKZF1	Q13422	HARNGLSLKEEHRAY	FSSHITRGEHRFHMS	99
HR3005B-IDT-14	IKZF1	Q13422	EKMNGSHRDQGSSAL	DRLASNVAKRKSSMP	190
HR3064-99-509-15	IKZF3	Q9UKT9	IKLERHVVSFDSSRP	FSSHIARGEHRALLK	411
HR6479A-436-509-NHT	IKZF3	Q9UKT9	RDSVKVINKEGEVMD	FSSHIARGEHRALLK	74
HR7992A-150-221-15	IKZF4	Q9H2S9	MGGIRLPNGKLKCDV	KLHSGEKPFKCPFCN	73
HR7992A-150-232-15	IKZF4	Q9H2S9	MGGIRLPNGKLKCDV	PFCNYACRRRDALTG	84
HR7992A-150-246-15	IKZF4	Q9H2S9	MGGIRLPNGKLKCDV	GHLRTHSVSSPTVGK	98
HR7992A-155-216-15	IKZF4	Q9H2S9	MPNGKLKCDVCGMVC	LLRHIKLHSGEKPFK	63
HR7992A-155-216-15.7-15TEV	IKZF4	Q9H2S9	MPNGKLKCDVCGMVC	LLRHIKLHSGEKPFK	63
HR7992A-155-239-15	IKZF4	Q9H2S9	MPNGKLKCDVCGMVC	RRRDALTGHLRTHSV	86
HR7992B-513-585-15	IKZF4	Q9H2S9	MSKEVLRVVGESGEP	YEFSSHIVRGEHKVG	74

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7992B-518-585-15	IKZF4	Q9H2S9	MRVVGESGEPVKAFK	YEFSSHIVRGEHKVG	69
HR7992B-523-585-15	IKZF4	Q9H2S9	MSGEPVKAFKCEHCR	YEFSSHIVRGEHKVG	64
HR7992B-528-585-15	IKZF4	Q9H2S9	MKAFKCEHCRILFLD	YEFSSHIVRGEHKVG	59
HR7992C-155-272-Av6HT	IKZF4	Q9H2S9	PNGKLKCDVCGMVCI	KQQSTLEEHKERCHN	118
HR7992C-159-272-Av6HT	IKZF4	Q9H2S9	LKCDVCGMVCIGPNV	KQQSTLEEHKERCHN	114
HR7992C-159-283-Av6HT	IKZF4	Q9H2S9	LKCDVCGMVCIGPNV	RCHNYLQSLSTEAQA	125
HR7992D-197-260-Av6HT	IKZF4	Q9H2S9	TQKGNLLRHIKLHSG	KPYKCNYCGRSYKQQ	64
HR7992D-208-283-Av6HT	IKZF4	Q9H2S9	LHSGEKPFKCPFCNY	RCHNYLQSLSTEAQA	76
HR7992D-210-260-Av6HT	IKZF4	Q9H2S9	SGEKPFKCPFCNYAC	KPYKCNYCGRSYKQQ	51
HR7630A-358-419-NHT	IKZF5	Q9H5V7	QDPQLLHHCQHCDMY	YDFACHFARGQHNQH	62
HR7614A-263-300-15	INSM1	Q01101	MPLGEFICQLCKEEY	KCSRIVRVEYRCPEC	39
HR7614A-263-319-15	INSM1	Q01101	MPLGEFICQLCKEEY	SCPANLASHRRWHKP	58
HR7614B-424-497-15	INSM1	Q01101	MGDGEGAGVLGLSAS	GLTRHINKCHPSENR	75
HR7614B-429-493-15	INSM1	Q01101	MAGVLGLSASAECHL	YSSPGLTRHINKCHP	66
HR7614B-432-497-15	INSM1	Q01101	MLGLSASAECHLCPV	GLTRHINKCHPSENR	67
HR8043A-261-315-Av6HT	INSM2	Q96T92	GEFICQLCKEQYADP	SCPANLASHRRWHKP	55
HR6405A-1-113-TEV	IRF1	P10914	PITRMRMRPWLEMQI	RNKGSSAVRVYRMLP	112
HR7043A-1-113-Av6HT	IRF2	P14316	MPVERMRMRPWLEEQ	IKKGNNAFRVYRMLP	113
HR7043A-1-113-TEV	IRF2	P14316	PVERMRMRPWLEEQI	IKKGNNAFRVYRMLP	112
HR7278A-1-113-TEV	IRF3	Q14653	GTPKPRILPWLVSQL	DPHDPHKIYEFVNSG	112
HR7278B-196-386-TEV	IRF3	Q14653	LVPGEEWEFEVTAFY	LRALVEMARVGGASS	191
HR3173-1-119-14	IRF5	Q13568	MNQSIPVAPTPRRRV	DGPRDMPPQPYKIYE	119

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR3173A-14	IRF5	Q13568	MNQSIPVAPTPPRRV	PPQPYKIYEVCSNGP	125
HR3173A-15	IRF5	Q13568	MNQSIPVAPTPPRRV	PPQPYKIYEVCSNGP	125
HR3173F-8-114-14	IRF5	Q13568	APTPPRRVRLKPWLV	FRLIYDGPRDMPPQP	107
HR3173G-232-477-TEV	IRF5	Q13568	EQLLPDLLISPHMLP	HIWQSQQRLQPVAQA	246
HR7755A-198-455-Av6HT	IRF6	O14896	LEMEVPQAPIQPFYS	RILQTQESWQPMQPT	258
HR5527A-14	IRF7	Q92985	MALAPERAAPRVLFG	RRFVMLRDNSGDPAD	117
HR5527A-15	IRF7	Q92985	MALAPERAAPRVLFG	RRFVMLRDNSGDPAD	117
HR8215A-8-154-TEV	IRF7	Q92985	AAPRVLFGEWLLGEI	EAEAPAAVPPPQGGP	147
HR7337A-9-115-TEV	IRF8	Q02556	RLRQWLIEQIDSSMY	LDISEPYKVYRIVPE	107
HR7302A-205-393-Av6HT	IRF9	Q00978	QRSLEFLLPPEPDYS	LEQTPEQQAAILSLV	189
HR7302A-209-393-Av6HT	IRF9	Q00978	EFLLPPEPDYSLLLT	LEQTPEQQAAILSLV	185
HR7431A-121-188-NHT	IRX1	P78414	GQFQYGDPGRPKNAT	VSTWFANARRRLKKE	68
HR7304A-126-209-NHT	IRX6	P78412	PYERTLGQYQYERYG	TWFANARRRLKKENK	84
HR8326A-180-244-TEV	ISL1	P61371	KTTRVRTVLNEKQLH	QNKRCKDKKRSIMMK	65
HR8291A-190-254-TEV	ISL2	Q96A47	KTTRVRTVLNEKQLH	QNKRCKDKKKSILMK	65
HR8400A-617-732-TEV	JARID2	Q92833	LGRRWGPNVQRLACI	RLEKEVLMEKEILEK	116
HR8400B-804-1099-Av6HT	JARID2	Q92833	KGVLNDFHKCIYKGR	LDELRDTELRQRRQL	296
HR8400B-809-1099-Av6HT	JARID2	Q92833	DFHKCIYKGRSVSLT	LDELRDTELRQRRQL	291
HR8400B-809-1104-Av6HT	JARID2	Q92833	DFHKCIYKGRSVSLT	DTELRQRRQLFEAGL	296
HR8400C-900-1086-Av6HT	JARID2	Q92833	GSILRHLGAVPGVTI	KENGPTLSTISALLD	187
HR8400C-900-1104-Av6HT	JARID2	Q92833	GSILRHLGAVPGVTI	DTELRQRRQLFEAGL	205
HR7951-28-149-Av6HT	JDP2	Q8WYK2	SALTVEELKYADIRN	HRPTCIVRTDSVKTP	122

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4484C-253-308-Av6HT	JUN	P05412	MIKAERKRMRNRIAA	LASTANMLREQVAQL	57
HR4484C-253-308-TEV	JUN	P05412	IKAERKRMRNRIAAS	LASTANMLREQVAQL	56
HR4765B-273-324-TEV	JUNB	P17275	RKRLRNRLAATKCRK	LSSTAGLLREQVAQL	52
HR4754B-269-324-TEV	JUND	P17535	IKAERKRLRNRIAAS	LASTASLLREQVAQL	56
HR2962A-5-79-Av6HT	KAT5	Q92993	GEIIEGCRLPVLRRN	LKKIQFPKKEAKTPT	75
HR7375A-94-208-TEV	KDM5B	Q9UGL1	EAQTRVKLNFLDQIA	LQKPNLTTDTKDKEY	115
HR7375B-685-750-Av6HT	KDM5B	Q9UGL1	LPDDERQCVKCKTTC	YTLDDLYPMMNALKL	66
HR7375B-696-750-Av6HT	KDM5B	Q9UGL1	KTTCFMSAISCSCKP	YTLDDLYPMMNALKL	55
HR7375C-1487-1544-Av6HT	KDM5B	Q9UGL1	CPAVSCLQPEGDEVD	YICVRCTVKDAPSRK	58
HR7375D-1485-1536-Av6HT	KDM5B	Q9UGL1	AICPAVSCLQPEGDE	PEMAEKEDYICVRCT	52
HR7375E-1123-1227-Av6HT	KDM5B	Q9UGL1	ESLSDLERALTESKE	LRIWLCPHCRRSEKP	105
HR7375E-1123-1241-Av6HT	KDM5B	Q9UGL1	ESLSDLERALTESKE	PPLEKILPLLASLQR	119
HR7375E-1132-1230-Av6HT	KDM5B	Q9UGL1	LTESKETASAMATLG	WLCPHCRRSEKPPLE	99
HR7375E-1134-1241-Av6HT	KDM5B	Q9UGL1	ESKETASAMATLGEA	PPLEKILPLLASLQR	108
HR7375E-1143-1230-Av6HT	KDM5B	Q9UGL1	ATLGEARLREMEALQ	WLCPHCRRSEKPPLE	88
HR7375E-1143-1241-Av6HT	KDM5B	Q9UGL1	ATLGEARLREMEALQ	PPLEKILPLLASLQR	99
HR7188A-306-385-TEV	KDM5D	Q9BY66	HSSAQFIDSYICQVC	EAFGFEQATQEYSLQ	80
HR7714A-77-157-Av6HT	KIAA1683	Q9H0B3	RRVPRLRAVVESQAF	RHILHSSKSLVKKTR	81
HR7682A-10-80-Av6HT	KIAA2018	Q68DE3	PTKKQHRKKNRETHN	ITELKRQNDELLLNG	71
HR8201A-51-160-TEV	KIN	O60870	QRQLLLASENPQQFM	PETIRRQLELEKKKK	110
HR7553A-272-338-15	KLF1	Q13351	MARKRQAAHTCAHPG	DELTRHYRKHTGQRP	68
HR7553A-292-335-Av6HT	KLF1	Q13351	KSSHLKAHLRTHTGE	ARSDELTRHYRKHTG	44

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7553B-319-362-Av6HT	KLF1	Q13351	RFARSDELTRHYRKH	FSRSDHLALHMKRHL	44
HR6400A-353-423-TEV	KLF10	Q13118	SAAKVTPQIDSSRIR	RFARSDELSRHRRTH	71
HR6390-1-497-15	KLF11	O14901	MHTPDFAGPDDARAV	PGWQAEVGKLNRIAS	497
HR6390-1-501-15	KLF11	O14901	MHTPDFAGPDDARAV	AEVGKLNRIASAESP	501
HR6390-12-512-15	KLF11	O14901	ARAVDIMDICESILE	AESPGSPLVSMPASA	501
HR6390-123-512-15	KLF11	O14901	VSPQVTDSKACTATD	AESPGSPLVSMPASA	390
HR6390-128-512-15	KLF11	O14901	TDSKACTATDVLQSS	AESPGSPLVSMPASA	385
HR6390-15	KLF11	O14901	MHTPDFAGPDDARAV	AESPGSPLVSMPASA	512
HR6390-7-512-15	KLF11	O14901	AGPDDARAVDIMDIC	AESPGSPLVSMPASA	506
HR6390A-379-501-15	KLF11	O14901	SQNCVPQVDFSRRRN	AEVGKLNRIASAESP	123
HR6390A-384-497-15	KLF11	O14901	PQVDFSRRRNYVCSF	PGWQAEVGKLNRIAS	114
HR6390B-397-462-15	KLF11	O14901	SFPGCRKTYFKSSHL	HTGEKKFVCPVCDRR	66
HR6390B-402-457-15	KLF11	O14901	RKTYFKSSHLKAHLR	RHRRTHTGEKKFVCP	56
HR7238A-306-400-TEV	KLF12	Q9Y4X4	SESPDSRKRRIHRCD	FSRSDHLALHRRRHM	95
HR8436A-125-193-Av6HT	KLF16	Q9BXK1	KSHRCPFPDCAKAYY	RTHTGEKRFSCPLCS	69
HR7123A-272-355-TEV	KLF2	Q9Y5W3	HTCSYAGCGKTYTKS	FSRSDHLALHMKRHM	84
HR7880A-251-343-TEV	KLF3	P57682	PDTQRKRRIHRCDYD	FSRSDHLALHRKRHM	93
HR4433-1-347-14	KLF5	Q13887	MATRVLSMSARLGPV	ASKLAIHNPNLPTTL	347
HR4433-9-342-14	KLF5	Q13887	MSARLGPVPQPPAPQ	YAATIASKLAIHNPN	335
HR4668C-168-283-21	KLF6	Q99612	MELPSPGKVRSGTSG	FSRSDHLALHMKRHL	117
HR4668C-173-283-21	KLF6	Q99612	MGKVRSGTSGKPGDK	FSRSDHLALHMKRHL	112
HR4668C-173-283-Av6HT	KLF6	Q99612	GKVRSGTSGKPGDKG	FSRSDHLALHMKRHL	111

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4668C-173-283-TEV	KLF6	Q99612	GKVRSGTSGKPGDKG	FSRSDHLALHMKRHL	111
HR4668C-191-283-21	KLF6	Q99612	MASPDGRRRVHRCHF	FSRSDHLALHMKRHL	94
HR4668C-196-283-21	KLF6	Q99612	MRRRVHRCHFNGCRK	FSRSDHLALHMKRHL	89
HR4668D-205-283-21	KLF6	Q99612	MNGCRKVYTKSSHLK	FSRSDHLALHMKRHL	80
HR4668D-210-283-21	KLF6	Q99612	MVYTKSSHLKAHQRT	FSRSDHLALHMKRHL	75
HR8165A-231-302-Av6HT	KLF7	O75840	TKSSHLKAHQRTHTG	FSRSDHLALHMKRHI	72
HR8376A-270-332-Av6HT	KLF8	O95600	RRRIHQCDFAGCSKV	SDELTRHFRKHTGIK	63
HR7597A-181-244-NHT	KLF9	Q13886	LKKFSRSDELTRHYR	PSMIKRSKKALANAL	64
HR6918A-877-937-TEV	KNL2	Q6P0N0	DKEWNEKELQKLHCA	MENPRGKGSQKHVTK	61
HR6489A-15	L3MBTL3	Q96JM7	RRKRRGDSAVLKQGL	QPPLSPLELMEASEH	346
HR6489A-Av6HT	L3MBTL3	Q96JM7	RRKRRGDSAVLKQGL	QPPLSPLELMEASEH	346
HR6489A-TEV	L3MBTL3	Q96JM7	RRKRRGDSAVLKQGL	QPPLSPLELMEASEH	346
HR6490A-15	L3MBTL4	Q8NA19	MKQPNRKRKLNMDSK	SAFGCPYSDMNLKKE	414
HR6490A-30-371-15	L3MBTL4	Q8NA19	MEKKPKDSTTPLSHV	TGHPLEVPQRTNDLK	343
HR6490A-30-371-Av6HT	L3MBTL4	Q8NA19	EKKPKDSTTPLSHVP	TGHPLEVPQRTNDLK	342
HR6490A-30-371-Na6HT	L3MBTL4	Q8NA19	EKKPKDSTTPLSHVP	TGHPLEVPQRTNDLK	342
HR6490A-30-371-TEV	L3MBTL4	Q8NA19	EKKPKDSTTPLSHVP	TGHPLEVPQRTNDLK	342
HR6490A-Av6HT	L3MBTL4	Q8NA19	KQPNRKRKLNMDSKE	SAFGCPYSDMNLKKE	413
HR6490A-TEV	L3MBTL4	Q8NA19	KQPNRKRKLNMDSKE	SAFGCPYSDMNLKKE	413
HR2473-14	LARP1	Q6PKG0	MNTLFRFWSFFLRDH	AKWTSQHSNTQTLGK	185
HR7995A-377-483-Av6HT	LARP1	Q6PKG0	ISLIFAALKDSKVVE	SASLPDLDSENWIEV	107
HR6994A-210-292-NHT	LARP1B	Q659C4	VEEALLKEYIKRQIE	EVEIVDEKMRKKIEP	83

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7969A-107-200-TEV	LARP4	Q71RC2	SGESNSAVSTEDLKE	DEKGEKVRPSHKRCI	94
HR6949A-152-237-NHT	LARP4B	Q92615	SQEDPREVLKKTLEF	DEKGEKVRPNQNRCI	86
HR7099A-69-134-NHT	LASS2	Q96G23	NIKEKTRLRAPPNAT	RRRNQDRPSLLKKFR	66
HR8001A-86-135-TEV	LASS5	Q8N5B7	AQPNAILEKVFISIT	KIQCWFRHRRNQDKP	50
HR6906A-77-127-TEV	LASS6	Q6ZMG9	APPNAILEKVFTAIT	IQRWFRQRRNQEKPS	51
HR6954A-124-198-NHT	LBX1	P52954	KRRKSRTAFTNHQIY	LEEMKADVESAKKLG	75
HR8118A-343-405-TEV	LCOR	Q96JN0	RGRYRQYNSEILEEA	GTLKNPPKKKMKLMR	63
HR7552A-519-579-TEV	LCORL	Q8N3X6	RGRYRQYDHEIMEEA	RSGTLKTPPKKKLRL	61
HR7767A-314-501-Av6HT	LENG9	Q96B70	APCQPRPTHFVALMV	RTGGPFQPLAEIRLE	188
HR8129A-23-126-Av6HT	LHX1	P48742	AWHVKCVQCCECKCN	FVCKEDYLSNSSVAK	104
HR7637A-262-334-TEV	LHX2	P50458	SSQKTKRMRTSFKHH	KFRRNLLRQENTGVD	73
HR7663A-23-150-TEV	LHX3	Q9UBR4	LARRADLRREIPLCA	FYLMEDSRLVCKADY	128
HR7789A-16-91-NHT	LHX4	Q969G2	LPEMLGVPMQQIPQC	CKEDFFKRFGTKCTA	76
HR7587A-24-119-NHT	LHX5	Q9H2C1	AWHIKCVQCCECKTN	LYVIDENKFVCKDDY	96
HR7172-1-267-TEV	LHX9	Q9NQ69	LNGTTLEAAMLFHGI	PPSQKTKRMRTSFKH	266
HR7172-1-270-15	LHX9	Q9NQ69	MLNGTTLEAAMLFHG	QKTKRMRTSFKHHQL	270
HR7525A-134-187-Av6HT	LIN28A	Q9H9Z2	MSKGDRCYNCGGLDH	SCPLKAQQGPSAQGK	55
HR7525A-134-187-TEV	LIN28A	Q9H9Z2	SKGDRCYNCGGLDHH	SCPLKAQQGPSAQGK	54
HR7198A-25-103-NHT	LIN28B	Q6ZN17	SQVLRGTGHCKWFNV	KSSKGLESIRVTGPG	79
HR7658-1-237-Av6HT	LMX1A	Q8TE12	LDGLKMEENFQSAID	KVRETLAAETGLSVR	236
HR7658-1-247-Av6HT	LMX1A	Q8TE12	LDGLKMEENFQSAID	GLSVRVVQVWFQNQR	246
HR7658-1-257-Av6HT	LMX1A	Q8TE12	LDGLKMEENFQSAID	FQNQRAKMKKLARRQ	256

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7658-13-303-Av6HT	LMX1A	Q8TE12	SAIDTSASFSSLLGR	PYTALPTPQQLLAIE	291
HR7658A-61-153-NHT	LMX1A	Q8TE12	QCASCKEPLETTCFY	EGQLLCKGDYEKERE	93
HR7658B-13-153-Av6HT	LMX1A	Q8TE12	SAIDTSASFSSLLGR	EGQLLCKGDYEKERE	141
HR7658B-32-153-Av6HT	LMX1A	Q8TE12	KSVCEGCQRVILDRF	EGQLLCKGDYEKERE	122
HR6403A-128-207-15	LYL1	P12980	RLKRRPSHCELDLAE	RLAMKYIGFLVRLLR	80
HR6403A-133-207-15	LYL1	P12980	PSHCELDLAEGHQPQ	RLAMKYIGFLVRLLR	75
HR6403A-146-207-15	LYL1	P12980	PQKVARRVFTNSRER	RLAMKYIGFLVRLLR	62
HR6403A-146-226-15	LYL1	P12980	PQKVARRVFTNSRER	ALAAGPTPPGPRKRP	81
HR7569A-1-80-TEV	MAEL	Q96JY0	PNRKASRNAYYFFVQ	GKDPGPSEKQKPVFT	79
HR4779B-255-300-14	MAF	O75444	MLHFDDRFSDEQLVT	VIRLKQKRRTLKNRG	47
HR7214A-221-319-Av6HT	MAFA	Q8NHW3	VRLEERFSDDQLVSM	KERDLYKEKYEKLAG	99
HR7214A-225-319-Av6HT	MAFA	Q8NHW3	ERFSDDQLVSMSVRE	KERDLYKEKYEKLAG	95
HR7214A-228-313-Av6HT	MAFA	Q8NHW3	SDDQLVSMSVRELNR	EVGRLAKERDLYKEK	86
HR7214A-236-319-Av6HT	MAFA	Q8NHW3	SVRELNRQLRGFSKE	KERDLYKEKYEKLAG	84
HR7214A-246-319-Av6HT	MAFA	Q8NHW3	GFSKEEVIRLKQKRR	KERDLYKEKYEKLAG	74
HR6931A-209-305-Av6HT	MAFB	Q9Y5Q3	DRFSDDQLVSMSVRE	RDAYKVKCEKLANSG	97
HR6931B-210-236-Av6HT	MAFB	Q9Y5Q3	RFSDDQLVSMSVREL	RELNRHLRGFTKDEV	27
HR6931B-210-251-Av6HT	MAFB	Q9Y5Q3	RFSDDQLVSMSVREL	IRLKQKRRTLKNRGY	42
HR8265A-31-74-Av6HT	MAFF	Q9ULX9	GLSVRELNRHLRGLS	KNRGYAASCRVKRVC	44
HR7795A-21-123-TEV	MAFG	O15525	GTSLTDEELVTMSVR	SKYEALQTFARTVAR	103
HR7958A-24-123-TEV	MAFK	O60675	LSDDELVSMSVRELN	SKYEALQTFARTVAR	100
HR8183A-390-479-Av6HT	MATR3	P43243	MQKGRVETSRVVHIM	PVRVHLSQKYKRIKK	91

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8183A-390-479-TEV	MATR3	P43243	QKGRVETSRVVHIMD	PVRVHLSQKYKRIKK	90
HR8110A-22-107-TEV	MAX	P61244	ADKRAHHNALERKRR	ALLEQQVRALEKARS	86
HR8332A-280-361-NHT	MAZ	P56270	ACEMCGKAFRDVYHL	SRPDHLNSHVRQVHS	82
HR8332A-280-361-TEV	MAZ	P56270	ACEMCGKAFRDVYHL	SRPDHLNSHVRQVHS	82
HR8039A-131-243-TEV	MBD1	Q9UIS9	GCCENCGISFSGDGT	RGCQTQEDCGHCPIC	113
HR8039A-131-262-TEV	MBD1	Q9UIS9	GCCENCGISFSGDGT	RPGLRRQWKCVQRRC	132
HR5530A-14	MBD2	Q9UBB5	MEPVPFPSGSAGPGP	NDPLNQNKGKPDLNT	118
HR5530A-15	MBD2	Q9UBB5	MEPVPFPSGSAGPGP	NDPLNQNKGKPDLNT	118
HR6416-1-220-15	MBD3	O95983	MERKRWECPALPQGW	VWLNTTQPLCKAFMV	220
HR6416-1-226-15	MBD3	O95983	MERKRWECPALPQGW	QPLCKAFMVTDEDIR	226
HR6416-1-261-15	MBD3	O95983	MERKRWECPALPQGW	MLAHVEELARDGEAP	261
HR6416-15	MBD3	O95983	MERKRWECPALPQGW	EEEEEPDPDPEMEHV	291
HR6416-33-291-15	MBD3	O95983	VFYYSPSGKKFRSKP	EEEEEPDPDPEMEHV	259
HR6416-55-291-15	MBD3	O95983	MGSMDLSTFDFRTGK	EEEEEPDPDPEMEHV	238
HR6416A-1-106-15	MBD3	O95983	MERKRWECPALPQGW	KPDLNTALPVRQTAS	106
HR6416A-1-111-15	MBD3	O95983	MERKRWECPALPQGW	TALPVRQTASIFKQP	111
HR6416A-1-117-15	MBD3	O95983	MERKRWECPALPQGW	QTASIFKQPVTKITN	117
HR6416B-1-72-15	MBD3	O95983	MERKRWECPALPQGW	DLSTFDFRTGKMLMS	72
HR6416B-1-77-15	MBD3	O95983	MERKRWECPALPQGW	DFRTGKMLMSKMNKS	77
HR4635B-14	MBD4	O95243	MTECRKSVPCGWERV	VLSKRGIKSRYKDCS	81
HR4635B-15	MBD4	O95243	MTECRKSVPCGWERV	VLSKRGIKSRYKDCS	81
HR4635C-14	MBD4	O95243	RSSECNPLLQEPIAS	FTVLSKRGIKSRYKD	101

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4635D-55-161-14	MBD4	O95243	MIKRSSECNPLLQEP	SKRGIKSRYKDCSMA	108
HR4635D-55-161-Av6HT	MBD4	O95243	IKRSSECNPLLQEPI	SKRGIKSRYKDCSMA	107
HR4635D-55-161-TEV	MBD4	O95243	IKRSSECNPLLQEPI	SKRGIKSRYKDCSMA	107
HR4635D-55-191-14	MBD4	O95243	MIKRSSECNPLLQEP	NLRTRSKCKKDVFMP	138
HR4635D-61-156-14	MBD4	O95243	MCNPLLQEPIASAQF	DFTVLSKRGIKSRYK	97
HR4635D-61-186-14	MBD4	O95243	MCNPLLQEPIASAQF	NNSNWNLRTRSKCKK	127
HR4635E-437-574-TEV	MBD4	O95243	KWTPPRSPFNLVQET	DHKLNKYHDWLWENH	138
HR8088A-178-246-TEV	MBNL1	Q9NR56	RTDRLEVCREYQRGN	EKCKYFHPPAHLQAK	69
HR7551A-175-243-TEV	MBNL2	Q5VZF2	RTDKLEVCREFQRGN	EKCKYFHPPAHLQAK	69
HR7762A-173-241-TEV	MBNL3	Q9NUK0	RCSREKCKYFHPPAH	NGATPVFNPTVFHCQ	69
HR3168-14	MDS1	Q13465	MRSKGRARKLATNNE	QADVYMPGLQCAFLS	169
HR7632A-77-171-TEV	MECP2	P51608	SEGSGSAPAVPEASA	VGDTSLDPNDFDFTV	95
HR4583C-2-78-TEV	MEF2A	Q02078	GRKKIQITRIMDERN	KVLLKYTEYNEPHES	77
HR8120A-2-94-TEV	MEF2B	Q02080	GRKKIQISRILDQRN	TNTDILETLKRRGIG	93
HR4550C-2-78-TEV	MEF2D	Q14814	GRKKIQIQRITDERN	KVLLKYTEYNEPHES	77
HR8225A-277-341-NHT	MEIS1	O00470	GIFPKVATNIMRAWL	RRRIVQPMIDQSNRA	65
HR8225A-277-341-TEV	MEIS1	O00470	GIFPKVATNIMRAWL	RRRIVQPMIDQSNRA	65
HR8514A-250-313-TEV	MEIS3P2	A8K0S8	GIFPKVATNIMRAWL	ARRRMVQPMIDQSNR	64
HR7119A-175-247-NHT	MEOX2	P50222	QEGNYKSEVNSKPRK	VWFQNRRMKWKRVKG	73
HR7798B-1047-1351-Av6HT	MET	P08581	LQNTVHIDLSALNPE	ISAIFSTFIGEHYVH	305
HR8521A-1-179-TEV	MGMT	P16455	DKDCEMKRTTLDSPL	KEWLLAHEGHRLGKP	178
HR7181A-199-243-TEV	MIER1	Q8N108	YKENEKVYENDDQLL	KDASRRTGDEKGVEA	45

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7181A-199-265-TEV	MIER1	Q8N108	YKENEKVYENDDQLL	KDNEQALYELVKCNF	67
HR7181A-203-243-TEV	MIER1	Q8N108	EKVYENDDQLLWDPE	KDASRRTGDEKGVEA	41
HR7181A-203-265-TEV	MIER1	Q8N108	EKVYENDDQLLWDPE	KDNEQALYELVKCNF	63
HR7181A-208-260-TEV	MIER1	Q8N108	NDDQLLWDPEYLPED	EGSHIKDNEQALYEL	53
HR3622D-1-299-14	MINK1	Q8N4C8	MGDPAPARSLDDIDL	KFPFIRDQPTERQVR	299
HR3622D-13-294-14	MINK1	Q8N4C8	MIDLSALRDPAGIFE	TEQLLKFPFIRDQPT	283
HR3622D-8-299-14	MINK1	Q8N4C8	MRSLDDIDLSALRDP	KFPFIRDQPTERQVR	293
HR3622D-9-294-14	MINK1	Q8N4C8	MSLDDIDLSALRDPA	TEQLLKFPFIRDQPT	287
HR3622E-1180-1284-14	MINK1	Q8N4C8	MIYGSSAGFHAVDVD	GEKAIEIRSVETGHL	106
HR3622E-1185-1282-14	MINK1	Q8N4C8	MAGFHAVDVDSGNSY	GWGEKAIEIRSVETG	99
HR7746A-86-152-Av6HT	MIXL1	Q9H2W2	QRRKRTSFSAEQLQL	RAKSRRQSGKSFQPL	67
HR7244A-248-367-NHT	MKRN1	Q9UHC7	DAAQRSQHIKSCIEA	EKQKLILKYKEAMSN	120
HR7430A-278-369-NHT	MKRN3	Q13064	DAAQREEHMRACIEA	NRIVKSCPQCRVTSE	92
HR7905A-54-146-Av6HT	MKX	Q8IYA7	NLGLRHRRTGARQNG	VRQPDLSWALRIKLY	93
HR4516M-1422-1490-15	MLL	Q03164	MILTSVPITPRVVCF	CRRCKFCHVCGRQHQ	70
HR4516M-1422-1514-15	MLL	Q03164	MILTSVPITPRVVCF	KCRNSYHPECLGPNY	94
HR4516M-1427-1486-15	MLL	Q03164	MPITPRVVCFLCASS	ENWCCRRCKFCHVCG	61
HR4516M-1427-1514-15	MLL	Q03164	MPITPRVVCFLCASS	KCRNSYHPECLGPNY	89
HR4516N-1476-1537-15	MLL	Q03164	MCRRCKFCHVCGRQH	KVWICTKCVRCKSCG	63
HR4516O-2012-2081-15	MLL	Q03164	MNGLEPENIHMMIGS	YTCKIVECRPPVVEP	71
HR4516O-2017-2076-15	MLL	Q03164	MENIHMMIGSMTIDC	RKRCVYTCKIVECRP	61
HR4516O-2017-2082-15	MLL	Q03164	MENIHMMIGSMTIDC	TCKIVECRPPVVEPD	67

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8195A-1-143-Av6HT	MLLT1	Q03111	DNQCTVQVRLELGHR	TEFRYKLLRAGGVMV	142
HR7909A-1-139-Av6HT	MLLT3	P42568	ASSCAVQVKLELGHR	NNPTEDFRRKLLKAG	138
HR7716A-121-220-NHT	MLX	Q9UH92	AYKESYKDRRRRAHT	TALKIMKVNYEQIVK	100
HR7887A-717-802-Av6HT	MLXIP	Q9HAP2	LKNRQMKHISAEQKR	EELNATIISCQQLLP	86
HR7887A-726-802-Av6HT	MLXIP	Q9HAP2	SAEQKRRFNIKMCFD	EELNATIISCQQLLP	77
HR7434A-647-736-NHT	MLXIPL	Q9NP71	TENRRITHISAEQKR	EELNAAINLCQQQLP	90
HR7223A-347-541-Av6HT	MRF	Q9Y2G1	NYQSIKWQPHQQNKW	IIVRASNPGQFESDS	195
HR7242A-112-186-TEV	MRRF	Q96E11	ESGMNLNPEVEGTLI	DTVSEDTIRLIEKQI	75
HR4485B-103-183-14	MSC	O60682	MECKQSQRNAANARE	ENGYVHPVNLTWPFV	82
HR4485B-103-188-14	MSC	O60682	MECKQSQRNAANARE	HPVNLTWPFVVSGRP	87
HR4485B-103-188-Av6HT	MSC	O60682	ECKQSQRNAANARER	HPVNLTWPFVVSGRP	86
HR4485B-103-188-TEV	MSC	O60682	ECKQSQRNAANARER	HPVNLTWPFVVSGRP	86
HR4485B-103-194-14	MSC	O60682	MECKQSQRNAANARE	WPFVVSGRPDSDTKE	93
HR4485B-103-199-14	MSC	O60682	MECKQSQRNAANARE	SGRPDSDTKEVSAAN	98
HR4485B-135-194-14	MSC	O60682	MPWVPPDTKLSKLDT	WPFVVSGRPDSDTKE	61
HR4485C-103-174-14	MSC	O60682	MECKQSQRNAANARE	ROLLQEDRYENGYVH	73
HR7186A-122-193-NHT	MSGN1	A6NI15	SVQRRRKASEREKLR	TDLLNRGREPRAQSA	72
HR7207A-25-769-TEV	MST1R	Q04912	EDWQCPRTPYAASRD	GAQVPGSWTFQYRED	745
HR4585B-167-224-TEV	MSX1	P28360	RKPRTPFTTAQLLAL	VKIWFQNRRAKAKRL	58
HR7691A-143-200-TEV	MSX2	P35548	RKPRTPFTTSQLLAL	VKIWFQNRRAKAKRL	58
HR4538-1-540-14	MTA1	Q13330	MAANMYRVGDYVYFE	LKQAVRKPLEAVLRY	540
HR4538-1-540-15	MTA1	Q13330	MAANMYRVGDYVYFE	LKQAVRKPLEAVLRY	540

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4538-1-545-14	MTA1	Q13330	MAANMYRVGDYVYFE	RKPLEAVLRYLETHP	545
HR4538-1-545-15	MTA1	Q13330	MAANMYRVGDYVYFE	RKPLEAVLRYLETHP	545
HR4538C-375-438-14	MTA1	Q13330	MGVVNGTGAPGQSPG	WKKYGGLKMPTRLDG	65
HR4538C-380-433-14	MTA1	Q13330	MTGAPGQSPGAGRAC	SCWTYWKKYGGLKMP	55
HR4538D-15	MTA1	Q13330	ALVPQGGPVLCRDEM	QVYIPNYNKPNPNQI	97
HR4538D-Av6HT	MTA1	Q13330	ALVPQGGPVLCRDEM	QVYIPNYNKPNPNQI	97
HR4538D-TEV	MTA1	Q13330	ALVPQGGPVLCRDEM	QVYIPNYNKPNPNQI	97
HR4621B-352-412-15	MTA2	O94776	MSKPGMNGAGFQKGL	WKKYGGLKTPTQLEG	62
HR4621B-357-407-15	MTA2	O94776	MNGAGFQKGLTCESC	SCWIYWKKYGGLKTP	52
HR4621C-1-140-15	MTA2	O94776	MAANMYRVGDYVYFE	CFFYSLVFDPVQKTL	140
HR4621C-1-145-15	MTA2	O94776	MAANMYRVGDYVYFE	LVFDPVQKTLLADQG	145
HR4621C-1-161-15	MTA2	O94776	MAANMYRVGDYVYFE	IRVGCKYQAEIPDRL	161
HR4621C-1-166-15	MTA2	O94776	MAANMYRVGDYVYFE	KYQAEIPDRLVEGES	166
HR4468C-268-324-TEV	MTA3	Q9BTC8	AISALVPQGGPVLCR	IQQDFLPWKSLTSII	57
HR6907A-148-207-NHT	MTF1	Q14872	PRTYSTAGNLRTHQK	VHTKEKPFECDVQGC	60
HR6878A-57-136-TEV	MXD1	Q05195	SRSTHNEMEKNRRAH	LQREQRHLKRQLEKL	80
HR6454A-24-140-14	MXD4	Q14582	EHGYASVLPFDGDFA	LKRRLEQLSVQSVER	117
HR6454A-44-143-14	MXD4	Q14582	AAGLVRKAPNNRSSH	RLEQLSVQSVERVRT	100
HR6454A-44-143-15	MXD4	Q14582	AAGLVRKAPNNRSSH	RLEQLSVQSVERVRT	100
HR6454A-49-140-14	MXD4	Q14582	RKAPNNRSSHNELEK	LKRRLEQLSVQSVER	92
HR6454A-56-136-14	MXD4	Q14582	SSHNELEKHRRAKLR	EHRFLKRRLEQLSVQ	81
HR6454A-56-136-15	MXD4	Q14582	SSHNELEKHRRAKLR	EHRFLKRRLEQLSVQ	81

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6436A-58-155-14	MXI1	P50539	SSGSSNTSTANRSTH	KWRLEQLQGPQEMER	98
HR6436A-63-149-14	MXI1	P50539	NTSTANRSTHNELEK	REQRFLKWRLEQLQG	87
HR6436A-63-155-14	MXI1	P50539	NTSTANRSTHNELEK	KWRLEQLQGPQEMER	93
HR6436A-68-149-14	MXI1	P50539	NRSTHNELEKNRRAH	REQRFLKWRLEQLQG	82
HR8112A-85-136-TEV	MYBL1	P10243	LIKGPWTKEEDQRVI	GKQCRERWHNHLNPE	52
HR3593B-28-78-TEV	MYBL2	P10244	SKCKVKWTHEEDEQL	RTDQQCQYRWLRVLN	51
HR4620B-353-438-TEV	MYC	P01106	NVKRRTHNVLERQRR	REQLKHKLEQLRNSC	86
HR7184A-279-364-NHT	MYCL1	P12524	DVTKRKNHNFLERKR	RQQQLQKRIAYLTGY	86
HR8502A-1-111-Av6HT	MYCL2	P12525	DRDSYHHYFYDYDGG	EPLERAVSDLLAVGA	110
HR6419A-367-464-14	MYCN	P04198	AKSLSPRNSDSEDSE	RQQQLLKKIEHARTC	98
HR7983A-4-117-TEV	MYNN	Q9NPC7	SHHCEHLLERLNKQR	KVEEVVTKCKIKMED	114
HR4693-66-224-14	MYOG	P15173	MLPWACKVCKRKSVS	VEDVSVAFPDETMPN	160
HR4693B-71-145-14	MYOG	P15173	MKVCKRKSVSVDRRR	RLQALLSSLNQEERD	76
HR4693B-73-156-14	MYOG	P15173	MCKRKSVSVDRRRAA	EERDLRYRGGGGPQP	85
HR4693B-76-138-14	MYOG	P15173	MKSVSVDRRRAATLR	SAIQYIERLQALLSS	64
HR4693B-78-156-14	MYOG	P15173	MVSVDRRRAATLREK	EERDLRYRGGGGPQP	80
HR7507A-115-181-TEV	MYSM1	Q5VVJ2	ASYSVKWTIEEKELF	VKCGLDKETPNQKTG	67
HR7507B-367-470-TEV	MYSM1	Q5VVJ2	HEEEELKPPEQEIEL	IGAINFGCEQAVYNR	104
HR4437B-298-605-NHT	MYST2	O95251	LENLTSEYDLDLFRR	RSNSNKTMDPSCLKW	308
HR8033A-559-650-TEV	MYT1	Q01538	SYRPNVAPATPRANL	LSTRCWEMPENLSTK	92
HR6948A-486-549-TEV	MYT1L	Q9UL68	HVKKPYYDPSRTEKK	PPEILAMHESVLKCP	64
HR7215A-37-128-TEV	MZF1	P28698	DPGPEAARLRFRCFR	EAAALVDGLRREPGG	92

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6963A-75-157-TEV	NANOG	Q9H9S0	KQPTSAEKSVAKKED	FQNQRMKSKRWQKNN	83
HR7935-106-160-Av6HT	NANOGNB	Q7Z5D8	KRLVSKSLMHTLWAK	TSQWFCKTRKKYNKE	55
HR8537A-41-101-Av6HT	NANOGP1	Q8N7R0	TRTVFSSTQLCVLND	QNQRMKSKRWQKNNW	61
HR3639F-24-96-15	NCOA1	Q15788	MCDTLASSTEKRRRE	RMEQEKSTTDDDVQK	74
HR3639F-29-91-15	NCOA1	Q15788	MSSTEKRRREQENKY	IQLMKRMEQEKSTTD	64
HR3639G-104-174-15	NCOA1	Q15788	MQGVIEKESLGPLLL	LHVGDHAEFVKNLLP	72
HR3639G-107-179-15	NCOA1	Q15788	MIEKESLGPLLLEAL	HAEFVKNLLPKSLVN	74
HR3639G-112-174-15	NCOA1	Q15788	MLGPLLLEALDGFFF	LHVGDHAEFVKNLLP	64
HR3639G-112-203-15	NCOA1	Q15788	MLGPLLLEALDGFFF	RRNSHTFNCRMLIHP	93
HR3639G-99-179-15	NCOA1	Q15788	MISSSSQGVIEKESL	HAEFVKNLLPKSLVN	82
HR3639H-1185-1441-15	NCOA1	Q15788	MSPFSQLAANPEASL	PQAQQKSLLQQLLTE	258
HR3639H-1190-1441-15	NCOA1	Q15788	MLAANPEASLANRNS	PQAQQKSLLQQLLTE	253
HR3639H-1205-1441-15	NCOA1	Q15788	MVSRGMTGNIGGQFG	PQAQQKSLLQQLLTE	238
HR3639H-1210-1441-15	NCOA1	Q15788	MTGNIGGQFGTGINP	PQAQQKSLLQQLLTE	233
HR3639H-1216-1441-15	NCOA1	Q15788	MQFGTGINPQMQQNV	PQAQQKSLLQQLLTE	227
HR4453I-100-258-Av6HT	NCOA3	Q9Y6Q9	MVSSTGQGVIDKDSL	SCMICVARRITTGER	160
HR4453I-100-258-NHT	NCOA3	Q9Y6Q9	VSSTGQGVIDKDSLG	SCMICVARRITTGER	159
HR7885A-433-486-TEV	NCOR1	O75376	DRQFMNVWTDHEKEI	PDCVLYYYLTKKNEN	54
HR4636E-602-671-14	NCOR2	Q9Y618	MAELASMELNESSRW	KKRQNLDEILQQHKL	71
HR4636E-608-670-14	NCOR2	Q9Y618	MELNESSRWTEEEME	YKKRQNLDEILQQHK	64
HR7360A-102-160-TEV	NEUROD1	Q13562	RRMKANARERNRMHG	AKNYIWALSEILRSG	59
HR7134A-122-180-TEV	NEUROD2	Q15784	RRQKANARERNRMHD	AKNYIWALSEILRSG	59

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7078A-87-146-TEV	NEUROD4	Q9HD90	ARRVKANARERTRMH	ARNYIWALSEVLETG	60
HR8276A-95-153-NHT	NEUROD6	Q96NK8	RRQEANARERNRMHG	AKNYIWALSEILRIG	59
HR8276A-95-153-TEV	NEUROD6	Q96NK8	RRQEANARERNRMHG	AKNYIWALSEILRIG	59
HR6971A-104-175-NHT	NEUROG2	Q9H2A3	ETVQRIKKTRRLKAN	IWALTETLRLADHCG	72
HR7673A-76-167-NHT	NEUROG3	Q9Y4Z2	ALSKQRRSRRKKAND	APHCGELGSPGGSPG	92
HR7259A-264-544-TEV	NFAT5	O94916	KKSPMLCGQYPVKSE	AGRSHDVQPFTYTPD	281
HR8282A-416-591-Av6HT	NFATC1	O95644	MDWQLPSHSGPYELR	SLQVASNPIECSQRS	177
HR8282A-416-591-TEV	NFATC1	O95644	DWQLPSHSGPYELRI	SLQVASNPIECSQRS	176
HR7889A-421-595-TEV	NFATC3	Q12968	DWPLPAHFGQCELKI	SLQIASIPVECSQRS	175
HR4653B-214-293-14	NFE2	Q16621	MAKPTARGEAGSRDE	AAQNCRKRKLETIVQ	81
HR4653B-218-293-14	NFE2	Q16621	MARGEAGSRDERRAL	AAQNCRKRKLETIVQ	77
HR4653B-223-293-14	NFE2	Q16621	MGSRDERRALAMKIP	AAQNCRKRKLETIVQ	72
HR4653B-234-293-14	NFE2	Q16621	MKIPFPTDKIVNLPV	AAQNCRKRKLETIVQ	61
HR4653C-259-338-14	NFE2	Q16621	MLTESQLALVRDIRR	QQLTELYRDIFQHLR	81
HR4653C-274-338-14	NFE2	Q16621	MGKNKVAAQNCRKRK	QQLTELYRDIFQHLR	66
HR7672A-605-674-NHT	NFE2L1	Q14494	DFLDKQMSRDEHRAR	RRGKNKMAAQNCRKR	70
HR3520B-21	NFE2L2	Q16236	MMDLELPPPGLPSQQ	TSGSANYSQVAHIPK	110
HR3520F-21	NFE2L2	Q16236	DMDLIDILWRQDIDL	TSGSANYSQVAHIPK	95
HR3520L-455-594-14	NFE2L2	Q16236	TRDELRAKALHIPFP	EYSLQQTRDGNVFLV	140
HR3520L-455-599-14	NFE2L2	Q16236	TRDELRAKALHIPFP	QTRDGNVFLVPKSKK	145
HR3520M-435-523-15	NFE2L2	Q16236	MGHRKTPFTKDKHSS	VAAQNCRKRKLENIV	90
HR3520M-440-523-15	NFE2L2	Q16236	MPFTKDKHSSRLEAH	VAAQNCRKRKLENIV	85

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR3520N-489-570-15	NFE2L2	Q16236	MQFNEAQLALIRDIR	QLSTLYLEVFSMLRD	83
HR3520O-445-529-15	NFE2L2	Q16236	MKHSSRLEAHLTRDE	VAAQNCRKRKLENIV	80
HR7720A-530-598-NHT	NFE2L3	Q9Y4A8	DTDRNLSRDEQRAKA	RRGKNKVAAQNCRKR	69
HR7383A-63-165-TEV	NFIA	Q12857	EKPEVKQKWASRLLA	VKSPQCSNPGLCVQP	103
HR7383A-63-184-TEV	NFIA	Q12857	EKPEVKQKWASRLLA	VSVKELDLYLAYFVH	122
HR7383A-7-165-TEV	NFIA	Q12857	LTQDEFHPFIEALLP	VKSPQCSNPGLCVQP	159
HR7279A-8-166-Av6HT	NFIB	O00712	LTQDEFHPFIEALLP	MKSPHCTNPALCVQP	159
HR7320A-61-170-TEV	NFIC	P08651	LLGEKPEVKQKWASR	QCGHPVLCVQPHHIG	110
HR7320A-61-186-TEV	NFIC	P08651	LLGEKPEVKQKWASR	AVKELDLYLAYFVRE	126
HR7320A-64-166-TEV	NFIC	P08651	EKPEVKQKWASRLLA	VKAAQCGHPVLCVQP	103
HR7320A-8-166-TEV	NFIC	P08651	LTQDEFHPFIEALLP	VKAAQCGHPVLCVQP	159
HR3633C-804-893-TEV	NFKB1	P19838	AQGDMKQLAEDVKLQ	MGYTEAIEVIQAASS	90
HR3633D-248-354-Av6HT	NFKB1	P19838	SNLKIVRMDRTAGCV	ETSEPKPFLYYPEIK	107
HR5561A-15	NFKB1	P19838	MQLVRDLLEVTSGLI	GADPLVENFEPLYDL	201
HR4541C-445-696-14	NFKB2	Q00653	MEYNARLFGLAQRSA	TLTRLLLKAGADIHA	253
HR4541C-477-696-14	NFKB2	Q00653	MQRHLLTAQDENGDT	TLTRLLLKAGADIHA	221
HR4541C-482-696-14	NFKB2	Q00653	MTAQDENGDTPLHLA	TLTRLLLKAGADIHA	216
HR4541D-37-329-TEV	NFKB2	Q00653	GPYLVIVEQPKQRGF	GDVSDSKQFTYYPLV	293
HR6920A-980-1063-NHT	NFX1	Q12986	SKFSDSLKEDARKDL	DSEPKRNVVVTAIRG	84
HR6427-1-270-14	NFYA	P23511	MEQYTANSNSSTEQI	GAEMLEEEPLYVNAK	270
HR6427-1-290-14	NFYA	P23511	MEQYTANSNSSTEQI	LKRRQARAKLEAEGK	290
HR6427-1-295-14	NFYA	P23511	MEQYTANSNSSTEQI	ARAKLEAEGKIPKER	295

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6427A-38-145-14	NFYA	P23511	EAQVASASGQQVQTL	QIIIQQPQTAVTAGQ	108
HR6427A-42-150-14	NFYA	P23511	ASASGQQVQTLQVVQ	QPQTAVTAGQTQTQQ	109
HR6427A-47-145-14	NFYA	P23511	QQVQTLQVVQGQPLM	QIIIQQPQTAVTAGQ	99
HR4613B-51-143-TEV	NFYB	P25208	SFREQDIYLPIANVA	SYVEPLKLYLQKFRE	93
HR3512B-166-288-15	NKRF	O15226	VVAEKQYFIEKLTAT	LQKRIEVRVVRRKFK	123
HR3512B-166-293-15	NKRF	O15226	VVAEKQYFIEKLTAT	EVRVVRRKFKHTFGE	128
HR3512B-188-288-15	NKRF	O15226	PEMTSGSDKINYTYM	LQKRIEVRVVRRKFK	101
HR4600-161-227-Av6HT	NKX2-1	P43699	RRKRRVLFSQAQVYE	RYKMKRQAKDKAAQQ	67
HR7114A-123-196-TEV	NKX2-2	O95096	GDAGKKRKRRVLFSK	KMKRARAEKGMEVTP	74
HR4758B-148-211-TEV	NKX2-3	Q8TAU0	RRKPRVLFSQAQVFE	QNRRYKCKRQRQDKS	64
HR7998A-189-255-TEV	NKX2-4	Q9H2Z4	RRKRRVLFSQAQVYE	RYKMKRQAKDKAAQQ	67
HR5518A-127-207-14	NKX2-5	P52952	MADNAERPRARRRRK	CKRQRQDQTLELVGL	82
HR5518A-127-207-Av6HT	NKX2-5	P52952	ADNAERPRARRRRKP	CKRQRQDQTLELVGL	81
HR5518A-127-207-TEV	NKX2-5	P52952	ADNAERPRARRRRKP	CKRQRQDQTLELVGL	81
HR5518A-14	NKX2-5	P52952	MRPRARRRRKPRVLF	DQTLELVGLPPPPPP	83
HR5518A-15	NKX2-5	P52952	MRPRARRRRKPRVLF	DQTLELVGLPPPPPP	83
HR5518B-128-224-14	NKX2-5	P52952	MDNAERPRARRRRKP	PPPPPARRIAVPVLV	98
HR5518B-128-233-14	NKX2-5	P52952	MDNAERPRARRRRKP	AVPVLVRDGKPCLGD	107
HR5518B-143-224-14	NKX2-5	P52952	MVLFSQAQVYELERR	PPPPPARRIAVPVLV	83
HR5518B-143-228-14	NKX2-5	P52952	MVLFSQAQVYELERR	PARRIAVPVLVRDGK	87
HR5518B-143-233-14	NKX2-5	P52952	MVLFSQAQVYELERR	AVPVLVRDGKPCLGD	92
HR7861A-83-143-TEV	NKX2-8	O15522	KRKKRRVLFSKAQTL	KIWFQNHRYKLKRAR	61

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6470A-127-195-15	NKX3-1	Q99801	SRAAFSHTQVIELER	RKQLSSELGDLEKHS	69
HR6470A-132-189-15	NKX3-1	Q99801	SHTQVIELERKFSHQ	RRYKTKRKQLSSELG	58
HR6470A-97-163-15	NKX3-1	Q99801	RHLGSYLLDSENTSG	LSAPERAHLAKNLKL	67
HR6470A-97-168-15	NKX3-1	Q99801	RHLGSYLLDSENTSG	RAHLAKNLKLTETQV	72
HR6470A-97-189-15	NKX3-1	Q99801	RHLGSYLLDSENTSG	RRYKTKRKQLSSELG	93
HR6470A-97-195-15	NKX3-1	Q99801	RHLGSYLLDSENTSG	RKQLSSELGDLEKHS	99
HR8303A-212-271-Av6HT	NKX3-2	P78367	AFSHAQVFELERRFN	RRYKTKRRQMAADLL	60
HR6930A-227-296-NHT	NKX6-1	P78426	SILLDKDGKRKHTRP	VWFQNRRTKWRKKHA	70
HR7948A-199-323-Av6HT	NOC3L	Q8WTT2	TIEEHLIERKKKLQE	LENLEQMVKDWKQRK	125
HR8350A-301-458-Av6HT	NOC4L	Q9BVI4	GGALSLLALNGLFIL	HYHPEVSKAASVINQ	158
HR6913A-130-201-NHT	NPAS1	Q99742	VSEVFEQHLGGHILQ	HPGDHSEVLEQLGLR	72
HR8115A-87-146-Av6HT	NPAS2	Q99743	QLMLEALDGFIIAVT	FLPEQEHSEVYKILS	60
HR7386A-142-213-NHT	NPAS3	Q8IXF0	AIEVFEAHLGSHILQ	HPGDHVEMAEQLGMK	72
HR7814A-184-329-NHT	NPAS4	Q8IUM7	GNPVFTAFCAPLEPR	SDMEAWSLRQQLNSE	146
HR7372A-210-470-15	NR0B1	P51843	MGSTLYCVPTSTNQA	VSMDDMMLEMLCTKI	262
HR7372A-210-470-Av6HT	NR0B1	P51843	GSTLYCVPTSTNQAQ	VSMDDMMLEMLCTKI	261
HR7372A-210-470-TEV	NR0B1	P51843	GSTLYCVPTSTNQAQ	VSMDDMMLEMLCTKI	261
HR7372A-237-470-15	NR0B1	P51843	MDTSSGALRPVALKS	VSMDDMMLEMLCTKI	235
HR7372A-237-470-Av6HT	NR0B1	P51843	DTSSGALRPVALKSP	VSMDDMMLEMLCTKI	234
HR7372A-237-470-TEV	NR0B1	P51843	DTSSGALRPVALKSP	VSMDDMMLEMLCTKI	234
HR7372A-245-470-15	NR0B1	P51843	MPVALKSPQVVCEAA	VSMDDMMLEMLCTKI	227
HR7372A-245-470-Av6HT	NR0B1	P51843	PVALKSPQVVCEAAS	VSMDDMMLEMLCTKI	226

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7372A-245-470-TEV	NR0B1	P51843	PVALKSPQVVCEAAS	VSMDDMMLEMLCTKI	226
HR8369A-14-257-15	NR0B2	Q15466	MAASRPAILYALLSS	DVDIAGLLGDMLLLR	245
HR8278A-123-216-15	NR1D1	P20393	MTKLNGMVLLCKVCG	AVRFGRIPKREKQRM	95
HR8278A-123-216-Av6HT	NR1D1	P20393	TKLNGMVLLCKVCGD	AVRFGRIPKREKQRM	94
HR8278A-123-216-TEV	NR1D1	P20393	TKLNGMVLLCKVCGD	AVRFGRIPKREKQRM	94
HR8341A-100-171-15	NR1D2	Q14995	MVLLCKVCGDVASGF	QQCRFKKCLSVGMSR	73
HR8341A-100-171-Av6HT	NR1D2	Q14995	VLLCKVCGDVASGFH	QQCRFKKCLSVGMSR	72
HR8341A-100-171-TEV	NR1D2	Q14995	VLLCKVCGDVASGFH	QQCRFKKCLSVGMSR	72
HR8341A-100-181-15	NR1D2	Q14995	MVLLCKVCGDVASGF	VGMSRDAVRFGRIPK	83
HR8341A-100-181-Av6HT	NR1D2	Q14995	VLLCKVCGDVASGFH	VGMSRDAVRFGRIPK	82
HR8341A-100-181-TEV	NR1D2	Q14995	VLLCKVCGDVASGFH	VGMSRDAVRFGRIPK	82
HR8341A-100-200-15	NR1D2	Q14995	MVLLCKVCGDVASGF	RMLIEMQSAMKTMMN	102
HR8341A-100-200-Av6HT	NR1D2	Q14995	VLLCKVCGDVASGFH	RMLIEMQSAMKTMMN	101
HR8341A-100-200-TEV	NR1D2	Q14995	VLLCKVCGDVASGFH	RMLIEMQSAMKTMMN	101
HR8341B-381-579-15	NR1D2	Q14995	MHLVCPMSKSPYVDP	NNMHSEELLAFKVHP	200
HR8341B-381-579-Av6HT	NR1D2	Q14995	HLVCPMSKSPYVDPH	NNMHSEELLAFKVHP	199
HR8341B-381-579-TEV	NR1D2	Q14995	HLVCPMSKSPYVDPH	NNMHSEELLAFKVHP	199
HR7370A-209-460-15	NR1H2	P55055	MSQGSGEGEGVQLTA	DKKLPPLLSEIWDVH	253
HR7370A-209-460-Av6HT	NR1H2	P55055	SQGSGEGEGVQLTAA	DKKLPPLLSEIWDVH	252
HR7370A-209-460-TEV	NR1H2	P55055	SQGSGEGEGVQLTAA	DKKLPPLLSEIWDVH	252
HR8107E-205-447-TEV	NR1H3	Q13133	QLSPEQLGMIEKLVA	KKLPPLLSEIWDVHE	243
HR4469B-125-217-14	NR1H4	Q96RI1	MGASAGRIKGDELCV	LAECMYTGLLTEIQC	94

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4469B-125-234-14	NR1H4	Q96RI1	MGASAGRIKGDELCV	KRLRKNVKQHADQTV	111
HR4469B-129-217-14	NR1H4	Q96RI1	MGRIKGDELCVVCGD	LAECMYTGLLTEIQC	90
HR4469B-129-234-14	NR1H4	Q96RI1	MGRIKGDELCVVCGD	KRLRKNVKQHADQTV	107
HR4469B-130-213-14	NR1H4	Q96RI1	MRIKGDELCVVCGDR	EMGMLAECMYTGLLT	85
HR4469B-130-229-14	NR1H4	Q96RI1	MRIKGDELCVVCGDR	IQCKSKRLRKNVKQH	101
HR4469B-134-213-14	NR1H4	Q96RI1	MDELCVVCGDRASGY	EMGMLAECMYTGLLT	81
HR4469B-134-229-14	NR1H4	Q96RI1	MDELCVVCGDRASGY	IQCKSKRLRKNVKQH	97
HR7870A-37-107-15	NR1I2	O75469	MGPQICRVCGDKATG	QCQACRLRKCLESGM	72
HR7870A-37-107-Av6HT	NR1I2	O75469	GPQICRVCGDKATGY	QCQACRLRKCLESGM	71
HR7870A-37-107-TEV	NR1I2	O75469	GPQICRVCGDKATGY	QCQACRLRKCLESGM	71
HR7870A-37-130-15	NR1I2	O75469	MGPQICRVCGDKATG	EAVEERRALIKRKKS	95
HR7870A-37-130-Av6HT	NR1I2	O75469	GPQICRVCGDKATGY	EAVEERRALIKRKKS	94
HR7870A-37-130-TEV	NR1I2	O75469	GPQICRVCGDKATGY	EAVEERRALIKRKKS	94
HR7870B-130-434-15	NR1I2	O75469	MSERTGTQPLGVQGL	FATPLMQELFGITGS	306
HR7870B-130-434-Av6HT	NR1I2	O75469	SERTGTQPLGVQGLT	FATPLMQELFGITGS	305
HR7870B-130-434-TEV	NR1I2	O75469	SERTGTQPLGVQGLT	FATPLMQELFGITGS	305
HR7870C-142-434-15	NR1I2	O75469	MGLTEEQRMMIRELM	FATPLMQELFGITGS	294
HR7870C-142-434-Av6HT	NR1I2	O75469	GLTEEQRMMIRELMD	FATPLMQELFGITGS	293
HR7870C-142-434-TEV	NR1I2	O75469	GLTEEQRMMIRELMD	FATPLMQELFGITGS	293
HR7475A-6-105-15	NR1I3	Q14994	MDELRNCVVCGDQAT	RAKQAQRRAQQTPVQ	101
HR7475A-6-105-Av6HT	NR1I3	Q14994	DELRNCVVCGDQATG	RAKQAQRRAQQTPVQ	100
HR7475A-6-105-TEV	NR1I3	Q14994	DELRNCVVCGDQATG	RAKQAQRRAQQTPVQ	100

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7475A-6-120-15	NR1I3	Q14994	MDELRNCVVCGDQAT	LSKEQEELIRTLLGA	116
HR7475A-6-120-Av6HT	NR1I3	Q14994	DELRNCVVCGDQATG	LSKEQEELIRTLLGA	115
HR7475A-6-120-TEV	NR1I3	Q14994	DELRNCVVCGDQATG	LSKEQEELIRTLLGA	115
HR7475B-6-77-15	NR1I3	Q14994	MDELRNCVVCGDQAT	CPACRLQKCLDAGMR	73
HR7475B-6-77-Av6HT	NR1I3	Q14994	DELRNCVVCGDQATG	CPACRLQKCLDAGMR	72
HR7475B-6-77-TEV	NR1I3	Q14994	DELRNCVVCGDQATG	CPACRLQKCLDAGMR	72
HR7475B-6-82-15	NR1I3	Q14994	MDELRNCVVCGDQAT	LQKCLDAGMRKDMIL	78
HR7475B-6-82-Av6HT	NR1I3	Q14994	DELRNCVVCGDQATG	LQKCLDAGMRKDMIL	77
HR7475B-6-82-TEV	NR1I3	Q14994	DELRNCVVCGDQATG	LQKCLDAGMRKDMIL	77
HR7475C-103-352-15	NR1I3	Q14994	MPVQLSKEQEELIRT	QGLSAMMPLLQEICS	251
HR7475C-103-352-Av6HT	NR1I3	Q14994	PVQLSKEQEELIRTL	QGLSAMMPLLQEICS	250
HR7475C-103-352-TEV	NR1I3	Q14994	PVQLSKEQEELIRTL	QGLSAMMPLLQEICS	250
HR8155A-108-196-Av6HT	NR2C1	P13056	KVFDLCVVCGDKASG	SVQCERKPIEVSREK	89
HR6956-16-584-15	NR2C2	P49116	MAVASPQRIQGSEPA	RLMSSNITEELFFTG	570
HR6956-16-584-Av6HT	NR2C2	P49116	AVASPQRIQGSEPAS	RLMSSNITEELFFTG	569
HR6956-16-584-TEV	NR2C2	P49116	AVASPQRIQGSEPAS	RLMSSNITEELFFTG	569
HR6956A-115-584-15	NR2C2	P49116	MSASVERLLGKTDVQ	RLMSSNITEELFFTG	471
HR6956A-115-584-TEV	NR2C2	P49116	SASVERLLGKTDVQR	RLMSSNITEELFFTG	470
HR6956B-170-584-15	NR2C2	P49116	MYSCRSNQDCIINKH	RLMSSNITEELFFTG	416
HR6956B-170-584-Av6HT	NR2C2	P49116	YSCRSNQDCIINKHH	RLMSSNITEELFFTG	415
HR6956B-170-584-TEV	NR2C2	P49116	YSCRSNQDCIINKHH	RLMSSNITEELFFTG	415
HR6956B-181-584-15	NR2C2	P49116	MNKHHRNRCQFCRLK	RLMSSNITEELFFTG	405

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6956B-181-584-Av6HT	NR2C2	P49116	NKHHRNRCQFCRLKK	RLMSSNITEELFFTG	404
HR6956B-181-584-TEV	NR2C2	P49116	NKHHRNRCQFCRLKK	RLMSSNITEELFFTG	404
HR6956B-218-584-15	NR2C2	P49116	MEKPSNCAASTEKIY	RLMSSNITEELFFTG	368
HR6956B-218-584-Av6HT	NR2C2	P49116	EKPSNCAASTEKIYI	RLMSSNITEELFFTG	367
HR6956B-218-584-TEV	NR2C2	P49116	EKPSNCAASTEKIYI	RLMSSNITEELFFTG	367
HR6956C-110-188-15	NR2C2	P49116	MQIVTDSASVERLLG	SNQDCIINKHHRNRC	80
HR6956C-110-205-15	NR2C2	P49116	MQIVTDSASVERLLG	CRLKKCLEMGMKMES	97
HR6956C-115-183-15	NR2C2	P49116	MSASVERLLGKTDVQ	TYSCRSNQDCIINKH	70
HR6956C-115-183-Av6HT	NR2C2	P49116	SASVERLLGKTDVQR	TYSCRSNQDCIINKH	69
HR6956C-115-183-TEV	NR2C2	P49116	SASVERLLGKTDVQR	TYSCRSNQDCIINKH	69
HR6956C-115-200-15	NR2C2	P49116	MSASVERLLGKTDVQ	NRCQFCRLKKCLEMG	87
HR6956C-115-200-Av6HT	NR2C2	P49116	SASVERLLGKTDVQR	NRCQFCRLKKCLEMG	86
HR6956C-115-200-TEV	NR2C2	P49116	SASVERLLGKTDVQR	NRCQFCRLKKCLEMG	86
HR7378A-18-385-15	NR2E1	Q9Y466	MVCGDRSSGKHYGVY	PITRLLSDMYKSSDI	369
HR7378A-18-385-Av6HT	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	PITRLLSDMYKSSDI	368
HR7378A-18-385-TEV	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	PITRLLSDMYKSSDI	368
HR7378B-134-385-Av6HT	NR2E1	Q9Y466	VTQLEPHGLELAAVS	PITRLLSDMYKSSDI	252
HR7378B-134-385-TEV	NR2E1	Q9Y466	VTQLEPHGLELAAVS	PITRLLSDMYKSSDI	252
HR7378B-65-385-15	NR2E1	Q9Y466	MTHRNQCRACRLKKC	PITRLLSDMYKSSDI	322
HR7378B-65-385-Av6HT	NR2E1	Q9Y466	THRNQCRACRLKKCL	PITRLLSDMYKSSDI	321
HR7378B-65-385-TEV	NR2E1	Q9Y466	THRNQCRACRLKKCL	PITRLLSDMYKSSDI	321
HR7378B-76-385-15	NR2E1	Q9Y466	MKKCLEVNMNKDAVQ	PITRLLSDMYKSSDI	311

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7378B-76-385-Av6HT	NR2E1	Q9Y466	KKCLEVNMNKDAVQH	PITRLLSDMYKSSDI	310
HR7378B-76-385-TEV	NR2E1	Q9Y466	KKCLEVNMNKDAVQH	PITRLLSDMYKSSDI	310
HR7378C-18-109-15	NR2E1	Q9Y466	MVCGDRSSGKHYGVY	RTSTIRKQVALYFRG	93
HR7378C-18-109-Av6HT	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	RTSTIRKQVALYFRG	92
HR7378C-18-109-TEV	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	RTSTIRKQVALYFRG	92
HR7378C-18-83-15	NR2E1	Q9Y466	MVCGDRSSGKHYGVY	QCRACRLKKCLEVNM	67
HR7378C-18-83-Av6HT	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	QCRACRLKKCLEVNM	66
HR7378C-18-83-TEV	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	QCRACRLKKCLEVNM	66
HR7378C-18-96-15	NR2E1	Q9Y466	MVCGDRSSGKHYGVY	NMNKDAVQHERGPRT	80
HR7378C-18-96-Av6HT	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	NMNKDAVQHERGPRT	79
HR7378C-18-96-TEV	NR2E1	Q9Y466	VCGDRSSGKHYGVYA	NMNKDAVQHERGPRT	79
HR7906-Av6HT	NR2E3	Q9Y5X4	ETRPTALMSSTVAAA	NTPMEKLLCDMFKN*	410
HR7906B-101-410-Av6HT	NR2E3	Q9Y5X4	QACRLKKCLQAGMNQ	GNTPMEKLLCDMFKN	310
HR7906B-101-410-TEV	NR2E3	Q9Y5X4	QACRLKKCLQAGMNQ	GNTPMEKLLCDMFKN	310
HR7906B-114-410-15	NR2E3	Q9Y5X4	MNQDAVQNERQPRST	GNTPMEKLLCDMFKN	298
HR7906B-114-410-Av6HT	NR2E3	Q9Y5X4	NQDAVQNERQPRSTA	GNTPMEKLLCDMFKN	297
HR7906B-114-410-TEV	NR2E3	Q9Y5X4	NQDAVQNERQPRSTA	GNTPMEKLLCDMFKN	297
HR7906B-164-410-Av6HT	NR2E3	Q9Y5X4	SAARALGHHFMASLI	GNTPMEKLLCDMFKN	247
HR7906C-45-119-15	NR2E3	Q9Y5X4	MLQCRVCGDSSSGKH	LKKCLQAGMNQDAVQ	76
HR7906C-45-131-15	NR2E3	Q9Y5X4	MLQCRVCGDSSSGKH	AVQNERQPRSTAQVH	88
HR7906C-45-142-15	NR2E3	Q9Y5X4	MLQCRVCGDSSSGKH	AQVHLDSMESNTESR	99
HR7906C-50-114-15	NR2E3	Q9Y5X4	MCGDSSSGKHYGIYA	CQACRLKKCLQAGMN	66

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7906C-50-126-15	NR2E3	Q9Y5X4	MCGDSSSGKHYGIYA	GMNQDAVQNERQPRS	78
HR7906C-50-126-Av6HT	NR2E3	Q9Y5X4	CGDSSSGKHYGIYAC	GMNQDAVQNERQPRS	77
HR7906C-50-126-TEV	NR2E3	Q9Y5X4	CGDSSSGKHYGIYAC	GMNQDAVQNERQPRS	77
HR7906C-50-135-Av6HT	NR2E3	Q9Y5X4	CGDSSSGKHYGIYAC	ERQPRSTAQVHLDSM	86
HR7906C-50-135-TEV	NR2E3	Q9Y5X4	CGDSSSGKHYGIYAC	ERQPRSTAQVHLDSM	86
HR7906D-36-410-Av6HT	NR2E3	Q9Y5X4	EDPTGVSPSLQCRVC	GNTPMEKLLCDMFKN	375
HR3061D-77-164-TEV	NR2F1	P10589	SGQSQQHIECVVCGD	GMRREAVQRGRMPPT	88
HR6377A-77-157-TEV	NR2F2	P24468	IECVVCGDKSSGKHY	GMRREAVQRGRMPPT	81
HR7636A-56-394-Av6HT	NR2F6	P10588	CVVCGDKSSGKHYGV	TPIETLIRDMLLSGS	339
HR7636A-56-394-TEV	NR2F6	P10588	CVVCGDKSSGKHYGV	TPIETLIRDMLLSGS	339
HR7636B-113-394-15	NR2F6	P10588	MLKKCFRVGMRKEAV	TPIETLIRDMLLSGS	283
HR7636B-113-394-Av6HT	NR2F6	P10588	LKKCFRVGMRKEAVQ	TPIETLIRDMLLSGS	282
HR7636B-113-394-TEV	NR2F6	P10588	LKKCFRVGMRKEAVQ	TPIETLIRDMLLSGS	282
HR7636B-159-394-Av6HT	NR2F6	P10588	DLFPGQPVSELIAQL	TPIETLIRDMLLSGS	236
HR7636B-159-394-TEV	NR2F6	P10588	DLFPGQPVSELIAQL	TPIETLIRDMLLSGS	236
HR7636C-56-133-15	NR2F6	P10588	MCVVCGDKSSGKHYG	VGMRKEAVQRGRIPH	79
HR4533-15	NR3C1	P04150	MDSKESLTPGREENP	KYSNGNIKKLLFHQK	777
HR7785A-601-673-15	NR3C2	P08235	MKICLVCGDEASGCH	RLQKCLQAGMNLGAR	74
HR7785A-601-673-Av6HT	NR3C2	P08235	KICLVCGDEASGCHY	RLQKCLQAGMNLGAR	73
HR7785A-601-673-TEV	NR3C2	P08235	KICLVCGDEASGCHY	RLQKCLQAGMNLGAR	73
HR7785A-601-685-15	NR3C2	P08235	MKICLVCGDEASGCH	GARKSKKLGKLKGIH	86
HR7785A-601-685-Av6HT	NR3C2	P08235	KICLVCGDEASGCHY	GARKSKKLGKLKGIH	85

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7785A-601-685-TEV	NR3C2	P08235	KICLVCGDEASGCHY	GARKSKKLGKLKGIH	85
HR7785B-712-984-15	NR3C2	P08235	MAPAKEPSVNTALVP	KVESGNAKPLYFHRK	274
HR7785B-712-984-Av6HT	NR3C2	P08235	APAKEPSVNTALVPQ	KVESGNAKPLYFHRK	273
HR7785B-712-984-TEV	NR3C2	P08235	APAKEPSVNTALVPQ	KVESGNAKPLYFHRK	273
HR7785C-731-984-Av6HT	NR3C2	P08235	SRALTPSPVMVLENI	KVESGNAKPLYFHRK	254
HR7785C-731-984-TEV	NR3C2	P08235	SRALTPSPVMVLENI	KVESGNAKPLYFHRK	254
HR4793B-265-353-TEV	NR4A1	P22736	GRCAVCGDNASCQHY	TDSLKGRRGRLPSKP	89
HR8241A-261-342-15	NR4A2	P43354	MGLCAVCGDNAACQH	MVKEVVRTDSLKGRR	83
HR8241A-261-342-Av6HT	NR4A2	P43354	GLCAVCGDNAACQHY	MVKEVVRTDSLKGRR	82
HR8241A-261-342-TEV	NR4A2	P43354	GLCAVCGDNAACQHY	MVKEVVRTDSLKGRR	82
HR8241A-264-328-15	NR4A2	P43354	MAVCGDNAACQHYGV	RCQYCRFQKCLAVGM	66
HR8241A-264-328-Av6HT	NR4A2	P43354	AVCGDNAACQHYGVR	RCQYCRFQKCLAVGM	65
HR8241A-264-328-TEV	NR4A2	P43354	AVCGDNAACQHYGVR	RCQYCRFQKCLAVGM	65
HR8241B-328-598-15	NR4A2	P43354	MVKEVVRTDSLKGRR	PPAIIDKLFLDTLPF	271
HR8241B-328-598-Av6HT	NR4A2	P43354	VKEVVRTDSLKGRRG	PPAIIDKLFLDTLPF	270
HR8241B-328-598-TEV	NR4A2	P43354	VKEVVRTDSLKGRRG	PPAIIDKLFLDTLPF	270
HR7224A-291-626-15	NR4A3	Q92570	MTCAVCGDNAACQHY	PPSIIDKLFLDTLPF	337
HR7224B-363-626-15	NR4A3	Q92570	MRTDSLKGRRGRLPS	PPSIIDKLFLDTLPF	265
HR7224B-363-626-Av6HT	NR4A3	Q92570	RTDSLKGRRGRLPSK	PPSIIDKLFLDTLPF	264
HR7224B-363-626-TEV	NR4A3	Q92570	RTDSLKGRRGRLPSK	PPSIIDKLFLDTLPF	264
HR7224B-396-626-15	NR4A3	Q92570	MICMMNALVRALTDS	PPSIIDKLFLDTLPF	232
HR7224B-396-626-Av6HT	NR4A3	Q92570	ICMMNALVRALTDST	PPSIIDKLFLDTLPF	231

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7224B-396-626-TEV	NR4A3	Q92570	ICMMNALVRALTDST	PPSIIDKLFLDTLPF	231
HR7224B-411-626-15	NR4A3	Q92570	MPRDLDYSRYCPTDQ	PPSIIDKLFLDTLPF	217
HR7224B-411-626-Av6HT	NR4A3	Q92570	PRDLDYSRYCPTDQA	PPSIIDKLFLDTLPF	216
HR7224B-411-626-TEV	NR4A3	Q92570	PRDLDYSRYCPTDQA	PPSIIDKLFLDTLPF	216
HR7224C-291-374-15	NR4A3	Q92570	MTCAVCGDNAACQHY	EVVRTDSLKGRRGRL	85
HR7224C-291-374-Av6HT	NR4A3	Q92570	TCAVCGDNAACQHYG	EVVRTDSLKGRRGRL	84
HR7224C-291-374-TEV	NR4A3	Q92570	TCAVCGDNAACQHYG	EVVRTDSLKGRRGRL	84
HR7224C-294-356-15	NR4A3	Q92570	MVCGDNAACQHYGVR	NRCQYCRFQKCLSVG	64
HR7224C-294-356-Av6HT	NR4A3	Q92570	VCGDNAACQHYGVRT	NRCQYCRFQKCLSVG	63
HR7224C-294-356-TEV	NR4A3	Q92570	VCGDNAACQHYGVRT	NRCQYCRFQKCLSVG	63
HR7993A-220-461-15	NR5A1	Q13285	MGPNVPELILQLLQL	PRNNLLIEMLQAKQT	243
HR7993A-220-461-Av6HT	NR5A1	Q13285	GPNVPELILQLLQLE	PRNNLLIEMLQAKQT	242
HR7993A-220-461-TEV	NR5A1	Q13285	GPNVPELILQLLQLE	PRNNLLIEMLQAKQT	242
HR7993B-10-111-15	NR5A1	Q13285	MDELCPVCGDKVSGY	PMYKRDRALKQQKKA	103
HR7993B-10-111-Av6HT	NR5A1	Q13285	DELCPVCGDKVSGYH	PMYKRDRALKQQKKA	102
HR7993B-10-111-TEV	NR5A1	Q13285	DELCPVCGDKVSGYH	PMYKRDRALKQQKKA	102
HR8211A-79-187-Av6HT	NR5A2	O00482	MDEDLEELCPVCGDK	KRDRALKQQKKALIR	110
HR8211A-79-187-NHT	NR5A2	O00482	DEDLEELCPVCGDKV	KRDRALKQQKKALIR	109
HR8211A-79-187-TEV	NR5A2	O00482	DEDLEELCPVCGDKV	KRDRALKQQKKALIR	109
HR7049A-49-474-15	NR6A1	Q15406	MDRAEQRTCLICGDR	LFKVVLHSCKTSVGK	427
HR7049A-49-474-Av6HT	NR6A1	Q15406	DRAEQRTCLICGDRA	LFKVVLHSCKTSVGK	426
HR7049A-49-474-TEV	NR6A1	Q15406	DRAEQRTCLICGDRA	LFKVVLHSCKTSVGK	426

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7049B-117-474-15	NR6A1	Q15406	MLQMGMNRKAIREDG	LFKVVLHSCKTSVGK	359
HR7049B-117-474-Av6HT	NR6A1	Q15406	LQMGMNRKAIREDGM	LFKVVLHSCKTSVGK	358
HR7049B-117-474-TEV	NR6A1	Q15406	LQMGMNRKAIREDGM	LFKVVLHSCKTSVGK	358
HR7049C-49-143-15	NR6A1	Q15406	MDRAEQRTCLICGDR	DGMPGGRNKSIGPVQ	96
HR7049C-49-143-Av6HT	NR6A1	Q15406	DRAEQRTCLICGDRA	DGMPGGRNKSIGPVQ	95
HR7049C-49-143-TEV	NR6A1	Q15406	DRAEQRTCLICGDRA	DGMPGGRNKSIGPVQ	95
HR7049C-49-159-15	NR6A1	Q15406	MDRAEQRTCLICGDR	SEEEIERIMSGQEFE	112
HR7049C-49-159-Av6HT	NR6A1	Q15406	DRAEQRTCLICGDRA	SEEEIERIMSGQEFE	111
HR7049C-49-159-TEV	NR6A1	Q15406	DRAEQRTCLICGDRA	SEEEIERIMSGQEFE	111
HR7049C-58-143-15	NR6A1	Q15406	MICGDRATGLHYGII	DGMPGGRNKSIGPVQ	87
HR7049C-58-159-15	NR6A1	Q15406	MICGDRATGLHYGII	SEEEIERIMSGQEFE	103
HR7049C-58-159-Av6HT	NR6A1	Q15406	ICGDRATGLHYGIIS	SEEEIERIMSGQEFE	102
HR7049C-58-159-TEV	NR6A1	Q15406	ICGDRATGLHYGIIS	SEEEIERIMSGQEFE	102
HR8346A-59-490-Av6HT	NRF1	Q16656	LNSTAADEVTAHLAA	AMAPVTTRSDSAVT	432
HR7765A-130-171-Av6HT	NRL	P54845	ERFSDAALVSMSVRE	ALRLKQRRRTLKNRG	42
HR8036A-96-176-Av6HT	OLIG1	Q8TAK6	PDAKEEQQQQLRRKI	LLLGSSLQELRRALG	81
HR7010A-102-190-15	OLIG2	Q13516	MTEPELQQLRLKINS	IYGGHHAGFHPSACG	90
HR7010A-136-190-15	OLIG2	Q13516	MPYAHGPSVRKLSKI	IYGGHHAGFHPSACG	56
HR7010A-97-191-15	OLIG2	Q13516	MDKKQMTEPELQQLR	YGGHHAGFHPSACGG	96
HR6912A-76-168-NHT	OLIG3	Q7RTU3	LSEQDLQQLRLKING	GHHSAFHCGTVGHSA	93
HR4667C-291-437-TEV	ONECUT1	Q9UBC0	EINTKEVAQRITTEL	GLELSTVSNFFMNAR	147
HR8108A-313-459-TEV	ONECUT2	O95948	ERPPSSSSGSQVATS	FKENKRPSKEMQITI	147

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7555A-320-466-TEV	ONECUT3	O60422	EINTKEVAQRITAEL	GLELNTVSNFFMNAR	147
HR8321A-165-256-TEV	OSR1	Q8TAX0	GRLPSKTKKEFVCKF	QSRTLAVHKTLHSQV	92
HR6892A-160-254-TEV	OSR2	Q8N2R0	SRGRLPSKTKKEFIC	SRTLAVHKTLHMQES	95
HR7032A-39-109-TEV	OTX1	P32242	RRERTTFTRSQLDVL	QQQQSGSGTKSRPAK	71
HR7869A-39-106-TEV	OTX2	P32243	PAATPRKQRRERTTF	VWFKNRRAKCRQQQQ	68
HR8136A-96-170-Av6HT	OVOL1	O14753	RDHGFLRTKMKVTLG	NDTFDLKRHVRTHTG	75
HR8149A-102-172-Av6HT	OVOL2	Q9BRP0	ARSKIKFTTGTCSDS	DTFDLKRHVRTHTGI	71
HR8517-1-166-Av6HT	OVOL3	O00110	PRAFLVRSRRPQPPN	YRERREKLHVCEDCG	165
HR6980A-489-684-TEV	PARP12	Q9H0J9	DSSALPDPGFQKITL	YPEYVIQYTTSSKPS	196
HR8222A-355-438-TEV	PATZ1	Q9HBE1	VACEICGKIFRDVYH	RPDHLNGHIKQVHTS	84
HR7455A-96-228-NHT	PAX1	P15863	EQTYGEVNQLGGVFV	VSSISRILRNKIGSL	133
HR7856A-1-149-TEV	PAX5	Q02548	DLEKNYPTPRTSRTG	INRIIRTKVQQPPNQ	148
HR8074A-4-136-TEV	PAX6	P26367	SHSGVNQLGGVFVNG	SINRVLRNLASEKQQ	133
HR7676A-217-276-TEV	PAX7	P23759	QRRSRTTFTAEQLEE	QVWFSNRRARWRKQA	60
HR7297A-1-146-TEV	PAX8	Q06710	PHNSIRSGHGGLNQL	IRTKVQQPFNLPMDS	145
HR7882A-388-494-TEV	PBRM1	Q86U86	YYQQIKMPISLQQIR	RKSKKNIRKQRMKIL	107
HR7526A-233-295-TEV	PBX1	P40424	ARRKRRNFNKQATEI	SNWFGNKRIRYKKNI	63
HR7154A-244-306-TEV	PBX2	P40425	ARRKRRNFSKQATEV	SNWFGNKRIRYKKNI	63
HR7892A-235-297-Av6HT	PBX3	P40426	KTAVTAAHAVAAAVQ	NGDSYQGSQVGANVQ	63
HR7892A-235-297-TEV	PBX3	P40426	KTAVTAAHAVAAAVQ	NGDSYQGSQVGANVQ	63
HR7406A-210-272-Av6HT	PBX4	Q9BYU1	MARRKRRNFSKQATE	SNWFGNKRIRYKKNM	64
HR7406A-210-272-TEV	PBX4	Q9BYU1	ARRKRRNFSKQATEV	SNWFGNKRIRYKKNM	63

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7140A-124-182-TEV	PCGF6	Q9BYE7	NLSELTPYILCSICK	RCPKCNIVVHQTQPL	59
HR7140B-130-350-TEV	PCGF6	Q9BYE7	PYILCSICKGYLIDA	LVLHYGLVVSPLKIT	221
HR7140B-134-350-TEV	PCGF6	Q9BYE7	CSICKGYLIDATTIT	LVLHYGLVVSPLKIT	217
HR7140B-143-350-TEV	PCGF6	Q9BYE7	DATTITECLHTFCKS	LVLHYGLVVSPLKIT	208
HR7140B-182-350-TEV	PCGF6	Q9BYE7	LYNIRLDRQLQDIVY	LVLHYGLVVSPLKIT	169
HT6303A-249-350-Av6HT	PCGF6	Q9BYE7	IPPELDMSLLLEFIG	GLLVLHYGLVVSPLK	102
HT6303A-249-350-TEV	PCGF6	Q9BYE7	IPPELDMSLLLEFIG	GLLVLHYGLVVSPLK	102
HR7628A-146-206-TEV	PDX1	P52945	NKRTRTAYTRAQLLE	IWFQNRRMKWKKEED	61
HR4675D-563-641-TEV	PGR	P06401	PQKICLICGDEASGC	CCQAGMVLGGRKFKK	79
HR6832-1-194-15	PHB2	Q99623	MAQNLKDLAGRLPAG	DDVAITELSFSREYT	194
HR6832-1-194-Av6HT	PHB2	Q99623	AQNLKDLAGRLPAGP	DDVAITELSFSREYT	193
HR6832-1-291-15	PHB2	Q99623	MAQNLKDLAGRLPAG	NLVLNLQDESFTRGS	291
HR6832-1-291-Av6HT	PHB2	Q99623	AQNLKDLAGRLPAGP	NLVLNLQDESFTRGS	290
HR6832-15	PHB2	Q99623	MAQNLKDLAGRLPAG	SFTRGSDSLIKGKK*	300
HR6832-33-299-15	PHB2	Q99623	MAYGVRESVFTVEGG	ESFTRGSDSLIKGKK	268
HR6832-33-299-Av6HT	PHB2	Q99623	AYGVRESVFTVEGGH	ESFTRGSDSLIKGKK	267
HR6832-38-299-15	PHB2	Q99623	MESVFTVEGGHRAIF	ESFTRGSDSLIKGKK	263
HR6832-38-299-Av6HT	PHB2	Q99623	ESVFTVEGGHRAIFF	ESFTRGSDSLIKGKK	262
HR6832-Av6HT	PHB2	Q99623	AQNLKDLAGRLPAGP	SFTRGSDSLIKGKK*	299
HR6832A-33-194-15	PHB2	Q99623	MAYGVRESVFTVEGG	DDVAITELSFSREYT	163
HR6832A-33-194-Av6HT	PHB2	Q99623	AYGVRESVFTVEGGH	DDVAITELSFSREYT	162
HR6832A-33-207-15	PHB2	Q99623	MAYGVRESVFTVEGG	YTAAVEAKQVAQQEA	176

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6832A-33-207-Av6HT	PHB2	Q99623	AYGVRESVFTVEGGH	YTAAVEAKQVAQQEA	175
HR6832A-33-207-NHT	PHB2	Q99623	AYGVRESVFTVEGGH	YTAAVEAKQVAQQEA	175
HR6832A-72-194-15	PHB2	Q99623	MIPWFQYPIIYDIRA	DDVAITELSFSREYT	124
HR6832A-72-194-Av6HT	PHB2	Q99623	IPWFQYPIIYDIRAR	DDVAITELSFSREYT	123
HR6832A-72-207-15	PHB2	Q99623	MIPWFQYPIIYDIRA	YTAAVEAKQVAQQEA	137
HR6832A-72-207-Av6HT	PHB2	Q99623	IPWFQYPIIYDIRAR	YTAAVEAKQVAQQEA	136
HR7710A-641-719-NHT	PHF20	Q9BVI0	ELDGDDRYDFEVVRC	PGFKYWYDKEWLSRG	79
HR6973A-486-543-TEV	PHF21A	Q96BD5	IHEDFCSVCRKSGQL	CPRCQDQMLKKEEAI	58
HR6412A-62-149-14	PHOX2A	O14813	LRDHQPAPYSAVPYK	QVWFQNRRAKFRKQE	88
HR6412A-67-144-14	PHOX2A	O14813	PAPYSAVPYKFFPEP	TEARVQVWFQNRRAK	78
HR6412B-71-149-14	PHOX2A	O14813	SAVPYKFFPEPSGLH	QVWFQNRRAKFRKQE	79
HR6412B-76-144-14	PHOX2A	O14813	KFFPEPSGLHEKRKQ	TEARVQVWFQNRRAK	69
HR7334A-91-148-Av6HT	PHOX2B	Q99453	GLNEKRKQRRIRTTF	KIDLTEARVQVWFQN	58
HR8156A-1-65-TEV	PIAS1	O75925	ADSAELKQMVMSLRV	PAVQMKIKELYRRRF	64
HR7952-96-424-Av6HT	PIAS2	O75928	LAVAGIHSLPSTSVT	CSDVDEIKFQEDGSW	329
HR3483-121-628-14	PIAS3	Q9Y6X2	MQPVHPDVTMKPLPF	GPSLTGCRSDIISLD	509
HR3483-126-628-14	PIAS3	Q9Y6X2	MDVTMKPLPFYEVYG	GPSLTGCRSDIISLD	504
HR3042-1-425-14	PIAS4	Q8N2W9	MAAELVEAKNMVMSF	KERSCSPQGAILVLG	425
HR3042A-1-55-14	PIAS4	Q8N2W9	MAAELVEAKNMVMSF	RALQLVQFDCSPELF	55
HR3042A-1-68-14	PIAS4	Q8N2W9	MAAELVEAKNMVMSF	LFKKIKELYETRYAK	68
HR3042A-1-73-14	PIAS4	Q8N2W9	MAAELVEAKNMVMSF	KELYETRYAKKNSEP	73
HR7108A-119-222-Av6HT	PIKFYVE	Q9Y2I7	GHDPRTAVQLRSLST	RACTYCRKIALSYAH	104

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7108A-119-241-Av6HT	PIKFYVE	Q9Y2I7	GHDPRTAVQLRSLST	NSIGEDLNALSDSAC	123
HR7108A-143-222-Av6HT	PIKFYVE	Q9Y2I7	EGKSQDSDLKQYWMP	RACTYCRKIALSYAH	80
HR7108A-143-241-Av6HT	PIKFYVE	Q9Y2I7	EGKSQDSDLKQYWMP	NSIGEDLNALSDSAC	99
HR7108A-150-222-Av6HT	PIKFYVE	Q9Y2I7	DLKQYWMPDSQCKEC	RACTYCRKIALSYAH	73
HR7108A-150-241-NHT	PIKFYVE	Q9Y2I7	DLKQYWMPDSQCKEC	NSIGEDLNALSDSAC	92
HR7108B-586-943-Av6HT	PIKFYVE	Q9Y2I7	GWHHNNLELLREENG	LLELRIVFEKGEQEN	358
HR7108B-610-943-Av6HT	PIKFYVE	Q9Y2I7	SANHNHMMALLQQLL	LLELRIVFEKGEQEN	334
HR7108C-1761-2058-Av6HT	PIKFYVE	Q9Y2I7	KRETLRGADSAYYQV	DKKLEMVVKSTGILG	298
HR7108C-1761-2088-Av6HT	PIKFYVE	Q9Y2I7	KRETLRGADSAYYQV	TRFCEAMDKYFLMVP	328
HR7108C-1807-2058-Av6HT	PIKFYVE	Q9Y2I7	PHVELQFSDANAKFY	DKKLEMVVKSTGILG	252
HR7108C-1807-2088-Av6HT	PIKFYVE	Q9Y2I7	PHVELQFSDANAKFY	TRFCEAMDKYFLMVP	282
HR7108D-342-488-Av6HT	PIKFYVE	Q9Y2I7	TEDERKILLDSVQLK	DSDTEQIAEEGDDNL	147
HR7108D-353-488-Av6HT	PIKFYVE	Q9Y2I7	VQLKDLWKKICHHSS	DSDTEQIAEEGDDNL	136
HR8214A-89-149-TEV	PITX1	P78337	QRRQRTHFTSQQLQE	VWFKNRRAKWRKRER	61
HR4722B-274-317-Av6HT	PITX2	Q99697	RDTCNSSLASLRLKA	ASNLSACQYAVDRPV	44
HR7801A-85-145-TEV	PITX3	O75364	RYPDMSTREEIAVWT	GSFAAPLGGLVPPYE	61
HR7268A-257-319-TEV	PKNOX1	P55347	GSSKNKRGVLPKHAT	QVNNWFINARRRILQ	63
HR7428A-289-348-TEV	PKNOX2	Q96KN3	KNKRGVLPKHATNIM	QVNNWFINARRRILQ	60
HR7109-1-351-15	PLAG1	Q6DJT9	MATVIPGDLSEVRDT	SSTSYAISIPEKEQP	351
HR7109-1-356-15	PLAG1	Q6DJT9	MATVIPGDLSEVRDT	AISIPEKEQPLKGEI	356
HR7109-1-368-15	PLAG1	Q6DJT9	MATVIPGDLSEVRDT	GEIESYLMELQGGVP	368
HR7109-1-436-15	PLAG1	Q6DJT9	MATVIPGDLSEVRDT	FNFIPLNGPPYNPLS	436

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7109-1-441-15	PLAG1	Q6DJT9	MATVIPGDLSEVRDT	LNGPPYNPLSVGSLG	441
HR7109A-70-128-Av6HT	PLAG1	Q6DJT9	TKAFVSKYKLQRHMA	HDPNKETFKCEECGK	59
HR7109A-70-163-Av6HT	PLAG1	Q6DJT9	TKAFVSKYKLQRHMA	DLTCKVCLQTFESTG	94
HR7109A-72-107-Av6HT	PLAG1	Q6DJT9	AFVSKYKLQRHMATH	KCNYCEKMFHRKDHL	36
HR7109B-36-237-Av6HT	PLAG1	Q6DJT9	CQLCDKAFNSVEKLK	GRKDHLTRHMKKSHN	202
HR7109C-119-174-Av6HT	PLAG1	Q6DJT9	ETFKCEECGKNYNTK	ESTGVLLEHLKSHAG	56
HR7109C-122-170-Av6HT	PLAG1	Q6DJT9	KCEECGKNYNTKLGF	LQTFESTGVLLEHLK	49
HR7109D-183-233-Av6HT	PLAG1	Q6DJT9	KKHQCEHCDRRFYTR	AQRFGRKDHLTRHMK	51
HR7109D-183-237-Av6HT	PLAG1	Q6DJT9	KKHQCEHCDRRFYTR	GRKDHLTRHMKKSHN	55
HR7109D-188-233-Av6HT	PLAG1	Q6DJT9	EHCDRRFYTRKDVRR	AQRFGRKDHLTRHMK	46
HR7895-1-337-Av6HT	PLAGL1	Q9UM63	ATFPCQLCGKTFLTL	KGFCNISLFEDLPLQ	336
HR7895-1-397-Av6HT	PLAGL1	Q9UM63	ATFPCQLCGKTFLTL	LLGFWQLPPPATQNT	396
HR7895-159-463-Av6HT	PLAGL1	Q9UM63	DHCERCFYTRKDVRR	GTGSAILPHFHHAFR	305
HR7895-Av6HT	PLAGL1	Q9UM63	ATFPCQLCGKTFLTL	TGSAILPHFHHAFR*	463
HR7895A-149-212-15	PLAGL1	Q9UM63	MSGTKEKKHQCDHCE	HLTRHTKKTHSQELM	65
HR7895A-154-207-15	PLAGL1	Q9UM63	MKKHQCDHCERCFYT	FGRKDHLTRHTKKTH	55
HR7895A-159-199-Av6HT	PLAGL1	Q9UM63	DHCERCFYTRKDVRR	LCQFCAQRFGRKDHL	41
HR7895B-1-209-Av6HT	PLAGL1	Q9UM63	ATFPCQLCGKTFLTL	RKDHLTRHTKKTHSQ	208
HR7895C-1-69-Av6HT	PLAGL1	Q9UM63	ATFPCQLCGKTFLTL	THSPQKSHQCAHCEK	68
HR7895C-12-69-Av6HT	PLAGL1	Q9UM63	TFLTLEKFTIHNYSH	THSPQKSHQCAHCEK	58
HR7996A-189-243-Av6HT	PLAGL2	Q9UPG8	KKHPCDHCDRRFYTR	GRKDHLTRHVKKSHS	55
HR8052A-121-223-TEV	PLEK	P08567	PETIDLGALYLSMKD	FLDNPDAFYYFPDSG	103

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7739A-238-353-TEV	PLEK2	Q9NYT0	SLSTVELSGTVVKQG	KAERAEWIEAIKKLT	116
HR7495A-45-167-TEV	PLEKHA4	Q9H4M7	NALRRDPNLPVHIRG	RAEGDDYGQPRSPAR	123
HR8545A-1266-1889-Av6HT	PLXNA1	Q9UIW2	AYKRKSRDADRTLKR	QARRQRLRSKLEQVV	624
HR7225A-1259-1890-TEV	PLXNA2	O75051	AYKRKSRENDLTLKR	RQRLAYKVEQLINAM	632
HR8315A-1241-1867-TEV	PLXNA3	P51805	AYKRKTQDADRTLKR	KHKLRQKLEQIISLV	627
HR7815A-1736-1862-Av6HT	PLXNB1	O43157	DNRLLREDVEYRPLT	ALVPCLTKHVLRENQ	127
HR7815A-1736-1862-TEV	PLXNB1	O43157	DNRLLREDVEYRPLT	ALVPCLTKHVLRENQ	127
HR8081A-1224-1827-NHT	PLXNB2	O15031	SQQAEREYEKIKSQL	PAAQKMQLAFRLQQI	604
HR8081A-1224-1827-TEV	PLXNB2	O15031	SQQAEREYEKIKSQL	PAAQKMQLAFRLQQI	604
HR6985A-1280-1902-TEV	PLXNB3	Q9ULL4	GVGMGAAVLIAAVLL	LYNHIHRYYDQIISA	623
HR7941A-1198-1305-TEV	PLXNC1	O60486	TVALNVVFEKIPENE	HYEISNGSTIKVFKK	108
HR8405A-1553-1678-NHT	PLXND1	Q9Y4D7	AKPRNLNVSFQGCGM	RVKDLDTEKYFHLVL	126
HR6974A-571-650-TEV	PMS1	P54277	IKKPMSASALFVQDH	KRAIEQESQMSLKDG	80
HR6952A-34-103-NHT	POGK	Q9P215	QKVRICSEGGWVPAL	PFPKPDMITRLEGEE	70
HR7570A-37-117-NHT	POLE4	Q9NR33	RLSRLPLARVKALVK	IEAVDEFAFLEGTLD	81
HR7808-15	POLR2L	P62875	MIIPVRCFTCGKIVG	DLIEKLLNYAPLK*	68
HR8543A-1-154-Av6HT	POTEKP	Q9BYX7	DDDTAVLVIDNGSGM	LSLYTSGRTTGIVMD	153
HR4466B-129-273-TEV	POU1F1	P28069	IRELEKFANEFKVRR	RVWFCNRRQREKRVK	145
HR7752A-297-359-TEV	POU2F2	P09086	EKSFLANQKPTSEEI	APMLPSPGKPASYSP	63
HR7822A-187-257-TEV	POU2F3	Q9UKI9	DLEELEKFAKTFKQR	CKLKPLLEKWLNDAE	71
HR7177A-250-322-Av6HT	POU3F1	Q03052	PSSDDLEQFAKQFKQ	KLKPLLNKWLEETDS	73
HR6946A-348-418-15	POU3F2	P20265	MIAAQGRKRKKRTSI	NRRQKEKRMTPPGGT	72

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6946A-353-407-Av6HT	POU3F2	P20265	RKRKKRTSIEVSVKG	LEKEVVRVWFCNRRQ	55
HR6946A-353-414-15	POU3F2	P20265	MRKRKKRTSIEVSVK	VWFCNRRQKEKRMTP	68
HR6946A-356-432-15	POU3F2	P20265	MKKRTSIEVSVKGAL	TLPGAEDVYGGSRDT	78
HR6946A-361-427-15	POU3F2	P20265	MIEVSVKGALESHFL	TPPGGTLPGAEDVYG	68
HR6946A-377-427-15	POU3F2	P20265	MPKPSAQEITSLADS	TPPGGTLPGAEDVYG	52
HR8066A-320-388-TEV	POU3F3	P20264	DDLEQFAKQFKQRRI	CKLKPLLNKWLEEAD	69
HR8200A-192-260-TEV	POU3F4	P49335	DELEQFAKQFKQRRI	CKLKPLLNKWLEEAD	69
HR7341A-253-330-NHT	POU4F2	Q12837	ADPRDLEAFAERFKQ	KPILQAWLEEAEKSH	78
HR7479A-183-261-NHT	POU4F3	Q15319	DPRELEAFAERFKQR	VLQAWLEEAEAAYRE	79
HR7056A-224-289-TEV	POU5F1	Q01860	ETLVQARKRKRTSIE	RVWFCNRRQKGKRSS	66
HR8237A-224-288-NHT	POU5F1B	Q06416	ETLMQARKRKRTSIE	RVWFCNRRQKGKRSS	65
HR8392A-142-292-TEV	POU6F1	Q14863	INLEEIREFAKNFKI	VRVWFCNRRQTLKNT	151
HR7133A-97-168-15	PPARA	Q07869	MALNIECRICGDKAS	QYCRFHKCLSVGMSH	73
HR7133A-97-168-Av6HT	PPARA	Q07869	ALNIECRICGDKASG	QYCRFHKCLSVGMSH	72
HR7133A-97-168-TEV	PPARA	Q07869	ALNIECRICGDKASG	QYCRFHKCLSVGMSH	72
HR7133A-97-174-15	PPARA	Q07869	MALNIECRICGDKAS	KCLSVGMSHNAIRFG	79
HR7133A-97-174-Av6HT	PPARA	Q07869	ALNIECRICGDKASG	KCLSVGMSHNAIRFG	78
HR7133A-97-174-TEV	PPARA	Q07869	ALNIECRICGDKASG	KCLSVGMSHNAIRFG	78
HR7133A-97-187-15	PPARA	Q07869	MALNIECRICGDKAS	FGRMPRSEKAKLKAE	92
HR7133A-97-187-Av6HT	PPARA	Q07869	ALNIECRICGDKASG	FGRMPRSEKAKLKAE	91
HR7133A-97-187-TEV	PPARA	Q07869	ALNIECRICGDKASG	FGRMPRSEKAKLKAE	91
HR7133B-182-468-15	PPARA	Q07869	MAKLKAEILTGEHDI	AALHPLLQEIYRDMY	288

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7133B-182-468-Av6HT	PPARA	Q07869	AKLKAEILTCEHDIE	AALHPLLQEIYRDMY	287
HR7133B-182-468-TEV	PPARA	Q07869	AKLKAEILTCEHDIE	AALHPLLQEIYRDMY	287
HR7133C-192-468-15	PPARA	Q07869	MEHDIEDSETADLKS	AALHPLLQEIYRDMY	278
HR7133C-192-468-Av6HT	PPARA	Q07869	EHDIEDSETADLKSL	AALHPLLQEIYRDMY	277
HR7133C-192-468-TEV	PPARA	Q07869	EHDIEDSETADLKSL	AALHPLLQEIYRDMY	277
HR8028A-67-146-15	PPARD	Q03181	MCGSLNMECRVCGDK	KCLALGMSHNAIRFG	81
HR8028A-67-146-Av6HT	PPARD	Q03181	CGSLNMECRVCGDKA	KCLALGMSHNAIRFG	80
HR8028A-67-146-TEV	PPARD	Q03181	CGSLNMECRVCGDKA	KCLALGMSHNAIRFG	80
HR8028A-67-156-15	PPARD	Q03181	MCGSLNMECRVCGDK	AIRFGRMPEAEKRKL	91
HR8028A-67-156-Av6HT	PPARD	Q03181	CGSLNMECRVCGDKA	AIRFGRMPEAEKRKL	90
HR8028A-67-156-TEV	PPARD	Q03181	CGSLNMECRVCGDKA	AIRFGRMPEAEKRKL	90
HR8028A-73-146-15	PPARD	Q03181	MECRVCGDKASGFHY	KCLALGMSHNAIRFG	75
HR8028A-73-146-Av6HT	PPARD	Q03181	ECRVCGDKASGFHYG	KCLALGMSHNAIRFG	74
HR8028A-73-146-TEV	PPARD	Q03181	ECRVCGDKASGFHYG	KCLALGMSHNAIRFG	74
HR8028A-73-156-15	PPARD	Q03181	MECRVCGDKASGFHY	AIRFGRMPEAEKRKL	85
HR8028A-73-156-Av6HT	PPARD	Q03181	ECRVCGDKASGFHYG	AIRFGRMPEAEKRKL	84
HR8028A-73-156-TEV	PPARD	Q03181	ECRVCGDKASGFHYG	AIRFGRMPEAEKRKL	84
HR8028B-163-441-15	PPARD	Q03181	MNEGSQYNPQVADLK	TSLHPLLQEIYKDMY	280
HR8028B-163-441-Av6HT	PPARD	Q03181	NEGSQYNPQVADLKA	TSLHPLLQEIYKDMY	279
HR8028B-163-441-TEV	PPARD	Q03181	NEGSQYNPQVADLKA	TSLHPLLQEIYKDMY	279
HR4464B-222-504-TEV	PPARG	P37231	LAEISSDIDQLNPES	DMSLHPLLQEIYKDL	283
HR7373A-58-198-NHT	PPP1R10	Q96QC0	RSPEILVKFIDVGGY	AEEAPEKKREKPKSL	141

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7854A-111-311-Av6HT	PPP2R3B	Q9Y5P8	TRKEEPLPPATSQSI	LLEEEADINQLTEFF	201
HR6538A-2-187-TEV	PRDM1	O75626	LDICLEKRVGTTLAA	LAACQNGMNIYFYTI	186
HR7699A-188-339-TEV	PRDM10	Q9NQV6	KHGPLHPIPNRPVLT	YAASYAEFVNQKIHD	152
HR6506A-60-229-TEV	PRDM12	Q9H4Q4	KTAFTAEVLAQSFSG	VPGLEEDQKKNKHED	170
HR7923A-243-372-Av6HT	PRDM14	Q9GZV8	DKDSLQLPEGLCLMQ	QNQELLVWYGDCYEK	130
HR8160A-72-214-Av6HT	PRDM16	Q9HAZ2	VYIPEDIPIPADFEL	IEPGEELLVHVKEGV	143
HR4804B-1144-1216-14	PRDM2	Q13029	MLSIKDLTKHLSIHA	FLCNLQQHQRDLHPD	74
HR4804B-1144-1221-14	PRDM2	Q13029	MLSIKDLTKHLSIHA	QQHQRDLHPDKVCTH	79
HR4804B-1144-1230-14	PRDM2	Q13029	MLSIKDLTKHLSIHA	DKVCTHHEFESGTLR	88
HR4804B-1158-1216-14	PRDM2	Q13029	MEEWPFKCEFCVQLF	FLCNLQQHQRDLHPD	60
HR4804B-1158-1221-14	PRDM2	Q13029	MEEWPFKCEFCVQLF	QQHQRDLHPDKVCTH	65
HR4804B-1158-1230-14	PRDM2	Q13029	MEEWPFKCEFCVQLF	DKVCTHHEFESGTLR	74
HR4804C-342-415-14	PRDM2	Q13029	MQIPRTKEEANGDVF	TQINRRRHERRHEAG	75
HR4804C-356-417-14	PRDM2	Q13029	METFMFPCQHCERKF	INRRRHERRHEAGLK	63
HR4804C-360-422-14	PRDM2	Q13029	MFPCQHCERKFTTKQ	HERRHEAGLKRKPSQ	64
HR4804C-367-417-14	PRDM2	Q13029	MRKFTTKQGLERHMH	INRRRHERRHEAGLK	52
HR4804D-2-148-TEV	PRDM2	Q13029	NQNTTEPVAATETLA	EELLVWYNGEDNPEI	147
HR6504A-390-540-TEV	PRDM4	Q9UKN5	EHGPVTFVPDTPIES	FYYSRDYAQQIGVPE	151
HR7347A-1-135-NHT	PRDM5	Q9NQX1	LGMYVPDRFSLKSSR	IGYLDSDMEAEEEEQ	134
HR8295A-234-370-Av6HT	PRDM6	Q9NQX0	PPELPEWLRDLPREV	RGTELLVWYNDSYTS	137
HR7077A-196-395-NHT	PRDM7	Q9NQW5	EPQDDDYLYCEMCQN	VNCWSGMGMSMARNW	200
HR8098A-623-689-Av6HT	PRDM8	Q9NQV8	AQNWCAKCNASFRMT	FRERHHLSRHMTSHN	67

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8069A-197-382-Av6HT	PRDM9	Q9NQV7	PQDDDYLYCEMCQNF	KWGSKWKKELMAGRE	186
HR7506-125-390-Av6HT	PREB	Q9HCU5	EKKCGAETQHEGLEL	SRCQLHLLPSRRSVP	266
HR7506A-125-312-Av6HT	PREB	Q9HCU5	EKKCGAETQHEGLEL	CGHEVVSCLDVSESG	188
HR7506A-133-315-Av6HT	PREB	Q9HCU5	QHEGLELRVENLQAV	EVVSCLDVSESGTFL	183
HR7506A-155-312-15	PREB	Q9HCU5	MPLQKVVCFNHDNTL	CGHEVVSCLDVSESG	159
HR8329A-190-376-Av6HT	PREX2	Q70Z35	KHSDYAAVMEALQAM	RRKGLKLGMEQDTWV	187
HR8329B-674-751-Av6HT	PREX2	Q70Z35	PRETVKIPDSADGLG	AHVTACRKYRRPTKQ	78
HR8329B-674-760-Av6HT	PREX2	Q70Z35	PRETVKIPDSADGLG	RRPTKQDSIQWVYNS	87
HR2833A-1-98-NHT	PRKRIR	O43422	GQLKFNTSEEHHADM	ERYENGRKRLKAYLR	97
HR3353-149-736-14	PRKRIR	O43422	MDEDILPLTLEEKEN	LLNINFDIKHDLDLM	589
HR3353-154-731-14	PRKRIR	O43422	MPLTLEEKENKEYLK	SSNLALLNINFDIKH	579
HR4564B-74-136-14	PROP1	O75360	MTTFSPVQLEQLESA	AKQRKQERSLLQPLA	64
HR4660B-14	PROX1	Q92786	MAMQEGLSPNHLKKA	EIFKSPNCLQELLHE	164
HR4660B-15	PROX1	Q92786	MAMQEGLSPNHLKKA	EIFKSPNCLQELLHE	164
HR4660C-546-737-14	PROX1	Q92786	MTAEGLSLSLIKSEC	EIFKSPNCLQELLHE	193
HR4660C-550-737-14	PROX1	Q92786	MLSLSLIKSECGDLQ	EIFKSPNCLQELLHE	189
HR4660C-566-737-14	PROX1	Q92786	MSEISPYSGSAMQEG	EIFKSPNCLQELLHE	173
HR4660C-572-737-14	PROX1	Q92786	MSGSAMQEGLSPNHL	EIFKSPNCLQELLHE	167
HR4660C-577-737-14	PROX1	Q92786	MQEGLSPNHLKKAKL	EIFKSPNCLQELLHE	162
HR8100A-409-592-Av6HT	PROX2	Q3B8N5	LPLLPSVKMEQRGLH	DSDIPEIFKSSSYPQ	184
HR6440-1-237-14	PRRX1	P54821	MTSSYGHVLERQPAL	SIANLRLKAKEYSLQ	237
HR6440-14	PRRX1	P54821	MTSSYGHVLERQPAL	AKEYSLQRNQVPTVN	245

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6440-22-245-14	PRRX1	P54821	PGNLDTLQAKKNFSV	AKEYSLQRNQVPTVN	224
HR6440-27-245-14	PRRX1	P54821	TLQAKKNFSVSHLLD	AKEYSLQRNQVPTVN	219
HR7233A-95-168-Av6HT	PRRX2	Q99811	GSAAKRKKKQRRNRT	NRRAKFRRNERAMLA	74
HR7750A-100-341-Av6HT	PSMD11	O00231	EAATGQEVELCLFCI	YDNLLEQNLIRVIEP	242
HR7208-1-335-TEV	PSMD12	O00232	ADGGSERADGRIVKM	VEDYGMELRKGSLES	334
HR7208-11-335-TEV	PSMD12	O00232	GRIVKMEVDYSATVD	VEDYGMELRKGSLES	325
HR7208-TEV	PSMD12	O00232	ADGGSERADGRIVKM	THLIAKEEMIHNLQ*	456
HR7208A-338-425-15	PSMD12	O00232	MTDVFGSTEEGEKRW	GIINFQRPKDPNNLL	89
HR7208A-338-456-15	PSMD12	O00232	MTDVFGSTEEGEKRW	TTHLIAKEEMIHNLQ	120
HR7208A-343-420-15	PSMD12	O00232	MSTEEGEKRWKDLKN	VDRLAGIINFQRPKD	79
HR7208A-343-456-15	PSMD12	O00232	MSTEEGEKRWKDLKN	TTHLIAKEEMIHNLQ	115
HR7208A-371-456-15	PSMD12	O00232	MTRITMKRMAQLLDL	TTHLIAKEEMIHNLQ	87
HR7208B-300-417-TEV	PSMD12	O00232	PKYKDLLKLFTTMEL	FAKVDRLAGIINFQR	118
HR7208C-300-456-TEV	PSMD12	O00232	PKYKDLLKLFTTMEL	TTHLIAKEEMIHNLQ	157
HR5110A-437-891-Av6HT	Q6ZU11	Q6ZU11	LNKDQATALIQIAQM	HCEGREDGLQHANQY	455
HR7981A-147-199-TEV	RABGEF1	Q9UJ41	KTFHKTGQEIYKQTK	FYHNVAERMQTRGKV	53
HR8212A-1-114-NHT	RAD51	Q06609	AMQMQLEANADTSVE	IQITTGSKELDKLLQ	113
HR7358A-268-383-TEV	RAG1	P15918	NCSKIHLSTKLLAVD	LEKYNHHISSHKESK	116
HR7829A-213-923-TEV	RAPGEF3	O95398	PDALLTVALRKPPGQ	DNQRELSRLSRELEP	711
HR7829B-33-923-TEV	RAPGEF3	O95398	DVVPEGTLLNMVLRR	DNQRELSRLSRELEP	891
HR8365C-81-160-NHT	RARA	P10276	LPRIYKPCFVCQDKS	QKCFEVGMSKESVRN	80
HR8365C-81-160-TEV	RARA	P10276	LPRIYKPCFVCQDKS	QKCFEVGMSKESVRN	80

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6970C-82-160-TEV	RARB	P10826	FVCQDKSSGYHYGVS	MSKESVRNDRNKKKK	79
HR7515A-85-154-15	RARG	P13631	MRVYKPCFVCNDKSS	NRCQYCRLQKCFEVG	71
HR7515A-85-154-Av6HT	RARG	P13631	RVYKPCFVCNDKSSG	NRCQYCRLQKCFEVG	70
HR7515A-85-154-TEV	RARG	P13631	RVYKPCFVCNDKSSG	NRCQYCRLQKCFEVG	70
HR7515A-85-166-15	RARG	P13631	MRVYKPCFVCNDKSS	EVGMSKEAVRNDRNK	83
HR7515A-85-166-Av6HT	RARG	P13631	RVYKPCFVCNDKSSG	EVGMSKEAVRNDRNK	82
HR7515A-85-166-TEV	RARG	P13631	RVYKPCFVCNDKSSG	EVGMSKEAVRNDRNK	82
HR7515A-97-166-15	RARG	P13631	MSSGYHYGVSSCEGC	EVGMSKEAVRNDRNK	71
HR7515A-97-166-Av6HT	RARG	P13631	SSGYHYGVSSCEGCK	EVGMSKEAVRNDRNK	70
HR7515A-97-166-TEV	RARG	P13631	SSGYHYGVSSCEGCK	EVGMSKEAVRNDRNK	70
HR7515B-178-423-15	RARG	P13631	MDSYELSPQLEELIT	PPLIREMLENPEMFE	247
HR7515B-178-423-Av6HT	RARG	P13631	DSYELSPQLEELITK	PPLIREMLENPEMFE	246
HR7515B-178-423-TEV	RARG	P13631	DSYELSPQLEELITK	PPLIREMLENPEMFE	246
HR7515C-183-417-15	RARG	P13631	MSPQLEELITKVSKA	EIPGPMPPLIREMLE	236
HR7515C-183-417-Av6HT	RARG	P13631	SPQLEELITKVSKAH	EIPGPMPPLIREMLE	235
HR7515C-183-417-TEV	RARG	P13631	SPQLEELITKVSKAH	EIPGPMPPLIREMLE	235
HR7515D-84-169-15	RARG	P13631	MPRVYKPCFVCNDKS	MSKEAVRNDRNKKKK	87
HR7515D-84-169-Av6HT	RARG	P13631	PRVYKPCFVCNDKSS	MSKEAVRNDRNKKKK	86
HR7515D-84-169-TEV	RARG	P13631	PRVYKPCFVCNDKSS	MSKEAVRNDRNKKKK	86
HR8219A-137-208-Av6HT	RAX	Q9Y2V3	RRNRTTFTTYQLHEL	QEKLEVSSMKLQDSP	72
HR8168A-24-86-Av6HT	RAX2	Q96IS3	KKKHRRNRTTFTTYQ	QVWFQNRRAKWRRQE	63
HR7540-167-714-15	RBAK	Q9NYW8	MECGKTYHGEKMCEF	IHRRGNMNVLDVENL	549

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7540-171-714-15	RBAK	Q9NYW8	MTYHGEKMCEFNQNG	IHRRGNMNVLDVENL	545
HR7540-183-714-15	RBAK	Q9NYW8	MNGDTYSHNEENILQ	IHRRGNMNVLDVENL	533
HR7540-188-714-15	RBAK	Q9NYW8	MSHNEENILQKISIL	IHRRGNMNVLDVENL	528
HR7540-7-562-15	RBAK	Q9NYW8	MPVSFKDVAVDFTQE	FNELSYYTEHYRSHS	557
HR7540A-397-562-TEV	RBAK	Q9NYW8	GEKLYKCNECGKSYY	FNELSYYTEHYRSHS	166
HR7540A-397-591-TEV	RBAK	Q9NYW8	GEKLYKCNECGKSYY	SHNSSLFRHQRVHTG	195
HR7540A-428-562-TEV	RBAK	Q9NYW8	PYQCSECGKFFSRVS	FNELSYYTEHYRSHS	135
HR7540A-428-591-TEV	RBAK	Q9NYW8	PYQCSECGKFFSRVS	SHNSSLFRHQRVHTG	164
HR7540B-1-64-Av6HT	RBAK	Q9NYW8	NTLQGPVSFKDVAVD	DTTKPNVIIKLEQGE	63
HR7540C-633-701-Av6HT	RBAK	Q9NYW8	SRMSNLTVHYRSHSG	KFHHRSAFNSHQRIH	69
HR7540C-649-701-Av6HT	RBAK	Q9NYW8	KPYECNECGKVFSQK	KFHHRSAFNSHQRIH	53
HR7540C-653-701-Av6HT	RBAK	Q9NYW8	CNECGKVFSQKSYLT	KFHHRSAFNSHQRIH	49
HR8531A-503-605-Av6HT	RBM20	Q5T481	LASVGTTFAQRKGAG	SKRYKELQLKKPGKA	103
HR7548A-227-304-TEV	RBM22	Q9NW64	EDKTITTLYVGGLGD	KLIVNGRRLNVKWGR	78
HR7417A-596-679-NHT	RBM27	Q9P2N5	QYTNTKLEVKKIPQE	RFIRVLWHRENNEQP	84
HR7740A-184-325-NHT	RBM5	P52756	DWLCNKCCLNNFRKR	DFAKSARKDLVLSDG	142
HR6987A-23-453-TEV	RBPJ	Q06330	GERPPPKRLTREAMR	SLTFTYTPEPGPRPH	431
HR7414A-356-477-NHT	RBPJL	Q9UBG7	SSCWTIIGTESVEFS	DGLFYPSAFSFTYTP	122
HR8038A-1-79-Av6HT	RC3H2	Q9HBD1	PVQAAQWTEFLSCPI	PVNFALLQLVGAQVP	78
HR7631A-308-440-TEV	RCOR1	Q9UKL0	RKPPKGMFLSQEDVE	RRFNIDEVLQEWEAE	133
HR7631B-158-237-Av6HT	RCOR1	Q9UKL0	GYNMEQALGMLFWHK	IASLVKFYYSWKKTR	80
HR7631B-158-247-Av6HT	RCOR1	Q9UKL0	GYNMEQALGMLFWHK	WKKTRTKTSVMDRHA	90

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7631B-172-237-Av6HT	RCOR1	Q9UKL0	KHNIEKSLADLPNFT	IASLVKFYYSWKKTR	66
HR7631B-172-247-Av6HT	RCOR1	Q9UKL0	KHNIEKSLADLPNFT	WKKTRTKTSVMDRHA	76
HR7640A-292-387-NHT	RCOR2	Q8IZ40	ISLKRQVQSMKQTNS	YRRRFNLEEVLQEWE	96
HR2844A-14	REL	Q04864	EKDTYGNKAKKQKTT	NEQLSDSFPYEFFQV	337
HR2844B-14	REL	Q04864	EKDTYGNKAKKQKTT	NSQGIPPFLRIPVGN	171
HR2844C-14	REL	Q04864	DLNASNACIYNNADD	NEQLSDSFPYEFFQV	166
HR2845-14	RELA	Q04206	MDELFPLIFPAEPAQ	IADMDFSALLSQISS	551
HR2845B-14	RELA	Q04206	TDDRHRIEEKRKRTY	QFDDEDLGALLGNST	167
HR2845C-14	RELA	Q04206	DPAVFTDLASVDNSE	IADMDFSALLSQISS	93
HR2845D-191-291-Av6HT	RELA	Q04206	TAELKICRVNRNSGS	DRELSEPMEFQYLPD	101
HR2845D-191-291-TEV	RELA	Q04206	TAELKICRVNRNSGS	DRELSEPMEFQYLPD	101
HR2846-40-579-15	RELB	Q01201	MLSSLSLAVSRSTDE	AAFGGGLLSPGPEAT	541
HR2846B-14	RELB	Q01201	DHDSYGVDKKRKRGM	AAFGGGLLSPGPEAT	179
HR6006B-21	RERE	Q9P2R6	STQGEIRVGPSHQAK	LITFYYYWKKTPEAA	165
HR6006C-21	RERE	Q9P2R6	STQGEIRVGPSHQAK	RLVKKPVPKLIEKCW	116
HR6006D-21	RERE	Q9P2R6	STQGEIRVGPSHQAK	DLLMYLRAARSMAAP	60
HR6006E-21	RERE	Q9P2R6	VTQHEELVWMPGVND	ETGELITFYYYWKKT	132
HR6006F-21	RERE	Q9P2R6	VTQHEELVWMPGVND	RLVKKPVPKLIEKCW	87
HR6006H-21	RERE	Q9P2R6	QGEIRVGPSHQAKLP	ETGELITFYYYWKKT	159
HR6969A-342-413-NHT	REST	Q13127	SNQHEVTRHARQVHN	SKKGNLQYHFKSKHP	72
HR8119A-438-513-TEV	RFX1	P22670	TVQWLLDNYETAEGV	RGNSKYHYYGLRIKA	76
HR7754A-200-273-TEV	RFX2	P48378	LQWLLDNYETAEGVS	TRGNSKYHYYGIRLK	74

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7471A-184-257-TEV	RFX3	P48380	LQWLLDNYETAEGVS	TRGNSKYHYYGIRVK	74
HR8007A-101-167-15	RFX5	P48382	MEEHTDTGLPKQSVY	GRGQSKYCYSGIRRK	68
HR8007A-76-173-15	RFX5	P48382	MDKSSEPSTLSNEEY	YCYSGIRRKTLVSMP	99
HR8007A-81-167-15	RFX5	P48382	MPSTLSNEEYMYAYR	GRGQSKYCYSGIRRK	88
HR7289A-107-205-NHT	RFX6	Q8HWS3	KKTITQIVKDKKKQT	HYYGIGIKESSAYYH	99
HR7790-68-260-TEV	RFXANK	O14593	KHSTTLTNRQRGNEV	ILKLFQSNLVPADPE	193
HR7790-79-260-TEV	RFXANK	O14593	GNEVSALPATLDSLS	ILKLFQSNLVPADPE	182
HR7790A-101-248-15	RFXANK	O14593	MGELDQLKEHLRKGD	GYRKVQQVIENHILK	149
HR7790A-101-260-15	RFXANK	O14593	MGELDQLKEHLRKGD	ILKLFQSNLVPADPE	161
HR7790A-79-248-15	RFXANK	O14593	MGNEVSALPATLDSL	GYRKVQQVIENHILK	171
HR7790A-86-251-15	RFXANK	O14593	MPATLDSLSIHQLAA	KVQQVIENHILKLFQ	167
HR7790A-91-248-15	RFXANK	O14593	MSLSIHQLAAQGELD	GYRKVQQVIENHILK	159
HR7790A-91-260-15	RFXANK	O14593	MSLSIHQLAAQGELD	ILKLFQSNLVPADPE	171
HR7961A-325-470-TEV	RGS6	P49758	PSQQRVKRWGFSFDE	KGKSLAGKRLTGLMQ	146
HR6895A-323-449-TEV	RGS7	P49802	SQQRVKRWGFGMDEA	YPRFIRSSAYQELLQ	127
HR6935A-279-424-TEV	RGS9	O75916	DLNAKLVEIPTKMRV	YKDMLAKAIEPQETT	146
HR7291A-130-236-NHT	RHOXF2	Q9BQY4	PGNAQQPNVHAFTPL	PLFISGMRDDYFWDH	107
HR7532A-130-236-NHT	RHOXF2B	P0C7M4	PGNAQQPNVHAFTPL	PLFISGMRDDYFWDH	107
HR7189A-91-307-Av6HT	RIOK2	Q9BVS4	RQVVESVGNQMGVGK	EDTLDVEVSASGYTK	217
HR7201A-769-825-Av6HT	RLF	Q13129	LRYKCELNGCNIVFS	FYYSKIEYQNHLSMH	57
HR7201A-769-837-NHT	RLF	Q13129	LRYKCELNGCNIVFS	SMHNVENSNGDIKKS	69
HR7201B-707-825-Av6HT	RLF	Q13129	LDMKNRREKCTYCRR	FYYSKIEYQNHLSMH	119

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7201B-707-837-Av6HT	RLF	Q13129	LDMKNRREKCTYCRR	SMHNVENSNGDIKKS	131
HR7201B-716-825-Av6HT	RLF	Q13129	CTYCRRHFMSAFHLR	FYYSKIEYQNHLSMH	110
HR7201B-716-837-Av6HT	RLF	Q13129	CTYCRRHFMSAFHLR	SMHNVENSNGDIKKS	122
HR7201C-707-770-Av6HT	RLF	Q13129	LDMKNRREKCTYCRR	VNELLNHKQKHDDLR	64
HR7201C-725-770-Av6HT	RLF	Q13129	SAFHLREHEQVHCGP	VNELLNHKQKHDDLR	46
HR7461A-26-161-TEV	RNASE2	P10153	HVKPPQFTWAQWFET	PPQYPVVPVHLDRII	136
HR7865A-252-319-TEV	RNF113A	O15541	GSDDEEIPFKCFICR	GVFNPAKELIAKLEK	68
HR7107A-246-319-TEV	RNF113B	Q8IZP6	GSEEEEIPFRCFICR	KELMAKLQKLQAAEG	74
HR7482A-1-89-NHT	RNF114	Q9Y508	AAQQRDCGGAAQLAG	VRAVELERQIESTET	88
HR7645A-17-100-NHT	RNF125	Q96EQ8	ATARALERRRDPELP	ATDVAKRMKSEYKNC	84
HR4563B-87-210-14	RORA	P35398	MKEDKEVQTGYMNAQ	HRMQQQQRDHQQQPG	125
HR4563B-92-174-14	RORA	P35398	MVQTGYMNAQIEIIP	HCRLQKCLAVGMSRD	84
HR4563B-92-193-14	RORA	P35398	MVQTGYMNAQIEIIP	GRMSKKQRDSLYAEV	103
HR4563B-93-210-14	RORA	P35398	MQTGYMNAQIEIIPC	HRMQQQQRDHQQQPG	119
HR4563B-98-174-14	RORA	P35398	MNAQIEIIPCKICGD	HCRLQKCLAVGMSRD	78
HR4563B-98-193-14	RORA	P35398	MNAQIEIIPCKICGD	GRMSKKQRDSLYAEV	97
HR7194B-260-507-TEV	RORC	P51449	PEAPYASLTEIEHLV	VVQAAFPPLYKELFS	248
HR7255A-172-270-Av6HT	RPA2	P15927	ANSQPSAGRAPISNP	YSTVDDDHFKSTDAE	99
HR7255A-172-270-TEV	RPA2	P15927	ANSQPSAGRAPISNP	YSTVDDDHFKSTDAE	99
HR7006A-95-145-NHT	RREB1	Q92766	ADHSCSICGKSLSSA	GQSFTTNGNMHRHMK	51
HR4447C-46-185-Av6HT	RUNX1	Q01196	MSGDRSMVEVLADHP	DGPREPRRHRQKLDD	141
HR4447C-46-185-TEV	RUNX1	Q01196	SGDRSMVEVLADHPG	DGPREPRRHRQKLDD	140

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4568B-112-233-TEV	RUNX2	Q13950	ELVRTDSPNFLCSVL	VTVDGPREPRRHRQK	122
HR7324A-53-189-TEV	RUNX3	Q13761	QAAVGPGGRARPEVR	TQVATYHRAIKVTVD	137
HR4643B-135-200-TEV	RXRA	P19793	CAICGDRSSGKHYGV	RCQYCRYQKCLAMGM	66
HR8407C-205-270-NHT	RXRB	P28702	CAICGDRSSGKHYGV	RCQYCRYQKCLATGM	66
HR8407C-205-270-TEV	RXRB	P28702	CAICGDRSSGKHYGV	RCQYCRYQKCLATGM	66
HR4751B-139-204-TEV	RXRG	P48443	CAICGDRSSGKHYGV	RCQYCRYQKCLVMGM	66
HR7653A-909-977-NHT	SALL2	Q9Y467	SRKACEVCGQAFPSQ	HHQVQPFAPHGPQNI	69
HR7433A-975-1045-NHT	SALL3	Q9BXA9	PSTVCGVCGKPFACK	ELPSQLFDPNFALGP	71
HR6875A-376-433-Av6HT	SALL4	Q9UJQ4	MEAALYKHKCKYCSK	FTTKGNLKVHFHRHP	59
HR6875A-376-433-NHT	SALL4	Q9UJQ4	EAALYKHKCKYCSKV	FTTKGNLKVHFHRHP	58
HR4435B-174-250-14	SATB1	Q01826	MPKLEDLPPEQWSHT	FGRWYKHFKKTKDMM	78
HR4435B-174-254-14	SATB1	Q01826	MPKLEDLPPEQWSHT	YKHFKKTKDMMVEMD	82
HR4435B-179-244-14	SATB1	Q01826	MLPPEQWSHTTVRNA	AAKCQEFGRWYKHFK	67
HR4435B-179-250-14	SATB1	Q01826	MLPPEQWSHTTVRNA	FGRWYKHFKKTKDMM	73
HR4435C-368-452-TEV	SATB1	Q01826	NTEVSSEIYQWVRDE	AERDRIYQDERERSL	85
HR4435D-53-254-15	SATB1	Q01826	MQGVPLKHSGHLMKT	YKHFKKTKDMMVEMD	203
HR4435D-56-250-15	SATB1	Q01826	MPLKHSGHLMKTNLR	FGRWYKHFKKTKDMM	196
HR4435E-53-178-15	SATB1	Q01826	MQGVPLKHSGHLMKT	VTLKIQLHSCPKLEQ	127
HR4435E-56-175-15	SATB1	Q01826	MPLKHSGHLMKTNLR	YHVVTLKIQLHSCPK	121
HR7571A-350-437-NHT	SATB2	Q9UPW6	KPEPTNSSVEVSPDI	NLPEVERDRIYQDER	88
HR7571B-610-674-TEV	SATB2	Q9UPW6	SCAKKPRSRTKISLE	IKFFQNQRYHVKHHG	65
HR5092A-90-156-Av6HT	SCAPER	Q9BY12	RHPRKIDLRARYWAF	DFKALIDWIQLQEKL	67

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8394A-256-325-Av6HT	SCRT1	Q9BWW7	AFSRPWLLQGHMRSH	KSFALKSYLNKHYES	70
HR7196A-156-206-NHT	SCRT2	Q9NQ03	ACAECGKTYATSSNL	GKAYVSMPALAMHLL	51
HR3583D-653-871-15	SETDB1	Q15047	MLFLEMFCLDPYVLV	LDHIESVENFKEGYE	220
HR3583D-658-867-15	SETDB1	Q15047	MFCLDPYVLVDRKFQ	YFANLDHIESVENFK	211
HR3583E-555-676-15	SETDB1	Q15047	MLERAPAEPSYRAPM	PYVLVDRKFQPYKPF	123
HR3583E-584-681-15	SETDB1	Q15047	MSRVRPMRNEQYRGK	DRKFQPYKPFYYILD	99
HR3583E-590-676-15	SETDB1	Q15047	MRNEQYRGKNPLLVP	PYVLVDRKFQPYKPF	88
HR8073A-102-182-Av6HT	SHOX	O15266	EDVKSEDEDGQTKLK	RRAKCRKQENQMHKG	81
HR6933A-647-727-Av6HT	SHPRH	Q149N8	NTMSPFNTSDYRFEC	VSTRATLIISPSSIC	81
HR6933A-647-739-NHT	SHPRH	Q149N8	NTMSPFNTSDYRFEC	SICHQWVDEINRHVR	93
HR6933A-654-727-Av6HT	SHPRH	Q149N8	TSDYRFECICGELDQ	VSTRATLIISPSSIC	74
HR6933A-654-739-Av6HT	SHPRH	Q149N8	TSDYRFECICGELDQ	SICHQWVDEINRHVR	86
HR6933B-437-502-Av6HT	SHPRH	Q149N8	VQCPPTRVMILTAVK	KCLIFEGLVKQIKGH	66
HR6933B-437-512-Av6HT	SHPRH	Q149N8	VQCPPTRVMILTAVK	QIKGHGFSGTFTLGK	76
HR6933C-1495-1636-Av6HT	SHPRH	Q149N8	KANQEEDIPVKGSHS	GQTKPTIVHRFLIKA	142
HR6933C-1495-1646-Av6HT	SHPRH	Q149N8	KANQEEDIPVKGSHS	FLIKATIEERMQAML	152
HR6933C-1495-1659-Av6HT	SHPRH	Q149N8	KANQEEDIPVKGSHS	MLKTAERSHTNSSAK	165
HR7129A-227-345-NHT	SIM1	P81133	LHSNMFMFRASLDMK	VLTDTEYKGLQLSLD	119
HR8357A-72-194-Av6HT	SIM2	Q14190	PLDGVAKELGSHLLQ	YKVIHCSGYLKIRQY	123
HR7143A-205-281-Av6HT	SIX3	O95343	DGEQKTHCFKERTRS	KNRLQHQAIGPSGMR	77
HR7095A-206-294-Av6HT	SIX4	Q9UIU6	DKYRLRRKFPLPRTI	NPSETQSKSESDGNP	89
HR8072A-125-195-Av6HT	SIX6	O95475	IWDGEQKTHCFKERT	QRDRAAAAKNRLQQQ	71

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4810B-218-313-TEV	SKI	P12755	VRVYHECFGKCKGLL	RLGRCLDDVKEKFDY	96
HR8491B-28-132-Av6HT	SKOR2	Q2VWA4	QPRPGHANLKPNQVG	ITKREAERLCKSFLG	105
HR8491C-141-237-Av6HT	SKOR2	Q2VWA4	DNFAFDVSHECAWGC	ELVFAWEDVKAMFNG	97
HR8491C-141-244-Av6HT	SKOR2	Q2VWA4	DNFAFDVSHECAWGC	DVKAMFNGGSRKRAL	104
HR8491D-43-132-Av6HT	SKOR2	Q2VWA4	QVILYGIPIVSLVID	ITKREAERLCKSFLG	90
HR7664A-124-217-TEV	SLC30A9	Q6PML9	KYTQNNFITGVRAIN	ERLFRNQKILREYRD	94
HR8337A-9-132-TEV	SMAD1	Q15797	FTSPAVKRLLGWKQG	KEVCINPYHYKRVES	124
HR8337B-248-465-TEV	SMAD1	Q15797	APPLPSEINRGDVQA	LTQMGSPHNPISSVS	218
HR4670B-55-191-14	SMAD2	Q15796	MTGRLDELEKAITTQ	PVLVPRHTEILTELP	138
HR4670B-55-196-14	SMAD2	Q15796	MTGRLDELEKAITTQ	RHTEILTELPPLDDY	143
HR4670B-55-202-14	SMAD2	Q15796	MTGRLDELEKAITTQ	TELPPLDDYTHSIPE	149
HR4670B-55-202-Av6HT	SMAD2	Q15796	TGRLDELEKAITTQN	TELPPLDDYTHSIPE	148
HR4670B-55-202-TEV	SMAD2	Q15796	TGRLDELEKAITTQN	TELPPLDDYTHSIPE	148
HR4670B-6-173-14	SMAD2	Q15796	MPFTPPVVKRLLGWK	EVCVNPYHYQRVETP	169
HR4670C-101-173-14	SMAD2	Q15796	MLYSFSEQTRSLDGR	EVCVNPYHYQRVETP	74
HR4670C-106-173-14	SMAD2	Q15796	MEQTRSLDGRLQVSH	EVCVNPYHYQRVETP	69
HR4670D-261-456-Av6HT	SMAD2	Q15796	LDLQPVTYSEPAFWC	LNGPLQWLDKVLTQM	196
HR4503B-1-148-14	SMAD4	Q13485	MDNMSITNTPTSNDA	ERVVSPGIDLSGLTL	148
HR4503B-1-166-14	SMAD4	Q13485	MDNMSITNTPTSNDA	APSSMMVKDEYVHDF	166
HR4503B-10-160-14	SMAD4	Q13485	MPTSNDACLSIVHSL	LTLQSNAPSSMMVKD	152
HR4503B-11-142-14	SMAD4	Q13485	MTSNDACLSIVHSLM	VNPYHYERVVSPGID	133
HR4503C-9-149-14	SMAD4	Q13485	MTPTSNDACLSIVHS	RVVSPGIDLSGLTLQ	142

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4503D-314-552-Av6HT	SMAD4	Q13485	ISNHPAPEYWCSIAY	EVLHTMPIADPQPLD	239
HR5565A-14	SMAD6	O43541	MRLSPRDEYKPLDLS	TSCPCWLEILLNNPR	217
HR5560A-14	SMAD7	O15105	MVPSSAETGGTNYLA	ISSCPCWLEVIFNSR	199
HR5560A-15	SMAD7	O15105	MVPSSAETGGTNYLA	ISSCPCWLEVIFNSR	199
HR7626A-769-932-TEV	SMARCA1	P28370	VSEPKIPKAPRPPKQ	QIERGEARIQRRISI	164
HR7914A-754-917-TEV	SMARCA5	O60264	VSEPKAPKAPRPPKQ	QIERGEARIQRRISI	164
HR7256A-607-680-TEV	SMARCC1	Q92922	SKKTLAKSKGASAGR	IEDPYLENSDASLGP	74
HR7400B-419-526-Av6HT	SMARCC2	Q8TAQ2	EQTHHIIIPSYAAWF	YQVDAESRPTPMGPP	108
HR7400B-419-538-Av6HT	SMARCC2	Q8TAQ2	EQTHHIIIPSYAAWF	GPPPTSHFHVLADTP	120
HR7400B-421-520-Av6HT	SMARCC2	Q8TAQ2	THHIIIPSYAAWFDY	QWGLINYQVDAESRP	100
HR7400C-421-514-Av6HT	SMARCC2	Q8TAQ2	THHIIIPSYAAWFDY	VHAFLEQWGLINYQV	94
HR7811A-46-134-NHT	SMARCE1	Q969G3	GTNSRVTASSGITIP	EAEKIEYNESMKAYH	89
HR7811A-46-146-Av6HT	SMARCE1	Q969G3	GTNSRVTASSGITIP	AYHNSPAYLAYINAK	101
HR7520-1-242-15	SNAI1	O95863	MPRSFLVRKPSDPNR	QTHSDVKKYQCQACA	242
HR7520-1-247-15	SNAI1	O95863	MPRSFLVRKPSDPNR	VKKYQCQACARTFSR	247
HR7520-1-253-TEV	SNAI1	O95863	PRSFLVRKPSDPNRK	QACARTFSRMSLLHK	252
HR7520-15	SNAI1	O95863	MPRSFLVRKPSDPNR	LHKHQESGCSGCPR*	265
HR7520-34-264-Av6HT	SNAI1	O95863	PYDQAHLLAAIPPPE	LLHKHQESGCSGCPR	231
HR7520-40-264-15	SNAI1	O95863	MLLAAIPPPEILNPT	LLHKHQESGCSGCPR	226
HR7520-45-264-15	SNAI1	O95863	MRPPEILNPTASLPM	LLHKHQESGCSGCPR	221
HR7012A-122-179-NHT	SNAI2	O43623	AIEAEKFQCNLCNKT	DKEYVSLGALKMHIR	58
HR7849A-212-292-NHT	SNAI3	Q3KNW1	KICGKAFSRPWLLQG	SLLARHEESGCCPGP	81

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7971A-346-397-Av6HT	SNAPC4	Q5SXM2	RKEWTEEEDRMLTQL	DSMQLIYRWTKSLDP	52
HR7549A-165-287-NHT	SOHLH2	Q9NX45	EHLGYFPTDLFACSE	RFCKKQQTPIELSLP	123
HR7723A-49-131-TEV	SOX1	O00570	DRVKRPMNAFMVWSR	DYKYRPRRKTKTLLK	83
HR7246A-102-183-15	SOX10	P56693	MPHVKRPMNAFMVWA	PDYKYQPRRRKNGKA	83
HR7246A-109-178-15	SOX10	P56693	MNAFMVWAQAARRKL	HKKDHPDYKYQPRRR	71
HR7246A-127-178-15	SOX10	P56693	MPHLHNAELSKTLGK	HKKDHPDYKYQPRRR	53
HR7246A-91-179-15	SOX10	P56693	MPVRVNGASKSKPHV	KKDHPDYKYQPRRRK	90
HR7180A-31-110-Av6HT	SOX12	O15370	GWCKTPSGHIKRPMN	LRLKHMADYPDYKYR	80
HR7313A-421-500-TEV	SOX13	Q9UN79	SSHIKRPMNAFMVWA	EKYPDYKYKPRPKRT	80
HR7773A-2-88-TEV	SOX14	O95416	SKPSDHIKRPMNAFM	DYKYRPRRKPKNLLK	87
HR7489A-83-161-TEV	SOX18	P35713	SRIRRPMNAFMVWAK	RDHPNYKYRPRRKKQ	79
HR8317A-38-121-TEV	SOX2	P48431	PDRVKRPMNAFMVWS	DYKYRPRRKTKTLMK	84
HR8041A-6-88-TEV	SOX21	Q9Y651	DHVKRPMNAFMVWSR	DYKYRPRRKPKTLLK	83
HR7838A-132-219-TEV	SOX3	P41225	GGTDQDRVKRPMNAF	DYKYRPRRKTKTLLK	88
HR8424A-45-130-Av6HT	SOX4	Q06945	KADDPSWCKTPSGHI	RLKHMADYPDYKYRP	86
HR7351A-554-632-TEV	SOX5	P35711	PHIKRPMNAFMVWAK	EKYPDYKYKPRPKRT	79
HR7953A-619-697-TEV	SOX6	P35712	HNSNISKILGSRWKS	YKQLMRSRRQEMRQF	78
HR7275A-43-121-TEV	SOX7	Q9BT81	SRIRRPMNAFMVWAK	QDYPNYKYRPRRKKQ	79
HR7103A-102-174-Av6HT	SOX8	P57073	VKRPMNAFMVWAQAA	VQHKKDHPDYKYQPR	73
HR7103A-63-143-Av6HT	SOX8	P57073	RFPACIRDAVSQVLK	NAELSKTLGKLWRLL	81
HR7103A-63-150-Av6HT	SOX8	P57073	RFPACIRDAVSQVLK	LGKLWRLLSESEKRP	88
HR7103A-63-173-NHT	SOX8	P57073	RFPACIRDAVSQVLK	RVQHKKDHPDYKYQP	111

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7103A-82-143-Av6HT	SOX8	P57073	SLVPMPVRGGGGGAL	NAELSKTLGKLWRLL	62
HR7103A-82-150-Av6HT	SOX8	P57073	SLVPMPVRGGGGGAL	LGKLWRLLSESEKRP	69
HR7103A-82-173-Av6HT	SOX8	P57073	SLVPMPVRGGGGGAL	RVQHKKDHPDYKYQP	92
HR6433-64-495-15	SOX9	P48436	SEEDKFPVCIREAVS	ADTSGVPSIPQTHSP	432
HR6433-64-509-14	SOX9	P48436	SEEDKFPVCIREAVS	PQHWEQPVYTQLTRP	446
HR6433-68-490-15	SOX9	P48436	KFPVCIREAVSQVLK	MYTPIADTSGVPSIP	423
HR6433-68-509-14	SOX9	P48436	KFPVCIREAVSQVLK	PQHWEQPVYTQLTRP	442
HR4634C-496-558-14	SP1	P08047	MSSSNTTLTPIASAA	VHPIQGLPLAIANAP	64
HR4744B-38-153-14	SP100	P23497	MFTEDQGVDDRLLYD	HIYKGFENVIHDKLP	117
HR4744B-42-149-14	SP100	P23497	MQGVDDRLLYDIVFK	PDLIHIYKGFENVIH	109
HR4744B-70-135-14	SP100	P23497	MKTFPFLEGLRDRDL	VLEALFSDVNMQEYP	67
HR4744B-70-149-14	SP100	P23497	MKTFPFLEGLRDRDL	PDLIHIYKGFENVIH	81
HR4744C-595-684-TEV	SP100	P23497	DENINFKQSELPVTC	MENKFLPEPPSTRKK	90
HR7625A-677-739-NHT	SP140	Q13342	SQNNSSVDPCMRNLD	RTPWNCIFCRMKESP	63
HR7855A-523-581-Av6HT	SP2	Q02086	KKHVCHIPDCGKTFR	TRSDELQRHARTHTG	59
HR8154A-336-382-Av6HT	SP5	Q6BEB4	SFTRSDELQRHLRTH	SDHLAKHVKTHQNKK	47
HR7866A-256-329-Av6HT	SP6	Q3SY56	CHIPGCGKAYAKTSH	PCAVCSRVFMRSDHL	74
HR7872A-292-352-Av6HT	SP7	Q8TDD2	PIHSCHIPGCGKVYG	SDELERHVRTHTREK	61
HR7447A-131-213-TEV	SPDEF	O95238	LKDIETACKLLNITA	GDVLHAHLDIWKSAA	83
HR8383A-169-257-NHT	SPI1	P17947	KKIRLYQFLLDLLRS	KKVKKKLTYQFSGEV	89
HR4679B-130-262-14	SPIB	Q01892	MEEEDLPLDSPALEV	TYQFDSALLPAVRRA	134
HR4679B-134-262-14	SPIB	Q01892	MLPLDSPALEVSDSE	TYQFDSALLPAVRRA	130

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4679B-163-262-14	SPIB	Q01892	MAGTRKKLRLYQFLL	TYQFDSALLPAVRRA	101
HR4679B-168-262-14	SPIB	Q01892	MKLRLYQFLLGLLTR	TYQFDSALLPAVRRA	96
HR7954A-111-207-Av6HT	SPIC	Q8N5J4	LRLFEYLHESLYNPE	FSEAILQRLSPSYFL	97
HR7260A-815-910-NHT	SRCAP	Q6ZRS2	IEGSQEYNEGLVKRL	QLRKVCNHPNLFDPR	96
HR7260B-583-850-Av6HT	SRCAP	Q6ZRS2	EITDIAAAAESLQPK	FLLRRVKVDVEKQMP	268
HR7260B-597-840-Av6HT	SRCAP	Q6ZRS2	KGYTLATTQVKTPIP	VKRLHKVLRPFLLRR	244
HR7260B-601-850-Av6HT	SRCAP	Q6ZRS2	LATTQVKTPIPLLLR	FLLRRVKVDVEKQMP	250
HR7260B-607-840-Av6HT	SRCAP	Q6ZRS2	KTPIPLLLRGQLREY	VKRLHKVLRPFLLRR	234
HR7260B-607-850-Av6HT	SRCAP	Q6ZRS2	KTPIPLLLRGQLREY	FLLRRVKVDVEKQMP	244
HR4448F-14	SREBF1	P36956	VLLFVYGEPVTRPHS	VVRTSLWRQQQPPAP	432
HR4448G-521-624-14	SREBF1	P36956	MVYHSPGRNVLGTES	RALGRPLPTSHLDLA	105
HR4448G-521-643-14	SREBF1	P36956	MVYHSPGRNVLGTES	WNLIRHLLQRLWVGR	124
HR4448G-526-619-14	SREBF1	P36956	MGRNVLGTESRDGPG	LWLALRALGRPLPTS	95
HR4448G-526-638-14	SREBF1	P36956	MGRNVLGTESRDGPG	ACSLLWNLIRHLLQR	114
HR4448G-530-624-14	SREBF1	P36956	MLGTESRDGPGWAQW	RALGRPLPTSHLDLA	96
HR4448G-530-643-14	SREBF1	P36956	MLGTESRDGPGWAQW	WNLIRHLLQRLWVGR	115
HR4448G-535-619-14	SREBF1	P36956	MRDGPGWAQWLLPPV	LWLALRALGRPLPTS	86
HR4448G-535-638-14	SREBF1	P36956	MRDGPGWAQWLLPPV	ACSLLWNLIRHLLQR	105
HR4448H-319-400-TEV	SREBF1	P36956	QSRGEKRTAHNAIEK	SLRTAVHKSKSLKDL	82
HR4448H-TEV	SREBF1	P36956	QSRGEKRTAHNAIEK	SLRTAVHKSKSLKDL	82
HR6329A-1075-1134-14	SREBF2	Q12772	MPGQRERATAILLAC	RSCNDCQQMIVKLGG	61
HR4543C-132-223-TEV	SRF	P11831	SGAKPGKKTRGRVKI	ETGKALIQTCLNSPD	92

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6924A-56-131-Av6HT	SRY	Q05066	VQDRVKRPMNAFIVW	QAMHREKYPNYKYRP	76
HR6924A-56-131-TEV	SRY	Q05066	VQDRVKRPMNAFIVW	QAMHREKYPNYKYRP	76
HR7075A-105-202-Av6HT	SSB	P05455	KNDVKNRSVYIKGFP	YFAKKNEERKQNKVE	98
HR7075A-105-202-TEV	SSB	P05455	KNDVKNRSVYIKGFP	YFAKKNEERKQNKVE	98
HR7013A-305-448-TEV	SSH2	Q76I76	DSPTQIFEHVFLGSE	SFMRQLEEYQGILLA	143
HR7844A-287-452-NHT	SSH3	Q8TE77	DLESVTSKEIRQALE	QALRHVQELRPIARP	166
HR3575-1-551-14	SSRP1	Q08945	MAETLEFNDVYQEVK	EVKKGKDPNAPKRPM	551
HR3575-1-556-14	SSRP1	Q08945	MAETLEFNDVYQEVK	KDPNAPKRPMSAYML	556
HR3575-1-573-14	SSRP1	Q08945	MAETLEFNDVYQEVK	NASREKIKSDHPGIS	573
HR5522A-14	SSRP1	Q08945	MLKKAKMAKDRKSRK	RDYEKAMKEYEGGRG	101
HR5522A-15	SSRP1	Q08945	MLKKAKMAKDRKSRK	RDYEKAMKEYEGGRG	101
HR7020A-812-906-TEV	ST18	Q60284	PELKCPVIGCDGQGH	GCPLNAQVIKKGKVS	95
HR8389A-136-710-Av6HT	STAT1	P42224	LDKQKELDSKVRNVK	PKGTGYIKTELISVS	575
HR8389A-136-710-NHT	STAT1	P42224	LDKQKELDSKVRNVK	PKGTGYIKTELISVS	575
HR8389A-136-710-TEV	STAT1	P42224	LDKQKELDSKVRNVK	PKGTGYIKTELISVS	575
HR3569D-573-771-14	STAT2	P52630	NDGRIMGFVSRSQER	STLEPVIEPTLGMVS	199
HR3569E-1-131-14	STAT2	P52630	MAQWEMLQNLDSPFQ	RILIQAQRAQLEQGE	131
HR3569E-1-186-14	STAT2	P52630	MAQWEMLQNLDSPFQ	VFCFRYKIQAKGKTP	186
HR5539A-14	STAT2	P52630	MAQWEMLQNLDSPFQ	LEEKRILIQAQRAQL	127
HR5539A-15	STAT2	P52630	MAQWEMLQNLDSPFQ	LEEKRILIQAQRAQL	127
HR5535A-1-101-14	STAT3	P40763	MAQWNQLQQLDTRYL	KQFLQSRYLEKPMEI	101
HR5535A-14	STAT3	P40763	MAQWNQLQQLDTRYL	WEESRLLQTAATAAQ	124

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR5535B-1-116-14	STAT3	P40763	MAQWNQLQQLDTRYL	ARIVARCLWEESRLL	116
HR5535B-1-133-14	STAT3	P40763	MAQWNQLQQLDTRYL	AATAAQQGGQANHPT	133
HR3612-187-748-14	STAT4	Q14765	MNSAMVNQEVLTLQE	PTTIETAMKSPYSAE	563
HR5542A-14	STAT5A	P42229	MAGWIQAQQLQGDAL	NEQRLVREANNCSSP	129
HR5542A-15	STAT5A	P42229	MAGWIQAQQLQGDAL	NEQRLVREANNCSSP	129
HR5542B-128-712-NHT	STAT5A	P42229	SPAGILVDAMSQKHL	QIKQVVPEFVNASAD	585
HR5541A-1-102-14	STAT5B	P51692	MAVWIQAQQLQGEAL	GHYATQLQNTYDRCP	102
HR5541A-1-106-14	STAT5B	P51692	MAVWIQAQQLQGEAL	TQLQNTYDRCPMELV	106
HR5541A-14	STAT5B	P51692	MAVWIQAQQLQGEAL	NEQRLVREANNGSSP	129
HR5541A-15	STAT5B	P51692	MAVWIQAQQLQGEAL	NEQRLVREANNGSSP	129
HR5541B-1-127-14	STAT5B	P51692	MAVWIQAQQLQGEAL	LYNEQRLVREANNGS	127
HR5541B-1-135-14	STAT5B	P51692	MAVWIQAQQLQGEAL	REANNGSSPAGSLAD	135
HR5541C-1-684-TEV	STAT5B	P51692	AVWIQAQQLQGEALH	FPDRPKDEVYSKYYT	683
HR3396C-1-127-14	STAT6	P42226	MSLWGLVSKMPPEKV	QFRHLPMPFHWKQEE	127
HR3396C-1-169-14	STAT6	P42226	MSLWGLVSKMPPEKV	AEAGQVSLHSLIETP	169
HR3396C-1-174-14	STAT6	P42226	MSLWGLVSKMPPEKV	VSLHSLIETPANGTG	174
HR3396D-72-655-14	STAT6	P42226	GEGSTILQHISTLES	YVPATIKMTVERDQP	584
HR3396E-90-279-14	STAT6	P42226	RDPLKLVATFRQILQ	LRTLVTSCFLVEKQP	190
HR3396E-90-327-14	STAT6	P42226	RDPLKLVATFRQILQ	ADMVTEKQARELSVP	238
HR3396F-1-630-TEV	STAT6	P42226	SLWGLVSKMPPEKVQ	YPKKPKDEAFRSHYK	629
HR7864A-355-443-TEV	TADA2A	O75478	SNSGRRSAPPLNLTG	KIYDFLIREGYITKG	89
HR8503A-244-333-Av6HT	TADA2B	B3KX99	KEDGKDSEFAAIENL	LNSLTESGWISRDAS	90

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4753B-177-253-14	TAL1	P17542	MEITDGPHTKVVRRI	LAKLLNDQEEEGTQR	78
HR4753B-177-280-14	TAL1	P17542	MEITDGPHTKVVRRI	GGGGGGGGGAPPDDL	105
HR4753B-182-247-14	TAL1	P17542	MPHTKVVRRIFTNSR	MKYINFLAKLLNDQE	67
HR4753B-182-262-14	TAL1	P17542	MPHTKVVRRIFTNSR	EEGTQRAKTGKDPVV	82
HR4753B-182-262-Av6HT	TAL1	P17542	PHTKVVRRIFTNSRE	EEGTQRAKTGKDPVV	81
HR4753B-182-262-TEV	TAL1	P17542	PHTKVVRRIFTNSRE	EEGTQRAKTGKDPVV	81
HR4753B-182-287-14	TAL1	P17542	MPHTKVVRRIFTNSR	GGAPPDDLLQDVLSP	107
HR6460A-1-79-15	TAL2	Q16559	MTRKIFTNTRERWRQ	QTGVAAQGNILGLFP	79
HR6460A-1-84-15	TAL2	Q16559	MTRKIFTNTRERWRQ	AQGNILGLFPQGPHL	84
HR6460A-1-96-15	TAL2	Q16559	MTRKIFTNTRERWRQ	PHLPGLEDRTLLENY	96
HR6460A-34-96-15	TAL2	Q16559	PDKKLSKNETLRLAM	PHLPGLEDRTLLENY	63
hr464-14	TAX1BP1	Q86VP1	MTSFQEVPLQTSNFA	NSDMLVVTTKAGLLE	151
HR464-21	TAX1BP1	Q86VP1	MTSFQEVPLQTSNFA	NSDMLVVTTKAGLLE	151
HR7030-1-512-TEV	TAX1BP1	Q86VP1	TSFQEVPLQTSNFAH	TSASTVDVKPSPSAA	511
HR7030-1-529-TEV	TAX1BP1	Q86VP1	TSFQEVPLQTSNFAH	DFDIVTKGQVCEMTK	528
HR7030-1-588-TEV	TAX1BP1	Q86VP1	TSFQEVPLQTSNFAH	ENVKLELAEVQDNYK	587
HR7030-1-597-TEV	TAX1BP1	Q86VP1	TSFQEVPLQTSNFAH	VQDNYKELKRSLENP	596
HR7030A-15-466-TEV	TAX1BP1	Q86VP1	AHVIFQNVAKSYLPN	KFKECQRLQKQINKL	452
HR7030A-15-470-TEV	TAX1BP1	Q86VP1	AHVIFQNVAKSYLPN	CQRLQKQINKLSDQS	456
HR8311A-205-394-Av6HT	TBR1	Q16650	QVYLCNRPLWLKFHR	LKIDHNPFAKGFRDN	190
HR8240A-138-330-Av6HT	TBX18	O95935	APRVDLQGAELWKRF	RLKIDRNPFAKGFRD	193
HR7379A-91-283-TEV	TBX2	Q13207	SLKSLEPEDEVEDDP	DKITQLKIDNNPFAK	193

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7868A-127-326-Av6HT	TBX21	Q9UL17	LPAGLEVSGKLRVAL	QLKIDNNPFAKGFRE	200
HR7452A-91-277-NHT	TBX22	Q9Y458	DIQMELQGSELWKRF	QNQQITKLKIERNPF	187
HR7369A-99-311-TEV	TBX3	O15119	VEDDPKVHLEAKELW	LTLQSMRVFDERHKK	213
HR7232A-61-248-Av6HT	TBX4	P57082	EQTIENIKVGLHEKE	KITQLKIENNPFAKG	188
HR8313A-52-232-Av6HT	TBX5	Q99593	EGIKVFLHERELWLK	QNHKITQLKIENNPF	181
HR7389A-90-273-NHT	TBX6	O95947	GVSLSLENRELWKEF	QLKIAANPFAKGFRE	184
HR6334A-578-637-TEV	TCF12	Q99081	RRMANNARERLRVRD	AVAVILSLEQQVRER	60
HR6965A-1-144-Av6HT	TCF19	Q9Y242	MLPCFQLLRIGGGRG	DFAAITIPRSRGEAR	144
HR6965A-1-144-NHT	TCF19	Q9Y242	LPCFQLLRIGGGRGG	DFAAITIPRSRGEAR	143
HR8141A-73-167-Av6HT	TCF21	O43680	SQEGKQVQRNAANAR	PFMVAGKPESDLKEV	95
HR7160A-75-162-NHT	TCF23	Q7RTU1	SEASPENAARERSRV	LRYLHPLKKWPMRSR	88
HR7366A-178-588-Av6HT	TCF25	Q9BQ70	LYVEHRHLNPDTELK	DVTTQSVMGFDPLPP	411
HR4404E-550-609-TEV	TCF3	P15923	RRVANNARERLRVRD	AVSVILNLEQQVRER	60
HR4645C-565-624-TEV	TCF4	P15884	RRMANNARERLRVRD	AVAVILSLEQQVRER	60
HR8064A-27-105-TEV	TEAD1	P28347	IDNDAEGVWSPDIEQ	SSHIQVLARRKSRDF	79
HR7830A-40-115-TEV	TEAD2	Q15562	DAEGVWSPDIEQSFQ	SSHIQVLARRKSREI	76
HR7697A-27-104-TEV	TEAD3	Q99594	LDNDAEGVWSPDIEQ	VSSHIQVLARKKVRE	78
HR6976A-217-434-TEV	TEAD4	Q15561	RSVASSKLWMLEFSA	SEHGAQHHIYRLVKE	218
HR7931A-446-500-Av6HT	TERF2	Q15554	KKQKWTVEESEWVKA	MIKDRWRTMKRLGMN	55
HR7931A-446-500-TEV	TERF2	Q15554	KKQKWTVEESEWVKA	MIKDRWRTMKRLGMN	55
HR7939A-132-190-Av6HT	TERF2IP	Q9NYB0	GRIAFTDADDVAILT	SWQSLKDRYLKHLRG	59
HR7939A-132-190-TEV	TERF2IP	Q9NYB0	GRIAFTDADDVAILT	SWQSLKDRYLKHLRG	59

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8166A-153-218-TEV	TFAM	Q00059	GKPKRPRSAYNVYVA	AKEDETRYHNEMKSW	66
HR3078B-202-418-14	TFAP2A	P05549	GGVVNPNEVFCSVPG	EALKAMDKMYLSNNP	217
HR3078B-207-414-14	TFAP2A	P05549	PNEVFCSVPGRLSLL	NYLTEALKAMDKMYL	208
HR3162-15	TFAP2B	Q92481	MHSPPRDQAAIMLWK	GPGSKTGDKEEKHRK	460
HR7501-139-450-15	TFAP2C	Q92754	MRRDAYRRSDLLLPH	ADSNKTLEKMEKHRK	313
HR7501A-206-427-TEV	TFAP2C	Q92754	NLPCQKELVGAVMNP	QNYIKEALIVIDKSY	222
HR7501A-219-427-TEV	TFAP2C	Q92754	NPTEVFCSVPGRLSL	QNYIKEALIVIDKSY	209
HR7501B-128-430-Av6HT	TFAP2C	Q92754	LSGLEAGAVSARRDA	IKEALIVIDKSYMNP	303
HR7501B-128-450-Av6HT	TFAP2C	Q92754	LSGLEAGAVSARRDA	ADSNKTLEKMEKHRK	323
HR7501B-139-430-TEV	TFAP2C	Q92754	RRDAYRRSDLLLPHA	IKEALIVIDKSYMNP	292
HR7501B-206-450-TEV	TFAP2C	Q92754	NLPCQKELVGAVMNP	ADSNKTLEKMEKHRK	245
HR7501B-219-450-TEV	TFAP2C	Q92754	NPTEVFCSVPGRLSL	ADSNKTLEKMEKHRK	232
HR7272A-212-422-Av6HT	TFAP2E	Q6VUC0	TNPGEVFCSVPGRLS	YLLESLKGLDKMFLS	211
HR7122A-21-122-Av6HT	TFAP4	Q01664	EKEVIGGLCSLANIP	QQNTQLKRFIQELSG	102
HR7110A-303-394-15	TFCP2	Q12800	MLGEGNGSPNHQPEP	ALKGRMVRPRLTIYV	93
HR7110A-303-400-15	TFCP2	Q12800	MLGEGNGSPNHQPEP	VRPRLTIYVCQESLQ	99
HR7110A-303-404-15	TFCP2	Q12800	MLGEGNGSPNHQPEP	LTIYVCQESLQLREQ	103
HR7110A-332-395-15	TFCP2	Q12800	MEAQQWLHRNRFSTF	LKGRMVRPRLTIYVC	65
HR7110A-332-399-15	TFCP2	Q12800	MEAQQWLHRNRFSTF	MVRPRLTIYVCQESL	69
HR7110A-332-404-15	TFCP2	Q12800	MEAQQWLHRNRFSTF	LTIYVCQESLQLREQ	74
HR7022A-278-385-NHT	TFCP2L1	Q9NZI6	SPNSFGLGEGNASPT	MTIYVCQELEQNRVP	108
HR4671B-105-200-TEV	TFDP1	Q14186	RNRKGEKNGKGLRHF	KKEIKWIGLPTNSAQ	96

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7048A-121-215-TEV	TFDP2	Q14188	RRRVYDALNVLMAMN	QNQGPPALNSTIQLP	95
HR7261A-244-347-NHT	TFDP3	Q5H9I0	QRPLPNSVIHVPFII	AQGTFGGVFTTAGSR	104
HR4665C-333-388-14	TFE3	P19532	MISETEAKALLKERQ	PKSSDPEMRWNKGTI	57
HR4665C-333-443-14	TFE3	P19532	MISETEAKALLKERQ	ELELQAQIHGLPVPP	112
HR4665C-338-383-14	TFE3	P19532	MAKALLKERQKKDNH	LGTLIPKSSDPEMRW	47
HR4665C-338-438-14	TFE3	P19532	MAKALLKERQKKDNH	QLRQELELQAQIHG	102
HR7480A-223-309-NHT	TFEB	P19484	TDAESRALAKERQKK	SRELENHSRRLEMTN	87
HR4411B-151-237-14	TGIF1	Q15583	MDIPLDLSSSAGSGK	LPDMLRKDGKDPNQF	88
HR4411B-170-232-14	TGIF1	Q15583	MNLPKESVQILRDWL	ARRRLLPDMLRKDGK	64
HR4411B-170-252-14	TGIF1	Q15583	MNLPKESVQILRDWL	TISRRGAKISETSSV	84
HR4411B-171-248-14	TGIF1	Q15583	MLPKESVQILRDWLY	PNQFTISRRGAKISE	79
HR4411B-189-248-14	TGIF1	Q15583	MNAYPSEQEKALLSQ	PNQFTISRRGAKISE	61
HR4411C-171-241-14	TGIF1	Q15583	MLPKESVQILRDWLY	LRKDGKDPNQFTISR	72
HR4393-12-199-15	TGIF2	Q9GZN2	LLSLAGKRKRRGNLP	PTPPEQDKEDFSSFQ	188
HR4393-12-223-15	TGIF2	Q9GZN2	LLSLAGKRKRRGNLP	AAEMELQKQQDPSLP	212
HR4393-17-220-15	TGIF2	Q9GZN2	GKRKRRGNLPKESVK	LQRAAEMELQKQQDP	204
HR4393-6-199-15	TGIF2	Q9GZN2	LGEDEGLLSLAGKRK	PTPPEQDKEDFSSFQ	194
HR4393-6-223-15	TGIF2	Q9GZN2	LGEDEGLLSLAGKRK	AAEMELQKQQDPSLP	218
HR7881A-51-127-TEV	TGIF2LX	Q8IUE1	KKRKGNLPAESVKIL	LQQRRNDPIIGHKTG	77
HR8232A-51-127-NHT	TGIF2LY	Q8IUE0	KKRKGNLPAESVKIL	LQQRRNDPIIGHKTG	77
HR8232A-51-127-TEV	TGIF2LY	Q8IUE0	KKRKGNLPAESVKIL	LQQRRNDPIIGHKTG	77
HR7047A-1-87-TEV	THAP1	Q9NVV9	VQSCSAYGCKNRYDK	KENAVPTIFLCTEPH	86

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR1517A-1-91-Av6HT	THAP10	Q9P2Z0	PARCVAAHCGNTTKS	QRLRLVAGAVPTLHR	90
HR7799A-1-67-TEV	THAP11	Q96EK4	PGFTCCVPGCYNNSH	QPTTGHRLCSVHFQG	66
HR7799A-1-72-TEV	THAP11	Q96EK4	PGFTCCVPGCYNNSH	HRLCSVHFQGGRKTY	71
HR7799A-1-82-Av6HT	THAP11	Q96EK4	PGFTCCVPGCYNNSH	GRKTYTVRVPTIFPL	81
HR8301A-1-87-TEV	THAP2	Q9H0W7	PTNCAAAGCATTYNK	MDAVPTIFDFCTHIK	86
HR7028A-1-83-NHT	THAP5	Q7Z6K1	PRYCAAICCKNRRGR	RWGIRYLKQTAVPTI	82
HR8415A-1-149-Av6HT	THAP6	Q8TBB0	VKCCSAIGCASRCLP	QFIFEHSYSVMDSPK	148
HR7818A-1-92-Av6HT	THAP7	Q9BT49	PRHCSAAGCCTRDTR	ISGYHRLKEGAVPTI	91
HR6978A-1-82-15	THAP8	Q8NA92	MPKYCRAPNCSNTAG	QWRWGVRYLRPDAVP	82
HR6978A-1-87-15	THAP8	Q8NA92	MPKYCRAPNCSNTAG	VRYLRPDAVPSIFSR	87
HR6978A-16-87-15	THAP8	Q8NA92	MRLGADNRPVSFYKF	VRYLRPDAVPSIFSR	73
HR6978A-21-82-15	THAP8	Q8NA92	MNRPVSFYKFPLKDG	QWRWGVRYLRPDAVP	63
HR7271A-1-88-NHT	THAP9	Q9H5L6	TRSCSAVGCSTRDTV	YGIRRKLKKGAVPSV	87
HR7130A-202-461-15	THRB	P10828	MEELQKSIGHKPEPT	PTELFPPLFLEVFED	261
HR7130A-202-461-Av6HT	THRB	P10828	EELQKSIGHKPEPTD	PTELFPPLFLEVFED	260
HR7130A-202-461-TEV	THRB	P10828	EELQKSIGHKPEPTD	PTELFPPLFLEVFED	260
HR7130B-104-206-15	THRB	P10828	MDELCVVCGDKATGY	IEENREKRRREELQK	104
HR7130B-104-206-Av6HT	THRB	P10828	DELCVVCGDKATGYH	IEENREKRRREELQK	103
HR7130B-104-206-TEV	THRB	P10828	DELCVVCGDKATGYH	IEENREKRRREELQK	103
HR6921A-74-139-NHT	TIGD3	Q6B0B8	SKYSGIDEALLCWYH	VRWKRRNNVGFGARH	66
HR7457A-14-77-NHT	TIGD4	Q8IY51	TVKKKKSLSIEEKID	VLEAFESLRFDPKRK	64
HR7206A-68-132-NHT	TIGD6	Q17RP2	KRMRSALYDDIDKAV	QASVGWLNRFRDRHG	65

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7729A-62-136-NHT	TIGD7	Q6NT04	PLVGAEKRKRTTGAK	STGWLFRFRNRHAIG	75
HR7316A-298-418-NHT	TIPARP	Q7Z3E1	NDRMRMKYGGQEFWA	LFRSCFILLPYLQTL	121
HR7535A-206-260-TEV	TLX1	P31314	TSFTRLQICELEKRF	KTWFQNRRTKWRRQT	55
HR6480A-162-216-TEV	TLX2	O43763	TSFSRSQVLELERRF	KTWFQNRRTKWRRQT	55
HR7241A-171-225-TEV	TLX3	O43711	TSFSRVQICELEKRF	KTWFQNRRTKWRRQT	55
HR3551B-1-370-TEV	TNFAIP3	P21580	AEQVLPQALYLSNMR	WQENSEQGRREGHAQ	369
HR8218A-34-319-Av6HT	TOE1	Q96GM8	VPVVDVQSNNFKEMW	AYGWCPLGPQCPQSH	286
HR5174A-643-706-NHT	TOP3A	Q13472	QQEDIYPAMPEPIRK	PDSVLEASRDSSVCP	64
HR8243A-244-339-NHT	TOX	O94900	GKKPKTPKKKKKKDP	LAAYRASLVSKSYSE	96
HR8243A-244-339-TEV	TOX	O94900	GKKPKTPKKKKKKDP	LAAYRASLVSKSYSE	96
HR7258A-238-302-TEV	TOX2	Q96NM4	VASMWDSLGEEQKQA	STQANPPAKMLPPKQ	65
HR7680A-238-335-TEV	TOX3	O15405	GKKPKTPKKKKKKDP	AYRASLVSKAAAESA	98
HR7250A-206-291-TEV	TOX4	O94842	GKKQKAPKKRKKKDP	EAAKKEYLKALAAYK	86
HR6989-94-312-15-TEV	TP53	P04637	MLSSSVPSQKTYQGS	ELPPGSTKRALPNNT	221
HR6989-94-312-R175H-15-TEV	TP53	P04637	MLSSSVPSQKTYQGS	ELPPGSTKRALPNNT	221
HR6989A-20-73-Av6HT	TP53	P04637	SDLWKLLPENNVLSP	GPDEAPRMPEAAPPV	54
HR6989A-20-73-TEV	TP53	P04637	SDLWKLLPENNVLSP	GPDEAPRMPEAAPPV	54
HR3500C-14	TP63	Q9H3D4	MDALSPSPAIPSNTD	GTKRPFRQNTHGIQM	232
HR3500C-15	TP63	Q9H3D4	MDALSPSPAIPSNTD	GTKRPFRQNTHGIQM	232
HR3500D-540-614-TEV	TP63	Q9H3D4	PPPYPTDCSIVSFLA	GILDHRQLHEFSSPS	75
HR3466-110-636-14	TP73	O15350	MSPAPVIPSNTDYPG	RKQPIKEEFTEAEIH	528
HR3466-114-636-14	TP73	O15350	MVIPSNTDYPGPHHF	RKQPIKEEFTEAEIH	524

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR3466D-14	TP73	O15350	MSPAPVIPSNTDYPG	KADEDHYREQQALNE	209
HR3466D-15	TP73	O15350	MSPAPVIPSNTDYPG	KADEDHYREQQALNE	209
HR3466E-487-554-TEV	TP73	O15350	YHADPSLVSFLTGLG	TIWRGLQDLKQGHDY	68
HR8230A-78-139-TEV	TRAFD1	O14545	HEETEGPLRLAVCQH	VKDLKTHPEVCGREG	62
HR4455D-876-951-14	TRERF1	Q96PN7	MCHPLANYHYAGSDK	LGRKHRTRLAEIIDD	77
HR4455D-881-945-14	TRERF1	Q96PN7	MNYHYAGSDKWTSLE	WKKIMRLGRKHRTRL	66
HR4455E-773-1200-15	TRERF1	Q96PN7	MQTVDVEPRINIGLR	LDDQDSVLLQGDAEL	429
HR4455E-778-1200-15	TRERF1	Q96PN7	MEPRINIGLRFQAEI	LDDQDSVLLQGDAEL	424
HR4455F-773-841-15	TRERF1	Q96PN7	MQTVDVEPRINIGLR	ENLLNLCCSSALPGG	70
HR4455F-778-836-15	TRERF1	Q96PN7	MEPRINIGLRFQAEI	LQQRVENLLNLCCSS	60
HR7441A-22-107-NHT	TRIM23	P36406	GTAVVKVLECGVCED	FALLELLERLQNGPI	86
HR7486A-13-98-NHT	TRIM3	O75382	QPMDKQFLVCSICLD	SLMEAMQQAPDGAHD	86
HR7466A-4-90-TEV	TRIM32	Q13049	AAASHLNLDALREVL	LTDNLTVLKIIDTAG	87
HR5056A-48-428-Av6HT	TRIT1	Q9H3H1	GGEIVSADSMQVYEG	IKSKSHLNQLKKRRR	381
HR7683A-316-386-15	TSC22D4	Q9Y3Q8	MVGIDNKIEQAMDLV	EQLAQLPSSGVPRLG	72
HR7683A-320-381-15	TSC22D4	Q9Y3Q8	MNKIEQAMDLVKSHL	ALASPEQLAQLPSSG	63
HR7683A-320-395-Av6HT	TSC22D4	Q9Y3Q8	NKIEQAMDLVKSHLM	GVPRLGPPAPNGPSV	76
HR8019A-232-335-TEV	TSHZ1	Q6ZSZ6	DKDSEKTKRWSKPRK	EPAGMAAEVALSESA	104
HR7516A-824-951-NHT	TSHZ2	Q9NRE2	DVRRFEDVSSEVSTL	TPSTYISHLESHLGF	128
HR6901A-200-303-TEV	TSHZ3	Q63HK5	SSKLYGSIFTGASKF	DLSVHMIKTKHYQKV	104
HR7321A-615-661-Av6HT	TTF1	Q15361	NYKGRYSEGDTEKLK	ARSSLSVALKFSQIS	47
HR7321A-615-669-Av6HT	TTF1	Q15361	NYKGRYSEGDTEKLK	LKFSQISSQRNRGAW	55

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7321A-615-678-Av6HT	TTF1	Q15361	NYKGRYSEGDTEKLK	RNRGAWSKSETRKLI	64
HR7321A-615-687-Av6HT	TTF1	Q15361	NYKGRYSEGDTEKLK	ETRKLIKAVEEVILK	73
HR7321A-615-697-NHT	TTF1	Q15361	NYKGRYSEGDTEKLK	EVILKKMSPQELKEV	83
HR8382A-244-506-NHT	TUB	P50607	GISSSMSFDEDEEDE	QSYVLNFHGRVTQAS	263
HR8277A-291-536-TEV	TULP1	O00294	PREFVLRPAPQGRTV	LCALQAFAIALSSFD	246
HR7732A-263-520-Av6HT	TULP2	O00295	SPCPGLEEDMEAYVL	FSPLQAFSICLSSFN	258
HR6409A-100-175-14	TWIST1	Q15672	PQSYEELQTQRVMAN	QVLQSDELDSKMASC	76
HR6409A-102-171-14	TWIST1	Q15672	SYEELQTQRVMANVR	DFLYQVLQSDELDSK	70
HR7529A-43-146-TEV	U2AF1	Q01081	SQTIALLNIYRNPQN	NRWFNGQPIHAELSP	104
HR7415B-479-562-Av6HT	UBTF	P17480	MEMTWNNMEKKEKLM	NGELNHLPLKERMVE	85
HR7415B-479-562-TEV	UBTF	P17480	EMTWNNMEKKEKLMW	NGELNHLPLKERMVE	84
HR8089A-89-170-Av6HT	UNCX	A6NJT0	KLSDSGDPDKESPGC	RRAKWRKKENTKKGP	82
HR7768A-197-260-TEV	USF1	P22415	DEKRRAQHNEVERRR	KACDYIQELRQSNHR	64
HR6458A-220-346-15	USF2	Q15853	PYSPKIDGTRTPRDE	LQQHNLEMVGEGTRQ	127
HR6458A-226-279-15	USF2	Q15853	DGTRTPRDERRRAQH	CNADNSKTGASKGGI	54
HR6458A-226-346-15	USF2	Q15853	DGTRTPRDERRRAQH	LQQHNLEMVGEGTRQ	121
HR6458B-226-330-15	USF2	Q15853	DGTRTPRDERRRAQH	QQIEELKNENALLRA	105
HR6458B-231-325-15	USF2	Q15853	PRDERRRAQHNEVER	NELLRQQIEELKNEN	95
HR8005-106-565-15	USP39	Q53GS9	MPYLDTINRSVLDFD	IWKRRDNDETNQQGA	461
HR8005A-189-554-15	USP39	Q53GS9	MITYVLKPTFTKQQI	QMITLSEAYIQIWKR	367
HR8005A-189-565-15	USP39	Q53GS9	MITYVLKPTFTKQQI	IWKRRDNDETNQQGA	378
HR8005A-194-549-15	USP39	Q53GS9	MKPTFTKQQIANLDK	TDILPQMITLSEAYI	357

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8005A-194-555-15	USP39	Q53GS9	MKPTFTKQQIANLDK	MITLSEAYIQIWKRR	363
HR8005B-210-555-15	USP39	Q53GS9	MKLSRAYDGTTYLPG	MITLSEAYIQIWKRR	347
HR8005C-106-183-TEV	USP39	Q53GS9	PYLDTINRSVLDFDF	TLKFYCLPDNYEIID	78
HR8005C-129-183-TEV	USP39	Q53GS9	SHINAYACLVCGKYF	TLKFYCLPDNYEIID	55
HR6997A-81-167-NHT	VAX1	Q5SQQ9	DAKGSIREIILPKGL	TKQKKDQGKDSELRS	87
HR8032A-81-165-Av6HT	VAX2	Q9UIW0	VRDAKGTIREIVLPK	QNRRTKQKKDQSRDL	85
HR7564A-21-427-Av6HT	VDR	P11473	PRICGVCGDRATGFH	KLTPLVLEVFGNEIS	407
HR7564A-21-427-TEV	VDR	P11473	PRICGVCGDRATGFH	KLTPLVLEVFGNEIS	407
HR7564B-16-125-15	VDR	P11473	MFDRNVPRICGVCGD	KEEEALKDSLRPKLS	111
HR7564B-16-125-Av6HT	VDR	P11473	FDRNVPRICGVCGDR	KEEEALKDSLRPKLS	110
HR7564B-16-125-TEV	VDR	P11473	FDRNVPRICGVCGDR	KEEEALKDSLRPKLS	110
HR7703A-97-158-Av6HT	VENTX	O95231	AFTMEQVRTLEGVFQ	MKHKRQMQDPQLHSP	62
HR7928A-148-214-Av6HT	VSX1	Q9NZR4	EDRNDLKASPTLGKR	KTELPEDRIQVWFQN	67
HR8065A-153-211-Av6HT	VSX2	P58304	TIFTSYQLEELEKAF	QNRRAKWRKREKCWG	59
HR7106A-20-70-TEV	VTN	P04004	DQESCKGRCTEGFNV	AECKPQVTRGDVFTM	51
HR7713A-1017-1084-TEV	WDHD1	O75717	RPKTGFQMWLEENRS	KGETASEGTEAKKRK	68
HR7541A-815-891-NHT	WHSC1	O96028	AHFTARKGKRHHAHV	KLHFQDIIWVKLGNY	77
HR8130A-318-438-TEV	WT1	P19544	HSTGYESDNHTTPIL	HQRRHTGVKPFQCKT	121
HR3172-15	XBP1	P17861	MVVVAAAPNPADGTP	CQWGRHQPSWKPLMN	261
HR8228A-98-208-TEV	XPA	P23025	EFDYVICEEGKEFM	WGSQEALEEAKEVRQ	110
HR8027A-52-129-TEV	YBX1	P67809	KKVIATKVLGTVKWF	EGEKGAEAANVTGPG	78
HR7254A-87-164-TEV	YBX2	Q9Y2T7	KPVLAIQVLGTVKWF	EGEKGAEATNVTGPG	78

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7538A-193-335-NHT	YEATS2	Q9ULM3	RNADLTDETSRLFVK	AETVVDVELHRHSLG	143
HR8298A-14-147-15	YEATS4	O95619	MGRVKGVTIVKPIVY	TVVSEFYDEMIFQDP	135
HR8137A-293-414-TEV	YY1	P25490	PRTIACPHKGCTKMF	LKSHILTHAKAKNNQ	122
HR8207A-251-371-TEV	YY2	O15391	PKTVPCSYSGCEKMF	NLKTHILTHVKTKNN	121
HR7328A-23-83-TEV	ZBED1	O96006	SKVWKYFGFDTNAEG	KNHPEEFCEFVKSNT	61
HR7606A-49-119-TEV	ZBED2	Q9BTP6	NKGTRFSEAWEYFHL	MHREELEKSGHGQAG	71
HR8174A-11-69-15	ZBP1	Q9H171	MEGHLEQRILQVLTE	ELKVSLTSPATWCLG	60
HR8174A-26-69-15	ZBP1	Q9H171	MGSPVKLAQLVKECQ	ELKVSLTSPATWCLG	45
HR8174A-6-74-15	ZBP1	Q9H171	MADPGREGHLEQRIL	LTSPATWCLGGTDPE	70
HR7903A-21-117-Av6HT	ZBTB1	Q9Y2K1	GFLCDCCIAIDDIYF	YLQLYNVPDCLEDIQ	97
HR6940A-194-334-NHT	ZBTB11	O95625	PKHCQAVLKQLNEQR	KKGEVQTVASTQDLR	141
HR6940B-764-842-Av6HT	ZBTB11	O95625	RGYHCTQCEKSFFEA	GKEFYEKALFRRHVK	79
HR4454-14	ZBTB12	Q9Y330	MASGVEVLRFQLPGH	NVLEASVAEINVLIR	459
HR4454C-1-132-14	ZBTB12	Q9Y330	MASGVEVLRFQLPGH	KCRNALSQFIEPKIG	132
HR4454C-1-137-14	ZBTB12	Q9Y330	MASGVEVLRFQLPGH	LSQFIEPKIGLKEDG	137
HR4454C-1-151-14	ZBTB12	Q9Y330	MASGVEVLRFQLPGH	GVSEASLVSSISATK	151
HR4454D-328-447-14	ZBTB12	Q9Y330	MNPLKNIKCTKCPEV	HLKEQHGKTTAENVL	121
HR4454E-361-421-14	ZBTB12	Q9Y330	MCPRCGKQFNHSSNL	NLHSGARPYRCSYCD	62
HR4454E-366-418-14	ZBTB12	Q9Y330	MKQFNHSSNLNRHMN	DHLNLHSGARPYRCS	54
HR3471-145-673-15	ZBTB16	Q05516	MEEDRKARYLKNIFI	PDWRIEKTYLYLCYV	530
HR3471-150-673-15	ZBTB16	Q05516	MARYLKNIFISKHSS	PDWRIEKTYLYLCYV	525
HR3471-189-673-15	ZBTB16	Q05516	MTSFGLSAMSPTKAA	PDWRIEKTYLYLCYV	486

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR3471-194-673-15	ZBTB16	Q05516	MSAMSPTKAAVDSLM	PDWRIEKTYLYLCYV	481
HR4581F-2-115-TEV	ZBTB17	Q13105	DFPQHSQHVLEQLNQ	MQDITACHALKSLA	114
HR7182A-1-107-15	ZBTB2	Q8N680	MDLANHGLILLQQLN	VRLEQGIKFLHAYPL	107
HR7182A-1-113-15	ZBTB2	Q8N680	MDLANHGLILLQQLN	IKFLHAYPLIQEASL	113
HR7182B-1-117-15	ZBTB2	Q8N680	MDLANHGLILLQQLN	HAYPLIQEASLASQG	117
HR7182B-1-147-15	ZBTB2	Q8N680	MDLANHGLILLQQLN	YGIQIADHQLRQATK	147
HR7182B-1-152-15	ZBTB2	Q8N680	MDLANHGLILLQQLN	ADHQLRQATKIASAP	152
HR7182C-227-390-15	ZBTB2	Q8N680	MSDEQPASLTIAHVK	SHWREHMYIHTGKPF	165
HR7182C-232-385-15	ZBTB2	Q8N680	MASLTIAHVKPSIMK	KFIQKSHWREHMYIH	155
HR7182C-245-390-15	ZBTB2	Q8N680	MKRNGSFPKYYACHL	SHWREHMYIHTGKPF	147
HR7182C-248-385-15	ZBTB2	Q8N680	MGSFPKYYACHLCGR	KFIQKSHWREHMYIH	139
HR7182C-252-385-15	ZBTB2	Q8N680	MKYYACHLCGRRFTL	KFIQKSHWREHMYIH	135
HR8336A-81-201-Av6HT	ZBTB20	Q9HC78	INLHNFSNSVLETLN	DECTRIVSQNVGDVF	121
HR7741A-30-151-NHT	ZBTB22	O15209	AVVHVSFPEVTSALL	WHIVDKCTELLREGR	122
HR7877A-1-114-15	ZBTB25	P24278	MDTASHSLVLLQQLN	RFLHADYLSHIATEM	114
HR7877A-1-120-15	ZBTB25	P24278	MDTASHSLVLLQQLN	YLSHIATEMNQVFSP	120
HR7877A-1-138-15	ZBTB25	P24278	MDTASHSLVLLQQLN	QSSNLYGIQISTTQK	138
HR7877A-1-144-15	ZBTB25	P24278	MDTASHSLVLLQQLN	GIQISTTQKTVVKQG	144
HR7877B-231-376-15	ZBTB25	P24278	MTENSVKIHLCHYCG	SQLLEHMYTHKGKSY	147
HR7877B-236-373-15	ZBTB25	P24278	MKIHLCHYCGERFDS	PRKSQLLEHMYTHKG	139
HR7877B-244-373-15	ZBTB25	P24278	MGERFDSRSNLRQHL	PRKSQLLEHMYTHKG	131
HR7877B-261-373-15	ZBTB25	P24278	MVSGSLPFGVPASIL	PRKSQLLEHMYTHKG	114

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7877B-270-376-15	ZBTB25	P24278	MPASILESNDLGEVH	SQLLEHMYTHKGKSY	108
HR7877B-275-373-15	ZBTB25	P24278	MESNDLGEVHPLNEN	PRKSQLLEHMYTHKG	100
HR7422A-1-129-NHT	ZBTB26	Q9HCK0	SERSDLLHFKFENYG	IVERCTQALWKFIKP	128
HR6960A-46-179-NHT	ZBTB3	Q9H5J0	PSWGTMEFPEHSQQL	CKRRLQARALAEADS	134
HR7977A-1-110-Av6HT	ZBTB32	Q9Y2Y4	SLPPIRLPSPYGSDR	AARALGVQSLEEACW	109
HR7008A-1-116-TEV	ZBTB33	Q86T24	ESRKLISATDIQYSG	IKSGQLLGVKFIAEL	115
HR6893A-1-124-NHT	ZBTB34	Q8NCN2	DSSSFIQFDVPEYSS	QMQCVIDKCTQILES	123
HR7018A-1-125-NHT	ZBTB37	Q5TC79	EKGGNIQLEIPDFSN	MQHIIDKCTQILEGI	124
HR7837A-13-139-NHT	ZBTB38	Q8NAP3	DFHSDTVLSILNEQR	RNFSNSPGPYVFCIT	127
HR7896A-1-125-Av6HT	ZBTB39	O15060	GMRIKLQSTNHPNNL	MEDLLQACHSTFPDL	124
HR7527A-1-112-NHT	ZBTB40	Q9NUA8	ELPNYSRQLLQQLYT	DSLQMFDVAVSCKNL	111
HR8293A-24-183-Av6HT	ZBTB41	Q5SVQ8	EGNVAVECDQVTYTH	DAVKLLNNENVAPFH	160
HR7772A-367-467-TEV	ZBTB43	O43298	SATDKLYPCQCGKSF	SYEAAKAEQNTTEAN	101
HR7772B-1-126-15	ZBTB43	O43298	MEPGTNSFRVEFPDF	MWHVVDKCTEVLEGN	126
HR7772B-1-137-15	ZBTB43	O43298	MEPGTNSFRVEFPDF	LEGNPTVLCQKLNHG	137
HR7772C-10-126-Av6HT	ZBTB43	O43298	VEFPDFSSTILQKLN	MWHVVDKCTEVLEGN	117
HR7772C-10-131-Av6HT	ZBTB43	O43298	VEFPDFSSTILQKLN	DKCTEVLEGNPTVLC	122
HR7772C-29-131-15	ZBTB43	O43298	MQGQLCDVSIVVQGH	DKCTEVLEGNPTVLC	104
HR7772C-29-137-15	ZBTB43	O43298	MQGQLCDVSIVVQGH	LEGNPTVLCQKLNHG	110
HR7772C-33-126-15	ZBTB43	O43298	MCDVSIVVQGHIFRA	MWHVVDKCTEVLEGN	95
HR7772C-33-131-15	ZBTB43	O43298	MCDVSIVVQGHIFRA	DKCTEVLEGNPTVLC	100
HR8333A-1-128-15	ZBTB44	Q8NCP5	MGVKTFTHSSSSHSQ	FSVASTCSEFMKSSI	128

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8333A-1-133-15	ZBTB44	Q8NCP5	MGVKTFTHSSSSHSQ	TCSEFMKSSILWNTP	133
HR7817A-6-125-NHT	ZBTB45	Q96K62	AVHHIHLQNFSRSLL	IQTVIDECTQIIARA	120
HR7445A-291-405-NHT	ZBTB48	P10074	VECPTCHKKFLSKYY	KDLQSHMIKLHGAPK	115
HR7445A-291-405-TEV	ZBTB48	P10074	VECPTCHKKFLSKYY	KDLQSHMIKLHGAPK	115
HR7445B-4-120-TEV	ZBTB48	P10074	SFVQHSVRVLQELNK	EAVELCQSFKPKTSV	117
HR7910A-1-125-Av6HT	ZBTB49	Q6ZSB9	DPVATHSCHLLQQLH	SLCHTFLKSATVVQP	124
HR7620A-1-125-NHT	ZBTB5	O15062	DFPGHFEQIFQQLNY	VVKACKHYLTTRTLP	124
HR8300A-1-134-Av6HT	ZBTB6	Q15916	AAESDVLHFQFEQQG	TEALSKYLEIDLSMK	133
HR4695C-9-128-TEV	ZBTB7A	O95365	IGIPFPDHSSDILSG	LEIPAVSHVCADLLD	120
HR8347A-1-143-Av6HT	ZBTB7B	O15156	GSPEDDLIGIPFPDH	EIPCVIAACMEILQG	142
HR7365A-6-129-NHT	ZBTB7C	A1YPR0	DELIGIPFPNHSSEV	EIQCIVNVCLEIMEP	124
HR8095A-1-121-Av6HT	ZBTB8A	Q96BR9	EISSHQSHLLQQLNE	MTDVISVCKTFIKSS	120
HR7919A-1-131-Av6HT	ZBTB8B	Q8NAP8	EMQSYYAKLLGELNE	YIRSSLDICRKMEKE	130
HR7153A-20-140-NHT	ZBTB9	Q96C00	PRTIQIEFPQHSSSL	QMWQVVDQCSEILRE	121
HR6996-34-350-Av6HT	ZC3H15	Q8WU90	KKGAKQQKFIKAVTH	LYIPRDVDETGITVA	317
HR7121A-417-503-TEV	ZC3H4	Q9UPT8	ELPKKRELCKFYITG	GAEDEKEVEELKKQG	87
HR7136A-892-947-TEV	ZC3H7A	Q8IWR0	QYCWQHRFPTGYFSI	WEERRDALKMKLNKA	56
HR6981A-218-283-NHT	ZC3H8	Q8N5P1	EIEKKKEMCKFYVQG	APLTPETQELLAKVL	66
HR8421A-724-896-TEV	ZC3HAV1	Q7Z2W4	SSKKYKLSEIHHLHP	KDQVYPQYVIEYTED	173
HR4840-1-454-14	ZCCHC4	Q9H5U6	MAASRNGFEAVEAEG	GPKHGCFICGELDHK	454
HR4840-1-459-14	ZCCHC4	Q9H5U6	MAASRNGFEAVEAEG	CFICGELDHKRSTCP	459
HR4840-102-513-14	ZCCHC4	Q9H5U6	MLSRTQCVERYLKFI	RRKKRRERAHQYLGS	413

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR4840-41-454-14	ZCCHC4	Q9H5U6	MPHGPTLLFVKVTQG	GPKHGCFICGELDHK	415
HR4840-41-459-14	ZCCHC4	Q9H5U6	MPHGPTLLFVKVTQG	CFICGELDHKRSTCP	420
HR4840-41-513-14	ZCCHC4	Q9H5U6	MPHGPTLLFVKVTQG	RRKKRRERAHQYLGS	474
HR7192A-927-1161-Av6HT	ZCCHC6	Q5VYS8	SLKEENVCEEKNSPV	LCYTMKVFTKMCDIG	235
HR4656C-161-443-14	ZEB1	P37275	MGTPDAFSQLLTCPY	LENNQANLASKEQET	284
HR4656C-165-439-14	ZEB1	P37275	MAFSQLLTCPYCDRG	IRQVLENNQANLASK	276
HR4656D-885-1017-14	ZEB1	P37275	MNDSDSTPPKKKMRK	PEILSNEHVGARASP	134
HR4656D-885-992-14	ZEB1	P37275	MNDSDSTPPKKKMRK	HMNHRYSYCKREAEE	109
HR4656D-890-1014-14	ZEB1	P37275	MTPPKKKMRKTENGM	EAGPEILSNEHVGAR	126
HR4656D-890-987-14	ZEB1	P37275	MTPPKKKMRKTENGM	GSYSQHMNHRYSYCK	99
HR4656D-897-1014-14	ZEB1	P37275	MRKTENGMYACDLCD	EAGPEILSNEHVGAR	119
HR4656D-897-987-14	ZEB1	P37275	MRKTENGMYACDLCD	GSYSQHMNHRYSYCK	92
HR4656E-583-642-TEV	ZEB1	P37275	SPSQPPLKNLLSLLK	WFEKMQAGQISVQSS	60
HR4589D-647-707-Av6HT	ZEB2	O60315	MSPINPYKDHMSVLK	EQRKVYQYSNSRSPS	62
HR4589D-647-707-TEV	ZEB2	O60315	SPINPYKDHMSVLKA	EQRKVYQYSNSRSPS	61
HR6404A-128-196-TEV	ZFAND5	O76080	SPSVSQPSTSQSEEK	DKHNCPYDYKAEAAA	69
HR8363-123-190-Av6HT	ZFAND6	Q6FIF0	SVSDTAQQPSEEQSK	SDVHNCSYNYKADAA	68
HR7859A-697-770-TEV	ZFHX2	Q9C0A1	LGSSSDSLPTSPPPD	DTHRCKLCCYGTQLK	74
HR7545B-1398-1452-Av6HT	ZFHX3	Q15911	VYKYRCNQCSLAFKT	FRTFQALKKHLETSH	55
HR7545B-1398-1483-Av6HT	ZFHX3	Q15911	VYKYRCNQCSLAFKT	ANGDLLAMGDPTLAE	86
HR7545C-2943-2998-Av6HT	ZFHX3	Q15911	RPGQKRFRTQMTNLQ	GLPKRVVQVWFQNAR	56
HR7545C-2943-3010-Av6HT	ZFHX3	Q15911	RPGQKRFRTQMTNLQ	NARAKEKKSKLSMAK	68

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7545C-2954-3005-Av6HT	ZFHX3	Q15911	TNLQLKVLKSCFNDY	QVWFQNARAKEKKSK	52
HR7545C-2954-3010-Av6HT	ZFHX3	Q15911	TNLQLKVLKSCFNDY	NARAKEKKSKLSMAK	57
HR7573D-2953-3003-TEV	ZFHX4	Q86UP3	TPTMQECEMLGNEIG	INIGKPFMINQGGTE	51
HR8105A-360-407-Av6HT	ZFP1	Q6P2D0	TFSQRSTLRLHLRIH	SRLSVHQRVHIGEKP	48
HR7592A-1509-1806-Av6HT	ZFP106	Q9H2Y7	SSEISSEPGDDDEPT	MIYTGCYDGSIQAVR	298
HR7696A-237-310-NHT	ZFP112	O9UJU3	GEDIMKVSLLNQESI	NYSSLLHIHQNIERE	74
HR7929A-142-207-Av6HT	ZFP14	Q9HCL3	QVKITSEKMTTYKRH	IHTGEKPYKCKECGQ	66
HR7835A-8-127-NHT	ZFP161	O43829	SETIKYNDDDHKTLF	QILGIRFLDKLCSQK	120
HR7794A-82-137-NHT	ZFP2	Q6ZN57	DATQNSELIKTQRMF	IHTGEKPYKCNVCGK	56
HR7491A-224-296-NHT	ZFP28	Q8NHY6	LFETQPGLVTIKNLA	QEKEPWMVKRELTGS	73
HR7037A-395-483-Av6HT	ZFP3	Q96NJ6	CFECGKAFRRTSHLI	RIHTGEKPYECQECQ	89
HR7776A-306-378-NHT	ZFP30	Q9Y2G7	AFLCSTGLRLHHKLH	SRGYHLTLHQRIHTG	73
HR4743B-99-171-TEV	ZFP36	P26651	TTPSRYKTELCRTFS	PYGSRCHFIHNPSED	73
HR7685A-112-181-TEV	ZFP36L1	Q07352	SSRYKTELCRPFEEN	CPYGPRCHFIHNAEE	70
HR7167A-151-220-TEV	ZFP36L2	P47974	STRYKTELCRPFEES	CPYGPRCHFIHNADE	70
HR7105A-571-623-NHT	ZFP37	O9Y6Q3	KPYECNECEKAFNAK	TFKQNASLTKHVKTH	53
HR7063A-83-140-NHT	ZFP41	Q8N8Y5	RKKPYECSECGRIFK	HSSDVTKHQRTHTGE	58
HR7124A-165-241-NHT	ZFP42	Q96MM3	PKQLAEFARKKPPIN	VESSKLKRHFLVHTG	77
HR7554A-339-397-TEV	ZFP64	Q9NPA5	SEHPEKCSECSYSCS	PSNLSKHMKKFHGDM	59
HR7612A-318-373-Av6HT	ZFP82	Q8N141	AFLCGSGLRVHHKLH	IHTGEKPYECKECGK	56
HR7734A-483-537-NHT	ZFP90	Q8TF47	AFSQQAISHPGEKPY	KPYECNECGEAFSRR	55
HR7784A-364-427-15	ZFP91	Q96JP5	MKHHTDQRDYICEYC	ASLNWHMKKHDADSF	65

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7784A-370-422-15	ZFP91	Q96JP5	MRDYICEYCARAFKS	TCRQKASLNWHMKKH	54
HR7784A-370-456-15	ZFP91	Q96JP5	MRDYICEYCARAFKS	KDSVVAHKAKSHPEV	88
HR7784B-313-454-Av6HT	ZFP91	Q96JP5	CEMEGCGTVLAHPRY	EKKDSVVAHKAKSHP	142
HR7665A-150-204-NHT	ZFP92	A6NM28	KRYLCQQCGKAFSRS	RRSFALLEHQRIHSG	55
HR7876A-297-383-Av6HT	ZFPM1	Q8IX07	CRKSCPSASSLEIHM	TNHMVCQPGSKGEIY	87
HR7512A-673-743-NHT	ZFX	P17010	RPSELKKHVAAHKGK	RQQSELKKHMKTHSG	71
HR8053A-728-784-Av6HT	ZFYVE20	Q9H1K0	PEAEEPIEEELLLQQ	RELKHTLAKQKGGTD	57
HR8053B-133-256-15	ZFYVE20	Q9H1K0	MAFDRTNTESAKIRA	DEKDDDRIRCCTHCK	125
HR8053B-133-276-15	ZFYVE20	Q9H1K0	MAFDRTNTESAKIRA	REQQIDEKEHTPDIV	145
HR8053B-139-251-15	ZFYVE20	Q9H1K0	MTESAKIRAIEKSVV	VSSVLDEKDDDRIRC	114
HR8053B-139-273-15	ZFYVE20	Q9H1K0	MTESAKIRAIEKSVV	LLKREQQIDEKEHTP	136
HR8053B-150-251-15	ZFYVE20	Q9H1K0	MSVVPWVNDQDVPFC	VSSVLDEKDDDRIRC	103
HR8053B-150-273-15	ZFYVE20	Q9H1K0	MSVVPWVNDQDVPFC	LLKREQQIDEKEHTP	125
HR7907A-60-153-TEV	ZHX1	Q9UKY1	NQQNKKVEGGYECKY	NNQTIFEQTINDLTF	94
HR7907B-565-641-TEV	ZHX1	Q9UKY1	PDFTPQKFKEKTAEQ	EEKMEIDESNAGSSK	77
HR7907C-295-358-Av6HT	ZHX1	Q9UKY1	NNPLLLNTYNKFPYP	TPEEVEEARRKQFNG	64
HR7907D-768-820-Av6HT	ZHX1	Q9UKY1	NWDRGPSLIKFKTGT	KSHMGYEQVREWFAE	53
HR7907D-768-830-Av6HT	ZHX1	Q9UKY1	NWDRGPSLIKFKTGT	EWFAERQRRSELGIE	63
HR7907E-658-720-Av6HT	ZHX1	Q9UKY1	SGSTGKICKKTPEQL	SWFGDTRYAWKNGNL	63
HR7907E-658-728-Av6HT	ZHX1	Q9UKY1	SGSTGKICKKTPEQL	AWKNGNLKWYYYYQS	71
HR7907F-462-532-Av6HT	ZHX1	Q9UKY1	PDSFGIRAKKTKEQL	NQRNSKSNQCLHLNN	71
HR7907F-468-513-Av6HT	ZHX1	Q9UKY1	RAKKTKEQLAELKVS	KITGLTKGEIKKWFS	46

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7907F-468-532-Av6HT	ZHX1	Q9UKY1	RAKKTKEQLAELKVS	NQRNSKSNQCLHLNN	65
HR8292A-524-605-NHT	ZHX2	Q9Y6X8	AYPDFAPQKFKEKTQ	VLDSMGSGKKGQDVG	82
HR8292A-524-605-TEV	ZHX2	Q9Y6X8	AYPDFAPQKFKEKTQ	VLDSMGSGKKGQDVG	82
HR7743A-613-675-TEV	ZHX3	Q9H4I2	PTKYKERAPEQLRAL	SERRKKVNAEETKKA	63
HR7728A-219-303-TEV	ZIC1	Q15915	QPIKQELICKWIEPE	LVNHIRVHTGEKPFP	85
HR7748A-250-334-TEV	ZIC2	O95409	QCIKQELICKWIDPE	LVNHIRVHTGEKPFP	85
HR8404A-122-262-NHT	ZIC4	Q8N9L1	QPIKQELICKWLAAD	SSDRKKHSHVHTSDK	141
HR8404A-122-262-TEV	ZIC4	Q8N9L1	QPIKQELICKWLAAD	SSDRKKHSHVHTSDK	141
HR7356A-417-487-NHT	ZIK1	Q3SY52	SQSSILIQHRRIHTG	SQCSSLIHHQKCHNT	71
HR7796A-474-527-NHT	ZIM2	Q9NZV7	VFSRNSYLIQHYRTH	YQLHSQEKTVECDHC	54
HR8306A-277-331-Av6HT	ZIM3	Q96PE6	KSYQCNECEKSFRQN	IYKSDLVKHQRIHTG	55
HR8102A-61-140-Av6HT	ZKSCAN1	P17029	RFCYQNTFGPREALS	EAVTLLEDLELDLSG	80
HR8296A-7-131-Av6HT	ZKSCAN2	Q63HK3	SQIDAPLEVEGCLIM	VALVVHLEKETGRLR	125
HR7446A-37-132-NHT	ZKSCAN3	Q9BRR0	SPDLGSEGSRERFRG	VVLLEYLERQLDEPA	96
HR7362A-49-138-NHT	ZKSCAN4	Q969J2	PERSRQRFRGFRYPE	VVVLLEYLERQLDEP	90
HR7407A-25-130-NHT	ZKSCAN5	Q9Y2L8	LFIVKVEEEDCTWMQ	ESGEEAVAVIENIQR	106
HR7288A-62-140-NHT	ZMAT2	Q96NC0	LGKTIVITKTTPQSE	FEVNKKKMEEKQKDY	79
HR7490-503-839-Av6HT	ZMIZ1	Q9ULJ6	PVANYPHSPVPGNPT	VPIKSDLHIKDDPDG	337
HR7144A-212-289-NHT	ZNF10	P21506	SNECGQTFCQNIHLI	SWRSNLTRHQLIHTG	78
HR8024A-409-481-Av6HT	ZNF100	Q8IYN0	GFNWSSALTKHKRIH	NRSSQLTAHKMIHTG	73
HR8250A-364-426-Av6HT	ZNF101	Q8IZC7	KPYECTRCGKAFGWC	HERTHLAGRSQCFGR	63
HR7096A-555-620-15	ZNF107	Q9UII5	MEEHGKVFNQSSNLT	KPHKCEECGKAYNRF	67

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7096A-560-615-15	ZNF107	Q9UII5	MVFNQSSNLTTQKII	IYTGEKPHKCEECGK	57
HR7096B-607-719-Av6HT	ZNF107	Q9UII5	PHKCEECGKAYNRFS	SNLTTHKKIHTSEKP	113
HR7096C-611-676-15	ZNF107	Q9UII5	MEECGKAYNRFSNLT	KPYKCKECGKAFNLS	67
HR7096C-616-671-15	ZNF107	Q9UII5	MAYNRFSNLTIHKRI	IHTGEKPYKCKECGK	57
HR7096D-53-131-15	ZNF107	Q9UII5	MECTGHKGGHNTVNQ	SQLTQHRRIHTRVNS	80
HR7096D-58-126-15	ZNF107	Q9UII5	MKGGHNTVNQCLTAT	SFCVLSQLTQHRRIH	70
HR7096D-69-131-15	ZNF107	Q9UII5	MTATPSKIFQCNKYV	SQLTQHRRIHTRVNS	64
HR7096D-71-126-15	ZNF107	Q9UII5	MTPSKIFQCNKYVKV	SFCVLSQLTQHRRIH	57
HR7096E-158-269-TEV	ZNF107	Q9UII5	KPYKCEECGKAFNQS	LFSNLTNHKRIHAGE	112
HR7096F-672-727-Av6HT	ZNF107	Q9UII5	AFNLSSTLTAHKKIH	IHTSEKPYKCEECGK	56
HR7096G-659-772-Av6HT	ZNF107	Q9UII5	TGEKPYKCKECGKAF	NLSSNLTTHKKIHTG	114
HR7096H-702-776-Av6HT	ZNF107	Q9UII5	NQSSNLTTHKKIHTS	NLTTHKKIHTGEKPY	75
HR7096H-716-776-Av6HT	ZNF107	Q9UII5	SEKPYKCEECGKSFN	NLTTHKKIHTGEKPY	61
HR8087A-338-417-Av6HT	ZNF114	Q8NC26	GKAFRYSLHLNKHLR	LKKHLKTHKDEKPCE	80
HR7502A-318-390-NHT	ZNF121	P58317	GKAFATSSQLIEHIR	AYNRFYLLTKHLKTH	73
HR7834A-112-166-NHT	ZNF132	P52740	KANSCDMCGPFLKDI	WLNANLHQHQKEHSG	55
HR7007A-48-120-NHT	ZNF134	P52741	TALPCDICGPILKDI	LRRDKSEASIVKNCT	73
HR8079A-458-526-Av6HT	ZNF136	P52737	SYLNSFRTHEMIHTG	AYSCRASFQRHMLTH	69
HR8523A-1-79-Av6HT	ZNF137P	P52743	NVARFLVEKHTLHVI	IHGVGKLCKCNDCHK	78
HR6957A-157-213-NHT	ZNF140	P52738	VERPYGCHECGKTFG	SQISNLVKHQMIHTG	57
HR7470A-405-474-NHT	ZNF141	Q15928	RRSTDRSQHKKIHSA	FKRFSHLNKHKKIHT	70
HR8111A-337-394-Av6HT	ZNF143	P52747	SFTTSNIRKVHVRTH	IHTGEKPYVCTVPGC	58

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8395A-1-68-Av6HT	ZNF146	Q15072	SHLSQQRIYSGENPF	SQKQYVIKHQNTHTG	67
HR3636C-225-287-NHT	ZNF148	Q9UQR1	KPFRCDECGMRFIQK	HKRMCHENHDKKLNR	63
HR7492A-160-244-NHT	ZNF154	Q13106	CYICSECGKSFSKSY	SNLIKHRRVHTGERP	85
HR8099A-408-497-Av6HT	ZNF155	Q12901	GFYTNSQLSSHQRSH	PFKCEDCGKRLVHRT	90
HR7481A-394-466-NHT	ZNF157	P51786	AFYVKARLIEHQRMH	YVKVRLIEHQRIHTG	73
HR7426A-245-317-NHT	ZNF16	P17020	TFSQNSVLKNRHRSH	SQNSSLKKHQKSHMS	73
HR7677A-461-550-Av6HT	ZNF160	Q9HCG1	AFSMHSNLATHQVIH	PYKCIECGKSFTQKS	90
HR8279A-12-131-Av6HT	ZNF165	P49910	NSPEDEGLLIVKIEE	GEEAVTILEDLERGT	120
HR7169A-37-134-NHT	ZNF167	Q9P0L1	GQGSSLQKNYPPVCE	ESGEEAVAVVEDFQR	98
HR7079A-232-286-NHT	ZNF169	Q14929	KHHVCPECGRGFCQR	SQKASLSIHQRKHSG	55
HR8144A-504-578-Av6HT	ZNF17	P17021	GKSFRCRSTLDTHQR	SQNSHLIRHQKVHTR	75
HR7831A-37-132-TEV	ZNF174	Q18697	KNCPDPELCRQSFRR	VTLVEDFHRASKKPK	96
HR7382A-565-650-NHT	ZNF175	Q9Y473	GKAFTSKSQFKEHQR	THMGEKPYECLDCGK	86
HR8586A-246-321-Av6HT	ZNF177	Q13360	STGSYLIVHKRTHTG	LIMHKRIHNGQKLHE	76
HR8047A-6-143-Av6HT	ZNF18	P17022	GQALGLLPSLAKAED	WISIQVLGQDILSEK	138
HR7449A-517-602-Av6HT	ZNF180	Q9UJW8	GEKPFECNQCGKSFS	QSYVLVVHQRTHTGE	86
HR7508A-221-306-NHT	ZNF181	Q2M3W8	QGKSLTLPQTCNREK	PYKCIECGKAFSHVS	86
HR7707A-292-396-NHT	ZNF189	O75820	KCKKSFSRNSLLVEH	SQLCNLTRHQRIHTG	105
HR8281A-147-213-Av6HT	ZNF19	P17023	IQGKVPRIPCARKPF	NGNSSLIRHQRIHTG	67
HR8500A-45-132-Av6HT	ZNF192	O15776	LGQEVFRLRFRQLRY	NGEEVVTLLEDLERQ	88
HR7141A-10-132-NHT	ZNF193	O15535	SLGVQVPEAWEELLT	ESGEEAVILLEDLER	123
HR7496A-33-148-NHT	ZNF197	O14709	SSSSVWETSHLHFRQ	LVKDQDTLQKVVSAP	116

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7949A-310-387-Av6HT	ZNF20	P17024	PYECKQCGKAFRCGS	GKGFRCASQLQIHER	78
HR7725A-16-132-NHT	ZNF202	O95125	EGILMVKLEDDFTCR	VTLVEGLQKQPRRPR	117
HR7127A-305-371-NHT	ZNF205	O95201	RKSYRCEQCGKGFSW	IHTGEKPYTCPACRK	67
HR8501A-1095-1149-Av6HT	ZNF208	Q43345	KPYKCEECGKAFSTF	SWLSVFSKHKKIHTG	55
HR7117A-425-476-NHT	ZNF212	Q9UDV6	SSLICGYCGKSFSHP	KSFVQKQHLLQHQKI	52
HR7052A-50-129-NHT	ZNF213	O14771	QFCYGDVHGPHEAFS	EAVALVEDLQKQPVK	80
HR6908A-201-269-NHT	ZNF214	Q9UL59	VGVICQEDLLRDSME	CFSQRSDLYRHPRNH	69
HR8316A-49-130-Av6HT	ZNF215	O9UL58	QKFRHFQYLKVSGPH	SKDMVTLIEDVIEML	82
HR7041A-127-180-NHT	ZNF217	O75362	EFSCEVCGQTFRVAF	KEPWFLKNHMRTHNG	54
HR7381A-39-109-NHT	ZNF219	Q9P2Y4	SLGMGAVSWSESRAG	AQRALLRSHLRTHQP	71
HR7603A-137-200-Av6HT	ZNF22	P17026	MKPYQCDECGRCFSQ	MKVHKEEKPRKTRGK	65
HR7608A-137-200-NHT	ZNF22	P17026	KPYQCDECGRCFSQS	MKVHKEEKPRKTRGK	64
HR7848A-283-341-Av6HT	ZNF222	Q9UK12	KLYKSEKYGRGFIDR	YLLVHQRVHTGEKPY	59
HR8507A-136-200-Av6HT	ZNF223	Q9UK11	EGLSIMHTGQKPSNC	CYISALHIHQRVHLG	65
HR7500A-576-665-Av6HT	ZNF225	Q9UK10	SFSRASSILNHKRLH	LLQCEDCGKSIVHSS	90
HR7826A-501-553-NHT	ZNF226	O9NYT6	KPYKCNECGKSFRRN	GFSQSSYLQIHQKAH	53
HR7039-1-527-TEV	ZNF227	Q86WZ6	PSQNYDLPQKKQEKM	VHTGEKRFKCETCGK	526
HR7039-255-799-Av6HT	ZNF227	Q86WZ6	CGRGFSYSPRLPLHP	SRLTYHQKVHTGKKL	545
HR7039-320-799-Av6HT	ZNF227	Q86WZ6	GEKSYRCDSCGKGFS	SRLTYHQKVHTGKKL	480
HR7039A-21-80-Av6HT	ZNF227	Q86WZ6	EAVTFKDVAVVFSRE	PFQPDMVSQLEAEEK	60
HR7249A-552-607-NHT	ZNF229	Q9UJW7	SFGRSSDLHIHQRVH	VHTGERPYVCDVCGK	56
HR8056A-178-248-Av6HT	ZNF23	P17027	RCDSQLIQHQENNTE	SYSSHYITHQTIHSG	71

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7277A-201-287-Av6HT	ZNF230	Q9UIE0	RGKEFSQSSCLQTRE	IHTGEKPFKCEICGK	87
HR7779A-56-136-Av6HT	ZNF232	Q9UNY5	EEEQSCEYETRLPGN	LVLEQFLTILPEELQ	81
HR7083A-324-410-NHT	ZNF233	A6NK53	SQGSHLQPHQRVSTG	RACKCDVYDKGFSQT	87
HR7425A-606-678-Av6HT	ZNF234	Q14588	SQASSLQLHQSVHTG	RSNLVSHHKIHAAGT	73
HR6869A-681-738-NHT	ZNF235	Q14590	KPYTCQQCGKGFSQA	SHLIYHQRVHTGGNL	58
HR7932A-71-144-Av6HT	ZNF236	Q9UL36	CPQTFNVEFNLTLHK	FTLQSQLAVHMEEHR	74
HR7756A-1-111-NHT	ZNF238	Q99592	EFPDHSRHLLQCLSE	VLAAASYLHMYDIVK	110
HR7813A-381-458-NHT	ZNF239	Q16600	GKGFSQSSDLRIHLR	SNLHIHQRVHKKDPR	78
HR7147A-272-331-TEV	ZNF24	P17028	IHSGEKPYGCVECGK	SQNSGLINHQRIHTG	60
HR7147B-49-112-Av6HT	ZNF24	P17028	EIFRQRFRQFGYQDS	EQFVAILPKELQTWV	64
HR7147B-49-117-Av6HT	ZNF24	P17028	EIFRQRFRQFGYQDS	ILPKELQTWVRDHHP	69
HR7147B-49-138-Av6HT	ZNF24	P17028	EIFRQRFRQFGYQDS	VTVLEDLESELDDPG	90
HR7111A-518-570-Av6HT	ZNF248	Q8NDW4	KPYKCNECGKTFCEK	TFSQRSVLTKHQRIH	53
HR7751A-154-226-NHT	ZNF25	P17030	SFSKNEDLIRHQKIH	YQKPHLTEHQKTHTG	73
HR6988A-235-324-NHT	ZNF250	P15622	AFSQSSVLSKHRRIH	PYVCPLCGKAFNHST	90
HR7737A-208-296-Av6HT	ZNF253	O75346	AFNQSANLTTHKRIH	KPYKCEECGKAFKHP	89
HR6990A-584-652-NHT	ZNF254	O75437	NRSSTFTKHKVIHTG	AFNRSSHLTTDKITH	69
HR8384A-202-266-Av6HT	ZNF256	Q9Y2P7	VAFHSVKNHYNWGEC	CSLSDHLRVHTSEKP	65
HR7634A-464-533-Av6HT	ZNF26	P17031	PRKASLQIHQKTHSG	FCWNSGLRIHRKTHK	70
HR8255A-134-188-Av6HT	ZNF260	Q3ZCT1	KPYACKECGKAFNGK	SQKQYLIKHQNIHTG	55
HR7211A-35-111-NHT	ZNF263	O14978	PSPEASHLRFRRFRF	IQSRVQELHPESGEE	77
HR6888A-230-294-Av6HT	ZNF264	O43296	PYECTECGKTFIKST	IHSGEKPYKCNECGK	65

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8327A-276-365-Av6HT	ZNF266	Q14584	AFTVSSCLSQHMKIH	PYKCKDCGKAFTQNS	90
HR6896A-62-137-NHT	ZNF268	Q14587	LEWLFISQEQPKITK	QHTKPDIIFKLEQGE	76
HR8262A-158-241-Av6HT	ZNF273	Q14593	VHKRGYNGLNQCLTT	TATRVNFYKCKTCGK	84
HR7968A-77-161-Av6HT	ZNF276	Q8N554	GHCRLCHGKFSSRSL	HSLLKSFLQRVNASP	85
HR8173A-209-282-Av6HT	ZNF277	Q9NRM2	NCNEFLCTLQKKLDN	ELGKSWEEVQLEDDR	74
HR8391A-440-492-Av6HT	ZNF280C	Q8ND82	KNLLCPFCLKVSKMA	QFLTSKEKAEHKAQH	53
HR7724A-288-373-NHT	ZNF281	Q9Y2X9	PFQCSQCSMGFIQKY	RLLKHRRTCGEVIVK	86
HR7574A-86-180-Av6HT	ZNF282	Q9UDV7	REPQLPTAEISLWTV	RRLENLENLLRNRNF	94
HR6982A-569-623-NHT	ZNF283	Q8N7M2	KPFKCKECGKAFSWG	GSGYQLSVHQRFHTG	55
HR8101A-140-217-Av6HT	ZNF284	Q2VY69	IHIGETPSEHGKCKK	YKCDVCSKAFSQNSQ	78
HR7778A-510-590-NHT	ZNF285	Q96NJ3	KPYKCDECGKGFSRN	DLLTHQRLHEQRETL	81
HR7046A-197-250-NHT	ZNF286A	Q9HBT8	SFNQKSVLITEDRVP	TYKEKKPHKCNDCGE	54
HR7764A-1340-1400-TEV	ZNF292	O60281	PEKVKKDRGRGPNGK	NPRSLGGHLSKRSYC	61
HR7764B-550-593-Av6HT	ZNF292	O60281	EFLGHRIVRHAQKHY	NSKETFVPHVTLHVK	44
HR7764C-779-824-Av6HT	ZNF292	O60281	AKCMFPKCGRIFSEA	KFTGCGKVYRSQGEL	46
HR7764C-779-829-Av6HT	ZNF292	O60281	AKCMFPKCGRIFSEA	GKVYRSQGELEKHLD	51
HR7401A-713-806-TEV	ZNF295	Q9ULJ3	ASPVENKEVYQCRLC	RHQVEVHNQNNMAPT	94
HR7401B-907-958-Av6HT	ZNF295	Q9ULJ3	SLWPCEKCGKMFTVH	KAFRTNFRLWSHFQS	52
HR7401B-907-963-Av6HT	ZNF295	Q9ULJ3	SLWPCEKCGKMFTVH	NFRLWSHFQSHMSQA	57
HR7401C-1-125-Av6HT	ZNF295	Q9ULJ3	EGLLHYINPAHAISL	ISFLTNIVSKTPQAP	124
HR7401C-1-133-Av6HT	ZNF295	Q9ULJ3	EGLLHYINPAHAISL	SKTPQAPFPTCPNRK	132
HR7401D-1-110-Av6HT	ZNF295	Q9ULJ3	EGLLHYINPAHAISL	KSSLAAVQELGYSLG	109

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7401D-1-114-Av6HT	ZNF295	Q9ULJ3	EGLLHYINPAHAISL	AAVQELGYSLGISFL	113
HR7438A-396-451-NHT	ZNF296	Q8WUU4	TNSSNLTVHRRSHTG	GMTPGSTRFECPHCH	56
HR7980A-567-639-Av6HT	ZNF304	Q9HCX3	AYISSSHLVQHKKVH	SRSSHLVRHQKAHTG	73
HR6886A-245-327-NHT	ZNF311	Q5JNZ3	KLHECARCGKNFSWH	NSRSALCRHKKTHSG	83
HR8348A-456-510-Av6HT	ZNF319	Q9P2F9	KPLRCTLCERRFFSS	KYASDLQRHRRVHTG	55
HR7649A-335-419-Av6HT	ZNF320	A2RRD8	DKVFSRKSHLERHRR	KLHTGEKLYECEECD	85
HR7116A-290-344-NHT	ZNF322A	Q6U7Q0	THTFKCLEYEKSFNC	FLLGMDFVAQQKMRT	55
HR7719-1-389-Av6HT	ZNF322B	Q5SYY0	YTSEEKCNQRTQKRK	GEKPFVCNVSEKGLE	388
HR7473A-7-125-NHT	ZNF323	Q96LW9	QYDLKIVKVEEDPIW	VAVVEDLEQELSEPG	119
HR7209A-240-309-NHT	ZNF324	O75467	PSTWDELGEALHAGE	SQTSHLTQHQRIHSG	70
HR7920A-181-255-Av6HT	ZNF329	Q86UD4	ENIFTLSSSLNENQR	SKNYNLIVHQRIHTG	75
HR8085A-409-463-Av6HT	ZNF331	Q9NQX6	KPYGCTECGKSFSHG	NHLNHLREHQRIHNS	55
HR8071A-542-633-Av6HT	ZNF333	Q96JL9	VLSRLSTLKSHMRTH	QCNQCEKAFRHSSSL	92
HR7926A-511-568-Av6HT	ZNF334	Q9HCZ1	NTKENLYECSEHGHA	CRKSALTHHQRTHTG	58
HR8140A-560-612-Av6HT	ZNF335	Q9H4Z2	SSFPCPVCGRVYPMQ	SFKKRYTFKMHLLTH	53
HR7962A-683-744-Av6HT	ZNF337	Q9Y3M9	KPFVCQECKRGYTSK	KHLKRHLREKRFCTG	62
HR7329A-327-374-NHT	ZNF33B	Q06732	KHFECNECGKAFWEK	NQCGKTFWEKSNLTK	48
HR7973A-53-101-Av6HT	ZNF343	Q6P1L6	EGKAQIVVPVTFRDV	YKEVMLENYRNLLSL	49
HR7782A-408-478-NHT	ZNF345	Q14585	SSGSALNRHQRIHTG	GRDSEFQQHKKSHNG	71
HR8375A-52-156-Av6HT	ZNF346	Q9UL40	QPVGREEVEHMIQKN	TFSSPVVAQSHYLGK	105
HR7359A-200-274-NHT	ZNF35	P13682	GGKYSLNSGAVKNPK	IQSANLVVHQRIHTG	75
HR7521A-532-605-NHT	ZNF354A	O60765	GQSSALIQHRRIHTG	SSLTNHYKIHIEEDP	74

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8204A-166-238-Av6HT	ZNF354B	Q96LW1	NFYLKSVFIKQQRFA	IHNSSLRKHQKNHTG	73
HR7986A-494-548-Av6HT	ZNF354C	Q86Y25	KLYKCMECGKAYSYR	ICSSSLTQYQRFFKG	55
HR8492A-212-270-Av6HT	ZNF355P	Q9NSJ1	CKCEECGKACKQSLG	IHAGEKPYNCEKCGK	59
HR7559A-189-242-NHT	ZNF358	Q9NW07	SHGATLAQHRGIHTG	SHSGEKPHHCPVCGK	54
HR8534A-152-309-Av6HT	ZNF365	Q70YC5	DTKASFEAHVREKFN	QQASGFVRDLSGHVL	158
HR7222A-233-319-NHT	ZNF366	Q8N895	DVNVQIDDSYYVDVG	GTRPHKCQVCHKAFT	87
HR7913A-160-248-Av6HT	ZNF367	Q7RTV3	GEHSSSRIRCNICNR	SRFTHANRHCPKHPY	89
HR8143A-116-153-Av6HT	ZNF37A	P17032	EPSEYNKNGNSFWLN	IKNWEQSFEYNECGK	38
HR7024A-370-456-NHT	ZNF383	Q8NA42	ECGKAFTQSSQLRQH	RIHTGEKPYNCKECG	87
HR8244A-91-161-Av6HT	ZNF391	Q9UJN7	KDNSDLIKHQRLFSQ	SRSTHLIEHQRTHTG	71
HR7003A-55-179-NHT	ZNF394	Q53GI3	AASPDPETSRLHFRQ	TWEEWERLDPARRDF	125
HR7436A-17-136-NHT	ZNF397	Q8NF99	PEQELILVKVEDNFS	VTLLEDLEREFDDPG	120
HR6874A-291-375-Av6HT	ZNF41	P51814	RIHAGEKSRECDKSN	GKAFFQRSDLFRHLR	85
HR7062-124-478-15	ZNF410	Q86VK4	MLNLTRAGLGSSAEH	PQELLNQGDLTERRT	356
HR7062-129-478-15	ZNF410	Q86VK4	MAGLGSSAEHLVFVQ	PQELLNQGDLTERRT	351
HR7062-150-478-15	ZNF410	Q86VK4	MNDFLSSESTDSSIP	PQELLNQGDLTERRT	330
HR7062-155-478-15	ZNF410	Q86VK4	MSESTDSSIPWFLRV	PQELLNQGDLTERRT	325
HR7338A-520-571-NHT	ZNF416	Q9BWM5	RPYDCGQCGKSFIQK	KSFTQHSGLILHRKS	52
HR6922A-214-292-NHT	ZNF417	Q8TAU3	CGKRTKAFSTKHSVI	SRKSSLIQHQRVHTG	79
HR7936A-544-625-Av6HT	ZNF418	Q8TF45	GKSFHQSSSLLRHQK	RLHTRGKPYECSECG	82
HR8286A-145-234-Av6HT	ZNF420	Q8TAQ5	GKAFRRASHLTQHQS	EKPYKCEECGKAFIR	90
HR7298A-293-344-NHT	ZNF423	Q2M1K9	ADLQCIHCPEVFVDE	EQFSSVEGVYCHLDS	52

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7298B-1204-1284-Av6HT	ZNF423	Q2M1K9	NQMFDSPAKLLCHLI	FQTELQNHTMSQHAQ	81
HR7298C-136-178-Av6HT	ZNF423	Q2M1K9	LPYPCQFCDKSFIRL	LPFKCTYCSRLFKHK	43
HR7298D-627-684-Av6HT	ZNF423	Q2M1K9	ISNGEYPCNQCDLKF	DFDSQESLLQHLTVH	58
HR7298E-750-803-Av6HT	ZNF423	Q2M1K9	YRCTACNWDFRKEAD	TFSTEVELQCHITTH	54
HR7298F-923-981-Av6HT	ZNF423	Q2M1K9	AEFIKGSHKCNVCSR	RFPSLLTLTEHKVTH	59
HR7298F-928-981-Av6HT	ZNF423	Q2M1K9	GSHKCNVCSRTFFSE	RFPSLLTLTEHKVTH	54
HR8124A-692-742-Av6HT	ZNF425	Q6IV72	RPFQCPECGKGFLQK	GRSFTYVGALKTHIA	51
HR7371A-502-554-NHT	ZNF426	Q9BUY5	KPYECKECGKAFTCS	AYSHPRSLRRHEQIH	53
HR7017A-571-643-NHT	ZNF429	Q86V71	DKAFTHSSNLSSHKK	AFTRSSRLTQHKKIH	73
HR8161A-229-300-Av6HT	ZNF43	P17038	PYTCEECGKVFNWSS	YKCKECAKAFNQSSN	72
HR8378A-511-570-Av6HT	ZNF430	Q9H8G1	TSYKYLECDKAFSQS	LIEQSNSYWRETLQM	60
HR6876A-159-226-NHT	ZNF431	Q8TF32	EGYNELNQCLTTTQS	SFCMLLHLSQHKRIH	68
HR7979A-260-324-Av6HT	ZNF432	O94892	SFICSECGKVFTMKS	NHTGEKSYICSECGK	65
HR7340A-616-673-NHT	ZNF433	Q8N7K0	KPYKCKQCGKAFGCP	SQLQVHGRAHCIDTP	58
HR7145A-271-333-NHT	ZNF434	Q9NX65	SHQSFCARDKACTHI	SRSSYLVRHQRIHTG	63
HR6863A-401-470-NHT	ZNF436	Q9C0F3	ERSDLIKHQRTHTGE	SRSSALIKHKRVHTD	70
HR6993A-498-598-NHT	ZNF438	Q7Z4V0	GFSGIKKPWHRCHVC	GHLKEVHRVVISTEP	101
HR7001A-19-66-NHT	ZNF439	Q8NDP4	VAFKDVAVNFTQEEW	FWNLTSIGKKWKDQN	48
HR7627A-522-574-NHT	ZNF44	P15621	EPYECKECGKAFSSF	AFSRFSYLKTHERTH	53
HR7091A-1-51-NHT	ZNF440	Q8IYI8	DPVAFKDVAVNFTQE	FRNLTSLGKRWKDQN	50
HR6977A-625-693-NHT	ZNF441	Q8N8Z8	SHSSYLRIHERVHTG	AFHCISSFHKHEMTH	69
HR7410A-570-627-NHT	ZNF442	Q9H7R0	KSYECQQCGKAFTRS	SSLHRHKRTHWRDTL	58

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7294A-14-105-NHT	ZNF444	Q8N0Y2	LALDSPWHRFRRFHL	AVALLEELWGPAASP	92
HR8025A-61-139-Av6HT	ZNF445	P59923	LRYHESSGPLETLSR	EAVALLEELQRDLDG	79
HR8393A-22-126-Av6HT	ZNF446	Q9NWS9	PETARLRFRGFCYQE	LGWITAHVLKQEVLP	105
HR7503A-25-115-NHT	ZNF449	Q6P9G9	DCEVFRQRFRQFQYR	VVSLIEDLQRELEIP	91
HR7588A-340-406-Av6HT	ZNF45	Q02386	SFSYSSHLNIHCRIH	ECGKGFCRASNLLDH	67
HR7276A-252-326-Av6HT	ZNF454	Q8N9F8	AFSVSSSLTYHQKIH	RAHLTKHQNIHSGEK	75
HR7023A-306-367-NHT	ZNF460	Q14592	KPFACSECGKGFYES	QHERIHTGEKPFVCS	62
HR7305A-244-297-NHT	ZNF461	Q8TAF7	KCNECKECWKAFVHC	NYGSELTLHQRIHTG	54
HR8320A-1871-1948-TEV	ZNF462	Q96JM2	SRDLKRDFIILGNGP	KQKYADGAFADFKQE	78
HR8302A-424-504-15	ZNF467	Q7Z7K2	MAPSGERSFFCPDCG	AQCGRRFSRKSHLGR	82
HR8302A-429-499-15	ZNF467	Q7Z7K2	MRSFFCPDCGRGFSH	RPFACAQCGRRFSRK	72
HR8302B-485-539-15	ZNF467	Q7Z7K2	MRPFACAQCGRRFSR	SSKTNLVRHQAIHTG	56
HR8302C-540-595-TEV	ZNF467	Q7Z7K2	SRPFSCPQCGKSFSR	AWSAPPEVAPPPLFF	56
HR8302C-551-595-TEV	ZNF467	Q7Z7K2	SFSRKTHLVRHQLIH	AWSAPPEVAPPPLFF	45
HR8121A-1-49-Av6HT	ZNF468	Q5VIY5	ALPQGLLTFRDVAIE	DVMLENYRNLVSLDI	48
HR8083A-181-258-Av6HT	ZNF471	Q9BX82	TSDKKSFSKNSMVIK	KQRQHLAQHHRTHTG	78
HR7760A-205-288-Av6HT	ZNF473	Q8WTR7	GEKPYQCSECGKSFS	FSQSTYLWHQKTHTG	84
HR8431-87-306-Av6HT	ZNF474	Q6S9Z5	IPARRPGFRVCYICG	RIFTSDRLLVHQRSC	220
HR6879A-445-517-NHT	ZNF479	Q96JC4	AFSLSSTLTDHKRIH	KWHSSLAKHKIIHTG	73
HR8210A-201-255-Av6HT	ZNF480	Q8WV37	KPYECNEHSKVFRVS	SRNSHLAEHCRIHTG	55
HR7266A-379-438-TEV	ZNF483	Q8TF39	KRQKIHLGDRSQKCS	AALNKDEGNESGEKT	60
HR7735A-326-380-NHT	ZNF484	Q5JVG2	NYYKCSDYGRAFIQK	PQNSNLNIHKKIHTG	55

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7735B-1-66-Av6HT	ZNF484	Q5JVG2	TKSLESVSFKDVTVD	PKPEVIFSLEQEEPC	65
HR7735B-6-66-Av6HT	ZNF484	Q5JVG2	ESVSFKDVTVDFSRD	PKPEVIFSLEQEEPC	61
HR7735C-259-310-Av6HT	ZNF484	Q5JVG2	VFSPKSHAFAHESIC	GSQRVYAGICTEYEK	52
HR7735C-259-316-Av6HT	ZNF484	Q5JVG2	VFSPKSHAFAHESIC	AGICTEYEKDFSLKS	58
HR8114A-115-182-Av6HT	ZNF485	Q8NCK3	EKGLDWEGRSSTEKN	MNSSSLLNHHKVHAG	68
HR8541A-108-234-Av6HT	ZNF486	Q96H40	ILRKFEKCGHGNLHF	NRSSHLTTHKITHTR	127
HR7094A-456-529-NHT	ZNF490	Q9ULM2	IYFSHLRRHERSHTG	KSLHVHERTHSRQKP	74
HR8126A-332-407-Av6HT	ZNF491	Q8N8L2	CGKAFRSAKYIRIHG	TCSIYIRIHERIHTG	76
HR7429A-33-114-NHT	ZNF496	Q96IT1	GELPSPESSRRLFRR	SWVRAQEPESGEQAV	82
HR7643A-35-118-NHT	ZNF498	Q6NSZ9	DPSPETFRLRERQFR	EHGPESGKALAAMVE	84
HR7635A-55-136-NHT	ZNF500	O60304	LFCYQEVAGPREALS	VVLVEGLQRKPRKHR	82
HR7681A-163-225-NHT	ZNF501	Q96CX3	KCNECGKAFNQSAGL	THTGEKLYKCSECEK	63
HR8192A-151-240-Av6HT	ZNF502	Q8TBZ5	QKKSWKCNECGKTFT	LTQHQRIHTGEKPYK	90
HR7474A-5-630-TEV	ZNF503	Q96F45	PSLSALRSSKHSGGG	PVPVPAATGPYYSPY	626
HR7624A-141-197-NHT	ZNF506	Q5JVG8	QRKIFQCDEYVKFLH	NQSSTRTTYKKIDAG	57
HR7678A-639-723-NHT	ZNF507	Q8TCN5	RPYRCRLCHYTSGNK	KSQLRNHEREQHSLP	85
HR7670A-34-107-15	ZNF510	Q9Y2H8	MQEQQKMNISQASVS	EVIFKLEQGEEPWFS	75
HR7670A-40-102-15	ZNF510	Q9Y2H8	MNISQASVSFKDVTI	CCFKPEVIFKLEQGE	64
HR7670B-515-683-Av6HT	ZNF510	Q9Y2H8	SFQCNQCGKTFGQKS	TLSLYQKIQGEGNPY	169
HR7670B-521-652-Av6HT	ZNF510	Q9Y2H8	CGKTFGQKSNLRIHQ	GQKSNLRIHQRTHSG	132
HR7670B-521-683-Av6HT	ZNF510	Q9Y2H8	CGKTFGQKSNLRIHQ	TLSLYQKIQGEGNPY	163
HR7670C-582-683-Av6HT	ZNF510	Q9Y2H8	ARTSTLRVHQRIHTG	TLSLYQKIQGEGNPY	102

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7670C-595-683-Av6HT	ZNF510	Q9Y2H8	TGEKPFKCNECGKKF	TLSLYQKIQGEGNPY	89
HR7670D-552-607-Av6HT	ZNF510	Q9Y2H8	SFWRKDHLIQHQKTH	IHTGEKPFKCNECGK	56
HR8051A-493-578-TEV	ZNF512B	Q96KM6	PGGPEEQWQRAIHER	SAKPSDAEASEGGEQ	86
HR7686A-202-256-NHT	ZNF514	Q96K75	KSCKCNECGKSFHFQ	GHISSLIKHQRTHTG	55
HR7203A-240-299-NHT	ZNF516	Q92618	KPELSPGEFPCEVCG	FKEPWFLKNHMKAHG	60
HR8163A-367-458-Av6HT	ZNF517	Q6ZMY9	PHECPVCGRPFRHNS	RLHSGERPYRCRACG	92
HR6938A-228-328-NHT	ZNF518A	Q6AHZ1	RHNEIHYKCGKCHHV	ILKRYKIGASRKTFW	101
HR8175A-141-213-Av6HT	ZNF518B	Q9C0D4	RFSTKDPLQYKKHTL	AIRNDYIVKHTKRVH	73
HR8275A-281-327-Av6HT	ZNF519	Q8TB69	GHQKIHTGEKPYKCK	IHTGEKPFKCKECGK	47
HR8035A-114-170-Av6HT	ZNF521	Q96K83	PGLPYPCQFGDKSFS	KHKRSRDRHIKLHTG	57
HR8035B-928-981-Av6HT	ZNF521	Q96K83	GNYKCNVCSRTFFSE	RFPSLLTLTEHKVTH	54
HR8035C-1253-1292-Av6HT	ZNF521	Q96K83	GGTFKCPVCFTVFVQ	AHGQEDKIYDCTQCP	40
HR8035C-1253-1311-Av6HT	ZNF521	Q96K83	GGTFKCPVCFTVFVQ	FQTELQNHTMTQHSS	59
HR8035D-1177-1247-Av6HT	ZNF521	Q96K83	QVSPMPRISPSQSDE	TFDSPAKLQCHLIEH	71
HR8035D-1187-1247-Av6HT	ZNF521	Q96K83	SQSDEKKTYQCIKCQ	TFDSPAKLQCHLIEH	61
HR8035D-1193-1247-Av6HT	ZNF521	Q96K83	KTYQCIKCQMVFYNE	TFDSPAKLQCHLIEH	55
HR8035E-750-805-Av6HT	ZNF521	Q96K83	KVYRCTSCNWDFRNE	SFGTEVELQCHITTH	56
HR8035E-750-841-Av6HT	ZNF521	Q96K83	KVYRCTSCNWDFRNE	HLREKHCVFETKTPN	92
HR8035F-632-686-Av6HT	ZNF521	Q96K83	GEYICNQCGAKYTSL	EFPNQESLLKHVTIH	55
HR8035G-692-745-Av6HT	ZNF521	Q96K83	TYYICESCDKQFTSV	FDSKVSIQLHLAVKH	54
HR7651A-311-395-NHT	ZNF526	Q8TF50	QRSFSSANRLQAHGR	AHTANPLHRCRCGKT	85
HR8497A-368-478-Av6HT	ZNF527	Q8NB42	SRYAFLVEHQRIHTG	HTGEKPYECIKCGKF	111

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7761A-499-570-NHT	ZNF528	Q3MIS6	GKVFSRSSNLVCHQK	KAFRGCSGLTAHLAI	72
HR6966A-331-386-NHT	ZNF530	Q6P9A1	SFSHSTNLYRHRSAH	VHTGVRPYECSECGK	56
HR7961A-840-894-Av6HT	ZNF532	Q9HCE3	VGFRCVHCNVVYSDV	KSAPSTHSHAYTQHP	55
HR6910A-112-178-NHT	ZNF536	O15090	GIMSQMSDIEDDARK	DHRAAQKGNLKIHLR	67
HR7987A-333-390-Av6HT	ZNF540	Q8NDQ6	GKAFSVCGQLTRHQK	THAGKKPYECKECGK	58
HR8055A-1071-1130-Av6HT	ZNF541	Q9H0D2	EPHINIGSRFQAEIP	TQDRVTELCNVACSS	60
HR8506A-95-170-Av6HT	ZNF542	Q5EBM4	CTRNVCKECGNLYCH	CNECIKTFNQRAHLT	76
HR7303A-250-315-NHT	ZNF544	Q6NX49	SLNYGSSLCFHGRTF	DECRETCSESLCLVQ	66
HR7708A-449-521-NHT	ZNF546	Q86UE3	AFRLQTELTRHHRTH	SSRYHLTQHYRIHTG	73
HR8224A-347-390-Av6HT	ZNF547	Q8IVP9	TGERPYECSECGKAF	AAKQCSECGKFFRYN	44
HR7308A-303-359-NHT	ZNF552	Q9H707	KFFRHKYHLIAHQRV	VHTGQKPYECSECGK	57
HR7586A-351-442-Av6HT	ZNF554	Q86TJ5	PYECQECGRAFTHSS	RTHTGEKPYECSECG	92
HR6864A-534-597-NHT	ZNF555	Q8NEP9	KPYECKECGKVFKWP	VRIHTTEKQYKCNVG	64
HR7560A-285-337-NHT	ZNF556	Q9HAH1	RPYECKQCGKAYCWA	AFGWPSSLHKHARTH	53
HR7484-16-423-TEV	ZNF557	Q8N988	FPASQREGHTEGGEL	CGKSFTSNSYLSVHT	408
HR7484-32-423-TEV	ZNF557	Q8N988	NELLKSWLKGLVTFE	CGKSFTSNSYLSVHT	392
HR7484-TEV	ZNF557	Q8N988	AAVVLPPTAALSSLF	SYLSVHTRMHNRQM*	430
HR7484A-12-94-15	ZNF557	Q8N988	MLSSLFPASQREGHT	ASLGNQVDKPRLISQ	84
HR7484A-16-89-15	ZNF557	Q8N988	MFPASQREGHTEGGE	NCRNLASLGNQVDKP	75
HR7484A-32-89-15	ZNF557	Q8N988	MNELLKSWLKGLVTF	NCRNLASLGNQVDKP	59
HR7484B-344-408-15	ZNF557	Q8N988	MGEKPYTCNECGKSF	HMRTHTGKKPYECNY	66
HR7484B-344-423-15	ZNF557	Q8N988	MGEKPYTCNECGKSF	CGKSFTSNSYLSVHT	81

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7484B-349-403-15	ZNF557	Q8N988	MTCNECGKSFTNSFS	SSVKKHMRTHTGKKP	56
HR7484B-349-423-15	ZNF557	Q8N988	MTCNECGKSFTNSFS	CGKSFTSNSYLSVHT	76
HR8385A-150-204-Av6HT	ZNF558	Q96NG5	KLNECNQCFKVFSTK	SSRSYLTIHKRIHNG	55
HR7908A-195-264-Av6HT	ZNF559	Q9BR84	PSSSHLRECVRIYGG	FTESSYLTQHLRTHS	70
HR8030A-275-342-Av6HT	ZNF560	Q96MR9	RLILNVQVQRKCTQD	AFTHSTSHAVNVETH	68
HR8284A-153-246-Av6HT	ZNF561	Q8N587	KDTLSVHKEASTGQE	RAVTASSHLKQCVAV	94
HR7120A-315-406-NHT	ZNF562	Q6V9R5	PHKCTECGKAFTRST	RIHTGEKPYECVECG	92
HR7101A-177-267-NHT	ZNF563	Q8TA94	TFSSRRNLRRHMVVQ	YECKQCSKALPDSSS	91
HR7852A-258-333-NHT	ZNF564	Q8TBZ8	CGKAFDRPSLFRIHE	IFPSYVRKHERTHTG	76
HR7421A-165-219-Av6HT	ZNF565	Q8N9K5	KLMECHECGKAFSRG	SRASHLVQHQRIHTG	55
HR6953A-201-290-Av6HT	ZNF566	Q969W8	CKECGKSFRHPSRLT	IHTGEKPYECKECGK	90
HR7055-1-501-TEV	ZNF567	Q8N184	DVMLENYCHLISVGC	TNLNLHQRIHTGEKP	500
HR7055A-1-62-15	ZNF567	Q8N184	MDVMLENYCHLISVG	KAEDFLVKFKEHQEK	62
HR7055A-1-67-15	ZNF567	Q8N184	MDVMLENYCHLISVG	LVKFKEHQEKYSRSV	67
HR7513A-584-636-NHT	ZNF568	Q3ZCX4	KPYECNKCGKAFSQC	AFSQRASLSIHKRGH	53
HR8322A-184-236-Av6HT	ZNF569	Q5MCW4	TPFKCNHCGKGFNQT	AFSHKEKLIKHYKIH	53
HR7036A-222-276-NHT	ZNF57	Q68EA5	KTYKCEQCRMAFNGF	IYPSTFQRHMTTHTG	55
HR8437A-468-518-Av6HT	ZNF570	Q96NI8	KPYECTVCGKAFSYC	KKTFRQHAHLAHHQR	51
HR7898A-556-609-Av6HT	ZNF571	Q7Z3V5	KPYECKECGRAFSRG	FRCPSQLTQHTRLHN	54
HR7069A-130-212-NHT	ZNF572	Q7Z3I7	RPYKCSECWKSFSNS	SNTSHLIIHERTHTG	83
HR7339A-346-414-NHT	ZNF574	Q6ZN55	PSPSSLDQHLGDHSS	FVNLTKFLYHRRTHG	69
HR7766A-185-234-NHT	ZNF575	Q86XF7	AFSFPSKLAAHRLCH	QAFGQRRLLLLHQRS	50

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7135A-110-164-NHT	ZNF576	Q9H609	PTFPCPDCGKTFGQA	QDFAQEAGLHQHYIR	55
HR7392A-95-172-Av6HT	ZNF580	Q9UK33	PECARVFASPLRLQS	RFQDAAELAQHVRLH	78
HR7332A-85-197-NHT	ZNF581	Q9P0T4	KCYSCPVCSRVFEYM	MEQNTLQKHTRWKHP	113
HR7613A-143-226-NHT	ZNF582	Q96NG8	IIRHEEMPTFDQHAS	SRLIQHENIHSGKKP	84
HR8213A-490-546-Av6HT	ZNF583	Q96ND8	KPYECNVCGKAFSYS	RAHLAHHERIHTMES	57
HR7005A-111-199-NHT	ZNF584	Q8IVC4	EHLKSYRVIQHQDTH	RPFRCPTGRSAFKKS	89
HR7126A-700-769-NHT	ZNF585B	Q52M93	TKKSQLQVHQRIHTG	FVQKSVFSVHQSSHA	70
HR7959A-294-369-Av6HT	ZNF586	Q9NXT0	ECGKSFSLRSNLIHH	AENSSLIKHLRVHTG	76
HR8398-1-385-15	ZNF587	Q96SQ5	MAAAVPRRPTQQGTV	QRVHTGERPYKCGEC	385
HR8398A-13-68-15	ZNF587	Q96SQ5	MGTVTFEDVAVNFSQ	LGCWCGSKDEEAPCK	57
HR8398A-8-73-15	ZNF587	Q96SQ5	MRPTQQGTVTFEDVA	GSKDEEAPCKQRISV	67
HR8398B-85-147-15	ZNF587	Q96SQ5	MGVSPKKAHPCEMCG	AYLHQHQKQHIGEKF	64
HR8398B-90-144-15	ZNF587	Q96SQ5	MKAHPCEMCGLILED	DDTAYLHQHQKQHIG	56
HR7374A-223-283-NHT	ZNF589	Q86UQ0	AFNQKSNLFRQKAVT	THTGEKPYVCGECGR	61
HR7253A-8-134-TEV	ZNF593	O00488	GAHRAHSLARQMKAK	PTEVSTEVPEMDTST	127
HR7622A-658-732-NHT	ZNF594	Q96JF6	GKAFSQRSHLATHQK	MWHTAFLKHQRLHAG	75
HR8535A-211-338-Av6HT	ZNF595	Q8IYB9	RSTSLSKHKRIHTGE	SRSLNEHKNIHTGEK	128
HR7605A-165-247-NHT	ZNF596	Q8TC21	KSYGSHLFDYAFIQN	THCSDLRKHERTHTG	83
HR6958A-339-424-NHT	ZNF597	Q96LX8	KPLQCPDCDMTFPCF	LHLITHKRTHIKNTT	86
HR7504A-200-279-NHT	ZNF599	Q96NL3	TCTECGKGFSKKWAL	KRRFHLTEHQRIHTG	80
HR7536A-401-473-NHT	ZNF605	Q86T29	AFFKKSELIRHQKIH	TQKSSLISHQRTHTG	73
HR7780A-327-407-NHT	ZNF606	Q8WXB4	NQSPSFNEHPRLHVG	TYTAEKPYDYNECGT	81

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7972A-628-696-Av6HT	ZNF607	Q96SK3	CASYLVRHESVHADG	FRLRSILEVHQRIHI	69
HR7618A-201-256-NHT	ZNF610	Q8N9Z0	SYEYECSEDGEVFRV	SRNSHLVEHWRIHTG	56
HR7693A-602-688-NHT	ZNF611	Q8N823	TFSRRSSLHCHRRLH	AEKPYKCNECGKAFN	87
HR8205A-466-535-Av6HT	ZNF613	Q6PF04	SHKSGLINHQRIHTG	FSHLSCLVYHKGMLH	70
HR7312A-239-311-NHT	ZNF614	Q8N883	KLSRSVLFTKHLKTN	TMKRYLIAHQRTHSG	73
HR7638A-239-293-NHT	ZNF616	Q08AN1	KSYQCDVCGKIFRKN	SKSSHLAVHQRIHTG	55
HR8536A-215-271-Av6HT	ZNF619	Q8N2I2	PYTCKECGKTFRYNS	SHLLQHQKLHGGQRP	57
HR7004A-342-416-Av6HT	ZNF620	Q6ZNG0	GKRLSSNTALTQHQR	SWCGRFILHQKLHTQ	75
HR6865A-260-345-NHT	ZNF621	Q6ZSS3	EKLYKCKECWKAFGC	YGSFVQHQKLHPVEK	86
HR7076A-233-357-Av6HT	ZNF622	Q969S3	MQDAEEEEAEEGPPL	FADFYDFRSSYPDHK	126
HR7076A-233-357-NHT	ZNF622	Q969S3	QDAEEEEAEEGPPLG	FADFYDFRSSYPDHK	125
HR8004A-806-858-Av6HT	ZNF624	Q9P2J8	RPYKCEECGKAFRTN	AFRSSSSLTVHQRIH	53
HR8159A-235-289-Av6HT	ZNF625	Q96I27	KPYECKQCGKAFRSA	GCASSVKIHERTHTG	55
HR8312A-451-505-Av6HT	ZNF626	Q68DY1	KFYKCEECGKAFKCS	NQSSILTTHERIILE	55
HR7221A-166-234-Av6HT	ZNF627	Q7L945	PYDCKECGETFISLV	EKPYECKQCGKAFSC	69
HR7999A-830-869-Av6HT	ZNF629	Q9UEG4	GQNPKTLVEEKPYLC	AALLLHRSCHPGVSL	40
HR7098A-597-651-NHT	ZNF630	Q2M218	KTPECAESGMTFFWK	CQHVYFTGHQNPYRK	55
HR7646-132-485-Av6HT	ZNF639	Q9UID6	VHTAEDVPIAVEVHA	NERELISHLPVHETT	354
HR7646-158-485-Av6HT	ZNF639	Q9UID6	NSSESLQDQTDEEPP	NERELISHLPVHETT	328
HR7646-168-485-Av6HT	ZNF639	Q9UID6	DEEPPAKLCKILDKS	NERELISHLPVHETT	318
HR7646-24-485-Av6HT	ZNF639	Q9UID6	ISRIADGFNGIFSDH	NERELISHLPVHETT	462
HR7646-80-485-Av6HT	ZNF639	Q9UID6	RNQNYLVPSPVLRIL	NERELISHLPVHETT	406

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7646-Av6HT	ZNF639	Q9UID6	NEYPKKRKRKTLHPS	ERELISHLPVHETT*	485
HR7646A-406-471-15	ZNF639	Q9UID6	MDDCGKGFSSMLEYC	DLPHKCSDCLMRFGN	67
HR7646A-406-485-15	ZNF639	Q9UID6	MDDCGKGFSSMLEYC	NERELISHLPVHETT	81
HR7646A-411-466-15	ZNF639	Q9UID6	MGFSSMLEYCKHLNS	FKHSADLPHKCSDCL	57
HR7646A-411-485-15	ZNF639	Q9UID6	MGFSSMLEYCKHLNS	NERELISHLPVHETT	76
HR7646B-238-313-Av6HT	ZNF639	Q9UID6	NVCRVCKESFSTNML	SSSSELYLHFQEHSC	81
HR7646B-256-313-Av6HT	ZNF639	Q9UID6	EEDPYICKYCDYKTV	SSSSELYLHFQEHSC	58
HR7646C-372-425-Av6HT	ZNF639	Q9UID6	NFFVCQVCGFRSRLH	GFSSMLEYCKHLNSH	54
HR7646D-202-255-Av6HT	ZNF639	Q9UID6	GLYKCELCEFNSKYF	SFSTNMLLIEHAKLH	54
HR7858A-251-323-Av6HT	ZNF642	Q49AA0	RNTYKLDLINHPTSY	SQSASLSTHQRIHTG	73
HR7770A-52-130-NHT	ZNF645	Q8N7E2	LPIHFCDKCDLPIKI	IVQQCKRTYLSQKSL	79
HR7348A-260-345-NHT	ZNF648	Q5T619	QKPSKPLSPAETRGG	GEKPYPCPDCGKAFV	86
HR7533A-232-314-NHT	ZNF649	Q9BS31	KPHGCSLCGKAFYKR	SRKSLLVVHQRTHTG	83
HR7463A-365-451-NHT	ZNF652	Q9Y2D9	SFKRSMSLKVHSLQH	GEKPFICEICGKSFT	87
HR8324A-526-605-Av6HT	ZNF653	Q96CK0	REFTCETCGKSFKRK	CGKRFEKLDSVKFHT	80
HR8422A-173-222-15	ZNF655	Q8N720	MGKHEHLNLTEDFQS	TEKSYKCDVCGKIFH	51
HR8422A-173-239-15	ZNF655	Q8N720	MGKHEHLNLTEDFQS	SALTRHQRIHTREKP	68
HR8422A-178-218-15	ZNF655	Q8N720	MLNLTEDFQSSECKE	SIPNTEKSYKCDVCG	42
HR8422A-178-234-15	ZNF655	Q8N720	MLNLTEDFQSSECKE	IFHQSSALTRHQRIH	58
HR8422A-182-234-TEV	ZNF655	Q8N720	EDFQSSECKESLMDL	IFHQSSALTRHQRIH	53
HR8422B-430-491-TEV	ZNF655	Q8N720	HRKEKSYECNEYEGS	AHLVQHQSIHTKENS	62
HR8422B-434-491-TEV	ZNF655	Q8N720	KSYECNEYEGSFSHS	AHLVQHQSIHTKENS	58

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8422C-372-430-Av6HT	ZNF655	Q8N720	GIHFREKPYTCSECG	AFSQTSCLIQHHKMH	59
HR8422C-372-470-Av6HT	ZNF655	Q8N720	GIHFREKPYTCSECG	EVLTRQKAFDCDVWE	99
HR8422C-372-475-Av6HT	ZNF655	Q8N720	GIHFREKPYTCSECG	QKAFDCDVWEKNSSQ	104
HR8422D-378-430-Av6HT	ZNF655	Q8N720	KPYTCSECGKDFRLN	AFSQTSCLIQHHKMH	53
HR7623A-279-349-NHT	ZNF658	Q5TYW1	CDKTTAVEYNKVHMA	SQSSAHIVHQKTQAG	71
HR8538A-1-85-Av6HT	ZNF663	Q8NDT4	YTGEKPDECKENFKA	VLKESCLTPNQRIKT	84
HR6955A-195-249-NHT	ZNF664	Q8N3J9	GEKPYRCCGCGKAFS	AFSQSTSLCIHQRVH	55
HR7621A-171-230-NHT	ZNF667	Q5HYK9	PFECSNCRKAFRQIS	ILHMRIHDGKEILDC	60
HR6925A-83-165-NHT	ZNF670	Q9BS34	TFSQDSNLNLNKKVS	ISLTSVDRHMVTHTS	83
HR7883A-267-337-Av6HT	ZNF671	Q8TAW3	KPHKSTKLVSGFLMG	SQSYDLFKHQTVHTG	71
HR7237A-1-66-NHT	ZNF672	Q499Z4	FATSGAVAAGKPYSC	ARAADLRAHRRTHAG	65
HR7236A-365-437-NHT	ZNF674	Q2M3X9	ASDEKPSPTKHWRTH	SGKSHLSVHHRTHTG	73
HR7578A-1-67-15	ZNF675	Q8TD23	MGLLTFRDVAIEFSL	ITCLEQEKEPLTVKR	67
HR7578A-1-74-15	ZNF675	Q8TD23	MGLLTFRDVAIEFSL	KEPLTVKRHEMVNEP	74
HR7578B-131-199-15	ZNF675	Q8TD23	MGLNQCLPTMQSKMF	SHLTRHERNYTKVNF	70
HR7578B-136-194-15	ZNF675	Q8TD23	MLPTMQSKMFQCDKY	SFCMLSHLTRHERNY	60
HR7578B-137-199-15	ZNF675	Q8TD23	MPTMQSKMFQCDKYV	SHLTRHERNYTKVNF	64
HR7578B-140-194-15	ZNF675	Q8TD23	MQSKMFQCDKYVKVF	SFCMLSHLTRHERNY	56
HR7891A-111-163-Av6HT	ZNF676	Q8N7Q3	KVFQCGKYANVFHKC	SFCMLSHLSQHERIY	53
HR7577A-497-565-NHT	ZNF677	Q86XU0	TERSNLTQHKKIHTG	ALFQSSNLGDHQKSY	69
HR8530A-404-487-Av6HT	ZNF678	Q5SXM1	PYKCEECGKVFKQCS	FSSLTRHKRIHTGEK	84
HR7709A-360-411-NHT	ZNF679	Q8IYX0	AFAFSSTLNTHKRIH	NHKSMHTGEKPYKCE	52

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR8058A-136-204-Av6HT	ZNF680	Q8NEM1	KEGYNELNQCLRTTQ	SFCMLSHLTQHIRIH	69
HR7323A-486-575-NHT	ZNF681	Q96N22	AFNQSSILTTHKRIH	PYQCEECGKAFNQSS	90
HR7060A-141-225-NHT	ZNF682	O95780	PSKIFPYNKCVKVFS	KWFSYLTKHKRIHTG	85
HR7991A-155-211-Av6HT	ZNF684	Q5T5D7	VENAYECSECGKAFK	SRKAHLATHQKIHNG	57
HR7327A-962-1050-Na6HT	ZNF687	Q8N1G0	GWTCGLCHSWFPERD	SSRLILEKHVQVRHG	89
HR7862A-150-239-Av6HT	ZNF689	Q96CS4	ICPDCGCTFPDHQAL	VIHTGEKPYHCPDCG	90
HR8314A-328-397-NHT	ZNF692	Q9BU19	KKHLKEHMKLHSDTR	QKASLNWHQRKHAET	70
HR7296A-298-376-NHT	ZNF695	Q8IW36	CEECGKSFKLFPYLT	NQSSHLTEHRRIHTG	79
HR8310A-158-221-Av6HT	ZNF696	Q9H7X3	CGKAFIHSSHVVRHQ	KPYACADCGKAFGQR	64
HR7368A-352-405-NHT	ZNF697	Q5TEC3	PFACGECGKGFVRRS	SWRSDLVKHQRVHTG	54
HR8203A-585-642-Av6HT	ZNF699	Q32M78	KPFECLECGKAFSCP	AYFRRHVKTHTRENI	58
HR7900-118-686-15	ZNF7	P17097	MQNPGFGDVSDSEVW	NRSSRLTQHQKIHMG	570
HR7900-123-686-15	ZNF7	P17097	MGDVSDSEVWLDSHL	NRSSRLTQHQKIHMG	565
HR7900-176-686-15	ZNF7	P17097	MSSGLDCQPLESQGE	NRSSRLTQHQKIHMG	512
HR7900-181-686-15	ZNF7	P17097	MCQPLESQGESAEGM	NRSSRLTQHQKIHMG	507
HR7900-192-686-15	ZNF7	P17097	MEGMSQRCEECGKGI	NRSSRLTQHQKIHMG	496
HR7900-98-686-Av6HT	ZNF7	P17097	DILKSESYGTVVRIS	NRSSRLTQHQKIHMG	589
HR7900A-632-686-Av6HT	ZNF7	P17097	KLHQCEDCEKIFRWR	NRSSRLTQHQKIHMG	55
HR7964A-390-437-Av6HT	ZNF70	Q9UC06	KPYTCECGKAFRHRS	LCGKSFRGSSHLIRH	48
HR8076A-25-71-Av6HT	ZNF700	Q9H0M5	AFEDVAVNFTQEEWT	FRNLTSIGKKWSDQN	47
HR7819A-251-341-Av6HT	ZNF701	Q9NV72	DFHQKRYLACHRCHT	KCEECDKVFSRKSHL	91
HR6936A-105-196-NHT	ZNF705A	Q6ZN79	TMENSLILEDPFECN	TNCFRLRRHKMTHTG	92

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7717A-105-196-NHT	ZNF705G	A8MUZ8	TMENSLILEDPFECN	TNCFHLRRHKMTHTG	92
HR7599A-294-364-NHT	ZNF707	Q96C28	GKAFRTKENLSHHQR	GKGFRHLGFFTRHQR	71
HR8522A-127-192-Av6HT	ZNF708	P17019	GLNRCVTTTQSKIVQ	CMLSQLTQHEIIHTG	66
HR8299A-146-232-Av6HT	ZNF709	Q8N972	GKRFSFRSSFRIHER	HTGEKPYKCKECGKT	87
HR7598A-420-489-NHT	ZNF71	Q9NQZ8	SQSAYLIEHQRIHTG	FSRNTNLTRHLRIHT	70
HR7994A-273-347-Av6HT	ZNF710	Q8N1W2	RLDINVQIDDSYLVE	KQPSHLQTHLLTHQG	75
HR7730A-640-700-NHT	ZNF711	Q9Y462	IHKGRKIHQCRHCDF	RQQNELKKHMKTHTG	61
HR7315A-202-251-NHT	ZNF713	Q8N859	SIKHNSDLIYYQGNY	LTDHIHTAEKPSECG	50
HR8420A-149-221-Av6HT	ZNF718	Q3SXZ3	VKVFHKFSNSNKDKI	AFNWSSILTKHKRIH	73
HR7757A-1-62-NHT	ZNF720	Q7Z2F6	GLLTFRDVAIEFSRE	SKPDLITFLEQRKEP	61
HR8490A-143-197-Av6HT	ZNF730	Q6ZMV8	KIFQCDKYVKVFHKF	CILSHLAQHKKIHTG	55
HR8539A-16-91-Av6HT	ZNF738	Q8NE65	GYPGAERNLLEYSYF	DVSKPDLITCLEQGK	76
HR8339A-526-578-Av6HT	ZNF74	Q16587	KPYKCSECGRAFSQN	MFNWSSHLTEHQRLH	53
HR8360A-96-181-Av6HT	ZNF740	Q8NDX6	KIPKNFVCEHCFGAF	SRTDRLLRHKRMCQG	86
HR8176A-34-86-Av6HT	ZNF747	Q9BV97	PGAVSFADVAVYFSR	HLGALGESPTCLPGP	53
HR8509A-408-460-Av6HT	ZNF749	Q43361	RLYKCSECGKAFSLK	AFVRKSHLVQHQKIH	53
HR7901A-1-56-Av6HT	ZNF75A	Q96N20	YFSQEEWELLDPTQK	KVISCLEQGEEPWVQ	55
HR6964-1-358-Av6HT	ZNF76	P36508	ESLGLHTVTLSDGTT	PYTCSTCGKTYRQTS	357
HR6964-1-369-Av6HT	ZNF76	P36508	ESLGLHTVTLSDGTT	RQTSTLAMHKRSAHG	368
HR6964A-173-267-NHT	ZNF76	P36508	GRLYTTAHHLKVHER	RPFQCPFEGCGRSFT	95
HR6964B-161-251-Av6HT	ZNF76	P36508	GDRAFRCGYKGCGRL	KTSGDLQKHVRTHTG	91
HR6964C-227-276-Av6HT	ZNF76	P36508	CPEELCSKAFKTSGD	CGRSFTTSNIRKVHV	50

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR6964C-227-294-Av6HT	ZNF76	P36508	CPEELCSKAFKTSGD	TGERPYTCPEPHCGR	68
HR6964C-232-294-Av6HT	ZNF76	P36508	CSKAFKTSGDLQKHV	TGERPYTCPEPHCGR	63
HR6964C-235-276-Av6HT	ZNF76	P36508	AFKTSGDLQKHVRTH	CGRSFTTSNIRKVHV	42
HR7027A-389-461-NHT	ZNF765	Q7L2R6	SKTFSHKSSLTYHRR	YSFKSNLEIHQKIHT	73
HR7836A-303-379-NHT	ZNF766	Q5HY98	KCGKVYSSSSYLAQH	RHKFSLTVHQRNHNG	77
HR7774A-291-355-NHT	ZNF768	Q9H5H4	CEVCSKAFSQSSDLI	GQKPYKCPHCGKAFG	65
HR8123A-304-359-Av6HT	ZNF77	Q15935	SFSCYSSFRDHVRTH	THSGEKPYECKECGK	56
HR7610A-1-82-NHT	ZNF770	Q6IQ21	AENNLKMLKIQQCVV	VHLERHQLTHSLPFK	80
HR7742A-127-216-NHT	ZNF771	Q7L3S4	RFSAASNLRQHRRRH	PYACADCGTRFAQSS	90
HR7325A-385-458-NHT	ZNF772	Q68DY9	KYFGHKYRLIKHWSV	SHKHVLVQHHRIHTG	74
HR8057A-186-243-Av6HT	ZNF773	Q6PK81	AGKRHYKCSECGKAF	SHKSNLFIHQIVHTG	58
HR7824A-105-159-Av6HT	ZNF775	Q96BV0	GHFVCLDCGKRFSWW	SQKPNLARHQRHHTG	55
HR7510A-199-288-NHT	ZNF776	Q68DI1	IPLQGGKTHYICGES	WYKAHLTEHQRVHTG	90
HR7596A-588-631-NHT	ZNF780A	O75290	KPFECKECGKAFRLH	ECGKVFSLPTQLNRH	49
HR7666A-735-815-NHT	ZNF780B	Q9Y6R6	GLLTQLAQHQIIHTG	KLVQVRNPLNVRNVG	81
HR7344A-514-586-NHT	ZNF782	Q6ZMW2	AFKLKSGLRKHHRTH	SQKSNLRVHHRTHTG	73
HR8508A-34-122-Av6HT	ZNF783	Q6ZMS7	SYLYSTEITLWTVVA	LLQRRLENVENLLRN	89
HR8227A-257-318-Av6HT	ZNF785	A8K8V0	ACSDCKSRFTYPYLL	RIHTGEKPYPCPDCG	62
HR7825A-407-473-NHT	ZNF786	Q8N393	RLRRLLQVHQHAHGG	GRNFRQRGQLLRHQR	67
HR7915A-65-146-Av6HT	ZNF787	Q6DD87	PYICNECGKSFSHWS	SWSSNLMQHQRIHTG	82
HR8290A-153-225-Av6HT	ZNF789	Q5FWF6	GFLQNLNLIQDQNAQ	RRKAWFDQHQRIHFL	73
HR7454A-424-498-NHT	ZNF79	Q15937	KFFSESSALIRHHII	CSSAFVRHQRLHAGE	75

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7139A-544-636-NHT	ZNF790	Q6PG37	IWGSQLTRHKKIHTD	FEKAFSSSSHFISLL	93
HR8412A-506-576-Av6HT	ZNF791	Q3KP31	IYPTSFQGHMRMHTG	SVSTSLKKHMRMHNR	71
HR8238A-532-599-15	ZNF792	Q3KQV3	MRPYECSECGKTFRQ	IRERSMENVLLPCSQ	69
HR8238A-532-599-Av6HT	ZNF792	Q3KQV3	RPYECSECGKTFRQR	IRERSMENVLLPCSQ	68
HR7343A-484-570-NHT	ZNF799	Q96GE5	AFSCFQYLSQHRRTH	REKPYECQQCGKAFT	87
HR4794D-252-417-15	ZNF8	P17098	VQDKPYKCTDCGKSF	GKGFRHSSSLAQHQR	166
HR4794D-256-412-15	ZNF8	P17098	PYKCTDCGKSFNHNA	ECNHCGKGFRHSSSL	157
HR4794D-274-417-15	ZNF8	P17098	VHKRIHTGERPYMCK	GKGFRHSSSLAQHQR	144
HR4794D-280-412-15	ZNF8	P17098	TGERPYMCKECGKAF	ECNHCGKGFRHSSSL	133
HR4794E-339-417-15	ZNF8	P17098	KPYECQDCGRAFNQN	GKGFRHSSSLAQHQR	79
HR4794E-344-411-15	ZNF8	P17098	QDCGRAFNQNSSLGR	YECNHCGKGFRHSSS	68
HR7462A-85-152-NHT	ZNF80	P51504	AFPEKVDFVRPMRIH	CGKTFSYHSVFIQHR	68
HR8132A-480-640-Av6HT	ZNF800	Q2TB10	GFDFKQLYCKLCKRQ	AFAKKTYLEHHKKTH	161
HR7014A-407-492-NHT	ZNF808	Q8N4W9	AFNHQSSLARHHILH	TGEKTYKCNECRKTF	86
HR7427A-226-281-NHT	ZNF81	P51508	VFTQNSSYSHHENTH	FPIGEKANTCTEFGK	56
HR8002A-591-645-Av6HT	ZNF816A	Q0VGE8	KPYKCNECGKVFNQK	TGQSTLIHHQAIHGC	55
HR8532A-59-113-Av6HT	ZNF818P	Q6ZRF7	KRSLTNVCGKVLSQN	TQGSRFINHQIVHTG	55
HR7648A-533-610-NHT	ZNF823	P16415	GKAFSWLTCLLRHER	RSLHRHKRTHWKDTL	78
HR7405A-703-761-NHT	ZNF828	Q96JM3	KRGKGKYYCKICCCR	FLLESLLKNHVAAHG	59
HR8193A-154-208-Av6HT	ZNF829	Q3KNS6	KPWECKICGKTFNQN	SRGSLVTRHQRIHTG	55
HR7002A-127-184-15	ZNF83	P51522	MGKIFNKKSNLASHQ	IHTGEKPYKCNECGK	59
HR7002B-70-148-15	ZNF83	P51522	MTYECNFVDSLFTQK	SNLASHQRIHTGEKP	80

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7002B-74-145-15	ZNF83	P51522	MNFVDSLFTQKEKAN	NKKSNLASHQRIHTG	73
HR7002B-89-145-15	ZNF83	P51522	MGTEHYKCSERGKAF	NKKSNLASHQRIHTG	58
HR8533A-9-176-Av6HT	ZNF833P	Q6ZTB9	PYKCKFCGKAFDNLH	FSSFHSHEGVHTGEK	168
HR8077A-450-518-Av6HT	ZNF835	Q9Y2P0	SQGSSLALHQRTHTG	AFSFSSALIRHQRTH	69
HR8234A-206-287-15	ZNF836	Q6ZNA1	MTQLEKTHIREKPYM	PYQCGVCGKIFRQNS	83
HR8234A-206-287-Av6HT	ZNF836	Q6ZNA1	TQLEKTHIREKPYMC	PYQCGVCGKIFRQNS	82
HR7704A-427-482-NHT	ZNF837	Q96EG3	AFKGRSGLVQHQRAH	LHSGEKPYICRDCGK	56
HR8403A-681-738-Av6HT	ZNF84	P51523	KPYGCSECRKAFSQK	SQLINHQRTHTVKKS	58
HR8489A-197-283-Av6HT	ZNF841	Q6ZN19	RGKPYQCDVCGRIFR	SSSLATHQTVHTGDK	87
HR8361A-897-970-Av6HT	ZNF845	Q96IR2	NQQAHLACHHRIHTG	AKLARHHRIHTGKKH	74
HR7777A-476-525-NHT	ZNF846	Q147U1	KPYACKECGKAFRYS	CGKNFTQSSALAKHL	50
HR7585A-536-595-NHT	ZNF85	Q03923	KPYTCEECGKAFNQS	LTKHKIIHTGEKLQI	60
HR8493A-449-519-Av6HT	ZNF852	B4DLD7	SYNSSLMVHQRTHTG	SQRSTFNHHQRTHAG	71
HR8177-500-555-Av6HT	ZNF880	Q6PDB4	VFSHNSHLARHRQIH	IHTGEKPYRCHECGK	56
HR6923A-519-589-NHT	ZNF90	Q03938	KRSSVLSKHKIIHTG	NLSSDLNTHKRIHIG	71
HR8498A-486-572-Av6HT	ZNF98	A6NK75	GEKPYKCEECGKAFN	IAKISKYKRNCAGEK	87
HR8425A-706-753-Av6HT	ZNF99	A8MXY4	AFNNSSTLRKHEIIH	IHTGKKPYKCEECGK	48
HR7451A-925-1258-Av6HT	ZNFX1	Q9P2E3	LDLSSRWQLYRLWLQ	SKIIHTLRENNQIGP	334
HR6880A-166-356-NHT	ZRSR2	Q15696	EKDRANCPFYSKTGA	ANRDIYLSPDRTGSS	191
HR7933A-24-120-Av6HT	ZSCAN1	Q8NBB4	ADPGPASPRDTEAQR	CREAASLVEDLTQMC	97
HR7806A-1-70-NHT	ZSCAN10	Q96SZ4	GPRASLSRLRELCGH	DGEEVVLLLEGIHRE	69
HR8495A-9-132-Av6HT	ZSCAN12	O43309	AHMDQDEPLEVKIEE	VTVLEDLERELDEPG	124

Vector	HUGO	Uniprot	First 15aa	Last 15aa	Construct Length
HR7081A-224-291-TEV	ZSCAN16	Q9H4T2	GRSEWQQRERRRYKC	SHLIGHHRVHTGVKP	68
HR7530A-36-127-NHT	ZSCAN21	Q9Y5A6	KYLPSLEMFRQRFRQ	AEEAVTLLEDLEREL	92
HR7904A-40-135-Av6HT	ZSCAN22	P10073	DHIAHSEAARLRFRH	AVLVEDLTQVLDKRG	96
HR7247A-37-133-NHT	ZSCAN23	Q3MJ62	SRNNPHTREIFRRRF	AVTVLEDLRELDDP	97
HR8429A-9-104-Av6HT	ZSCAN29	Q8IWY8	ENGTNSETFRQRFRR	VTLVEDLEREPGRPR	96
HR6932A-12-134-NHT	ZSCAN30	Q86W11	APEEQEGLLVVKVEE	VTMLEELEKELEEPR	123
HR7089A-35-123-Av6HT	ZSCAN4	Q8NAM6	REEGISEFSRMVLNS	KSSGKNLERFIEDLT	89
HR7089A-35-130-NHT	ZSCAN4	Q8NAM6	REEGISEFSRMVLNS	ERFIEDLTDDSINPP	96
HR7089A-46-123-Av6HT	ZSCAN4	Q8NAM6	VLNSFQDSNNSYARQ	KSSGKNLERFIEDLT	78
HR7089A-46-130-Av6HT	ZSCAN4	Q8NAM6	VLNSFQDSNNSYARQ	ERFIEDLTDDSINPP	85
HR6950A-39-129-NHT	ZSCAN5A	Q9BUG6	DPEISHVNFRMFSCP	LEDLLRNNRRPKKWS	91
HR8432A-35-142-Av6HT	ZSCAN5B	A6NJL1	NHDRNPETWHMNFRM	WSIVNLLGKEYLMLN	108
HR7759A-37-130-NHT	ZSCAN5C	A6NGD5	DSDPETCHVNFRMFS	EDLLRNNRRPKKWSV	94
HR8021A-314-557-Av6HT	ZUFSP	Q96AP4	DGKTKTSGIIEALHR	LKHKQYQILAVEGAL	244
HR7812A-358-413-NHT	ZXDA	P98168	NSFKCEVCEESFPTQ	TFITVSALFSHNRAH	56
HR7168A-360-417-NHT	ZXDB	P98169	QENSFKCEVCEESFP	TFITVSALFSHNRAH	58
HR7131A-652-715-TEV	ZZZ3	Q8IYH5	NQLWTVEEQKKLEQL	KYFIKLTKAGIPVPG	64

References:

Acton, T. B., et al., 2011. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. Methods Enzymol. 493, 21-60.

Agaton et al., Molecular & Cellular Proteomics 2:405-414, 2003.

Bindewald, E., et al., CyloFold: secondary structure prediction including pseudoknots. Nucleic Acids Res. 38, W368-72.

Brodskii, L. I., et al., 1995. [GeneBee-NET: An Internet based server for biopolymer structure analysis]. Biokhimiia. 60, 1221-30.

Crowe, J., et al., 1994. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. Methods Mol Biol. 31, 371-87.

Ding, Y., et al., 2004. Sfold web server for statistical folding and rational design of nucleic acids. Nucleic Acids Res. 32, W135-41.

Do, C. B., et al., 2006. CONTRAfold: RNA secondary structure prediction without physics-based models. Bioinformatics. 22, e90-8.

Gonzalez de Valdivia, E. I., Isaksson, L. A., 2004. A codon window in mRNA downstream of the initiation codon where NGG codons give strongly reduced gene expression in Escherichia coli. Nucleic Acids Res. 32, 5198-205.

Gruber, A. R., et al., 2008. The Vienna RNA websuite. Nucleic Acids Res. 36, W70-4.

Hamada, M., et al., 2009. Predictions of RNA secondary structure by combining homologous sequence information. Bioinformatics. 25, i330-8.

Jansson, M.; et al., 1996. High-level production of uniformly ^{15}N- and ^{13}C-enriched fusion proteins in *Escherichia coli.* B. J. Biomol. NMR. 7, 131 - 141.

Kapust, R. B., et al., 2002. The P1' specificity of tobacco etch virus protease. Biochem Biophys Res Commun. 294, 949-55.

Kudla, G., et al., 2009. Coding-sequence determinants of gene expression in Escherichia coli. Science. 324, 255-8.

Lamla, T., Erdmann, V. A., 2004. The Nano-tag, a streptavidin-binding peptide for the purification and detection of recombinant proteins. Protein Expr Purif. 33, 39-47.

Lui et al., 2002, Loopy proteins appear conserved in evolution. J Mol Biol. 322~ 53-64)

Markham, N. R., Zuker, M., 2008. UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol. 453, 3-31.

Mathews, D. H., et al., 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc Natl Acad Sci U S A. 101, 7287-92.

Netzer and Hartl, 1997. Recombination of protein domains facilitated by co-translational folding in eukaryotes. Nature. 358~ 343-9.

Nomura, M., et al., 1984. Influence of messenger RNA secondary structure on translation efficiency. Nucleic Acids Symp Ser. 173-6.

Quan, J., et al., 2011. Parallel on-chip gene synthesis and application to optimization of protein expression. Nat Biotechnol. 29, 449-52.

Reeder, J., et al., 2007. pknotsRG: RNA pseudoknot folding including near-optimal structures and sliding windows. Nucleic Acids Res. 35, W320-4.

Rivas, E., Eddy, S. R., 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. J Mol Biol. 285, 2053-68.

Rocha, E. P., et al., 1999. Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. Nucleic Acids Res. 27, 3567-76.

Sharp, P. M., Li, W. H., 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15, 1281-95.

Scholle, M. D., et al., 2004. In vivo biotinylated proteins as targets for phage-display selection experiments. Protein Expr Purif. 37, 243-52.

Schroeder, S. J., et al., 2011. Ensemble of secondary structures for encapsidated satellite tobacco mosaic virus RNA consistent with chemical probing and crystallography constraints. Biophys J. 101, 167-75.

Voss, B., et al., 2006. Complete probabilistic analysis of RNA shapes. BMC Biol. 4, 5.

Xayaphoummine, A., et al., 2005. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. Nucleic Acids Res. 33, W605-10.

Xayaphoummine, A., et al., 2003. Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations. Proc Natl Acad Sci U S A. 100, 15310-5.

Zuker, M., Stiegler, P., 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 9, 133-48.

The foregoing examples and description of the preferred embodiments should be taken as illustrating, rather than as limiting the present invention as defined by the claims. As will be readily appreciated, numerous variations and combinations of the features set forth above can be utilized without departing from the present invention as set forth in the claims. Such variations are not regarded as a departure from the scope of the invention, and all such variations are intended to be included within the scope of the following claims. All references cited herein are incorporated herein in their entireties.

CLAIMS

WHAT IS CLAIMED IS:

1. A method of preparing an expression vector, wherein the expression vector comprises, in order of position: a first nucleic acid sequence encoding a 5' untranslated region of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag; and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame with the second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the method comprises specifically modifying the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

2. The method of claim 1, further comprising specifically modifying the second nucleic acid sequence to reduce the presence of rare codons.

3. The method of any one of claims 1-2, wherein nucleotides within about the last 100 nucleotides of the first nucleic acid sequence are modified.

4. The method of any one of claims 1-3, wherein nucleotides within about the first 90 nucleotides of the second nucleic acid sequence are modified.

5. The method of any one of claims 1-4, wherein the expression vector further comprises a target protein coding sequence inserted into the vector in-frame with the nucleic acid tag sequence to encode a fusion protein comprising the polypeptide tag and the target protein.

6. The method of claim 5, wherein the target protein coding sequence is not modified to minimize RNA secondary structure.

7. The method of any one of claims 5-6, wherein the target protein coding sequence is not modified to reduce the presence of rare codons.

8. The method of any one of claims 1-7, wherein the second nucleic acid sequence encodes at least one affinity purification tag.

9. The method of any one of claims 1-8, wherein the second nucleic acid sequence encodes more than one affinity purification tag.

10. The method of claim 9, wherein the second nucleic acid sequence encodes two affinity purification tags.

11. The method of any one of claims 8-10, wherein the encoded affinity purification tag(s) is/are selected from a Streptavidin binding moiety, a maltose binding protein moiety, and a HIS tag.

12. The method of claim 11, wherein the Streptavidin binding moiety is a Nano-tag or a biotinylated Avi-tag.

13. The method of any one of claims 1-12, wherein the second nucleic acid sequence encodes at least one solubility enhancement tag.

14. The method of any one of claims 1-13, wherein the second nucleic acid sequence encodes more than one solubility enhancement tag.

15. The method of claim 14, wherein the second nucleic acid sequence encodes two solubility enhancement tags.

16. The method of any one of claims 13-15, wherein the encoded solubility enhancement tag(s) is/are selected from a maltose binding protein tag, a protein G B1 domain tag, and a myxococcus protein S tag.

17. The method of any one of claim 1-16, wherein the second nucleic acid sequence further encodes at least one protease recognition site.

18. The method of claim 17, wherein the protease recognition site is a Tobacco Etch Virus (TEV), Thrombin, Factor Xa or a HRV 3C protease recognition site.

19. The method of any one of claims 5-18, wherein the target protein coding sequence encodes a transcription factor, a transcription factor domain, an epigenetic regulatory factor, or an epigenetic regulatory factor domain.

20. The method of any one of claims 5-19, wherein the target protein coding sequence encodes a polypeptide sequence as described in Table 2.

21. The method of any one of claims 5-20, wherein the target protein coding sequence encodes a protein antigen for producing an affinity capture reagent.

22. The method of claim 21, wherein the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

23. The method of any one of claims 5-20, wherein the target protein coding sequence encodes a protein antigen for producing an antibody or Fab by phage display.

24. The method of any one of claims 5-23, wherein the expression of the target protein is 1.5 fold greater than the expression of a target protein generated from an expression vector that was not modified as described in claim 1 or 2.

25. An expression vector prepared using the method of any one of claims 1-24.

26. An expression vector comprising, in order of position: a first nucleic acid sequence encoding a 5' untranslated region of an expressed mRNA that comprises a ribosome binding site (RBS); a second nucleic acid sequence encoding a polypeptide tag; and a cloning site, wherein the cloning site enables a target protein coding sequence to be inserted into the vector in-frame with the

second nucleic acid sequence to encode a fusion protein comprising the polypeptide tag and the target protein; and wherein the nucleic acid sequence encoding (i) the 5' untranslated region and (ii) the adjacent polypeptide tag has been specifically modified to minimize RNA secondary structure both within and/or between these two regions of the mRNA.

27. The expression vector of claim 26, wherein the second nucleic acid sequence has been specifically modified to reduce the presence of rare codons.

28. The expression vector of any one of claims 26-27, wherein nucleotides within about the last 100 nucleotides of the first nucleic acid sequence have been modified.

29. The expression vector of any one of claims 26-28, wherein nucleotides within about the first 90 nucleotides of the second nucleic acid sequence have been modified.

30. The expression vector of any one of claims 26-29, further comprising a target protein coding sequence inserted into the vector in-frame with the nucleic acid tag sequence to encode a fusion protein comprising the polypeptide tag and the target protein.

31. The expression vector of claim 30, wherein the target protein coding sequence has not been modified to minimize RNA secondary structure.

32. The expression vector of any one of claims 30-31, wherein the target protein coding sequence has not been modified to eliminate rare codons.

33. The expression vector of any one of claims 26-32, wherein the second nucleic acid sequence encodes at least one affinity purification tag.

34. The expression vector of any one of claims 26-33, wherein the second nucleic acid sequence encodes more than one affinity purification tag.

35. The expression vector of claim 34, wherein the second nucleic acid sequence encodes two affinity purification tags.

36. The expression vector of any one of claims 33-35, wherein the encoded affinity purification tag(s) is/are selected from a Streptavidin binding moiety, a maltose-binding protein moiety and a HIS tag.

37. The expression vector of claim 36, wherein the Streptavidin binding moiety is a Nano-tag or a biotinylated Avi-tag.

38. The expression vector of any one of claims 26-37, wherein the second nucleic acid sequence encodes at least one solubility enhancement tag.

39. The expression vector of any one of claims 26-38, wherein the second nucleic acid sequence encodes more than one solubility enhancement tag.

40. The expression vector of claim 39, wherein the second nucleic acid sequence encodes two solubility enhancement tags.

41. The expression vector of any one of claims 38-40, wherein the encoded solubility enhancement tag(s) is/are selected from a maltose binding protein tag, a protein G B1 domain tag, and a myxococcus protein S tag.

42. The expression vector of any one of claims 26-41, wherein the second nucleic acid sequence further encodes at least one protease recognition site.

43. The expression vector of claim 42, wherein the protease recognition site is a Tobacco Etch Virus (TEV), Thrombin, Factor Xa or a HRV 3C protease recognition site.

44. The expression vector of any one of claims 30-43, wherein the target protein coding sequence encodes a transcription factor, a transcription factor domain, an epigenetic regulatory factor, or an epigenetic regulatory factor domain.

45. The expression vector of any one of claims 30-44, wherein the target protein coding sequence encodes a polypeptide sequence as described in Table 2.

46. The expression vector of any one of claims 30-45, wherein the target protein coding sequence encodes a protein antigen for producing an affinity capture reagent.

47. The expression vector of claim 46, wherein the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

48. The expression vector of any one of claims 30-45, wherein the target protein coding sequence encodes a protein antigen for producing an antibody or Fab by phage display.

49. The expression vector of any one of claims 30-48, wherein the target protein is expressed at a 1.5-fold higher level than a target protein generated from an expression vector that was not modified as described in claim 26 or 27.

50. A host cell comprising the expression vector of any of claims 30-49.

51. A method for expressing a target protein in a host cell, comprising culturing the host cell of claim 50 for a period of time under conditions permitting expression of the target protein.

52. The method of claim 51, wherein the target protein is a protein antigen for producing an affinity capture reagent.

53. The method of claim 52, wherein the affinity capture reagent is an antibody, an antibody fragment, or an aptamer.

54. The method claim 51, wherein the target protein is a protein antigen for producing an antibody or Fab by phage display.

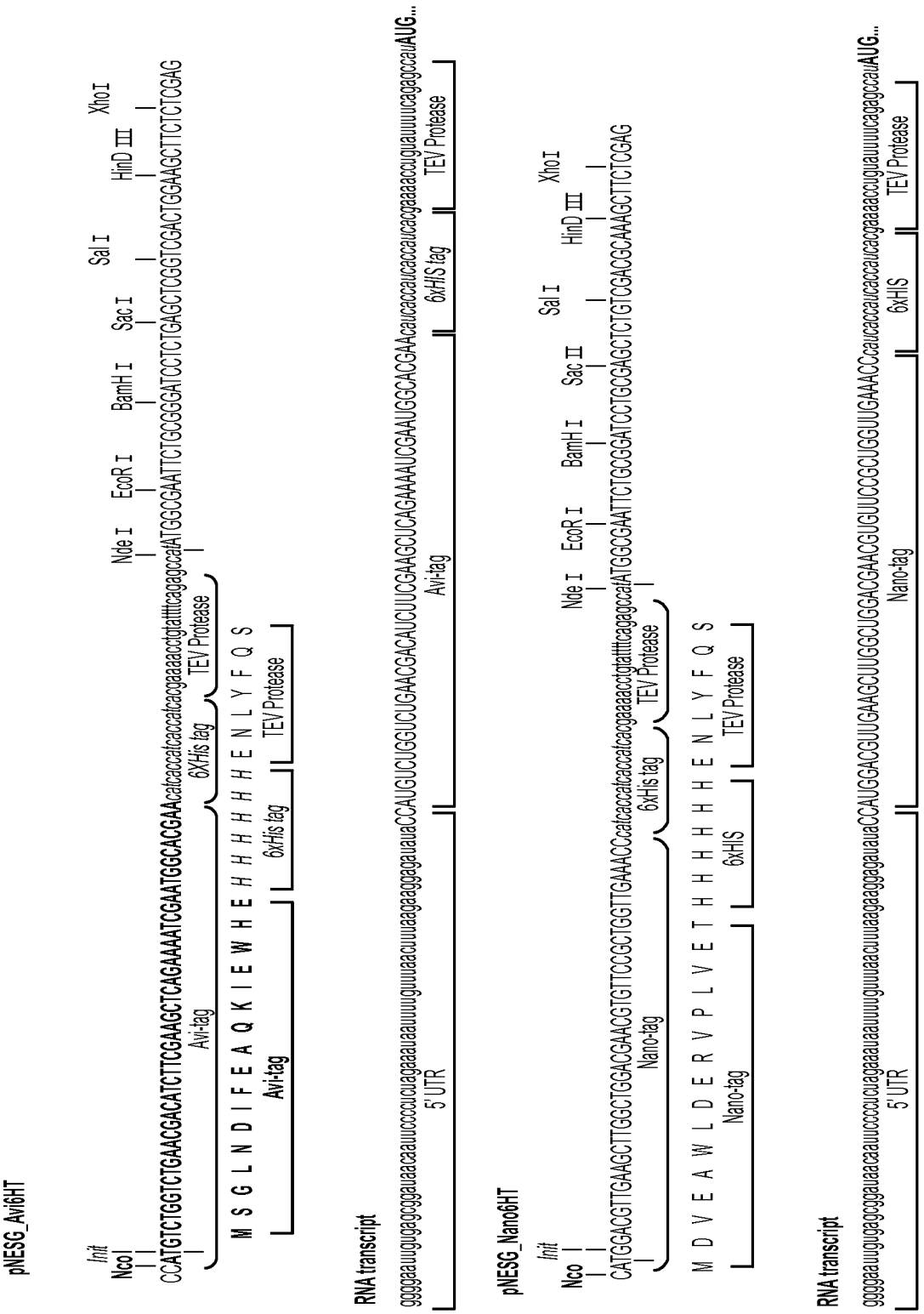

pNESG_Avi6HT
Init
NcoI
Nde I
EcoR I
BamH I
Sac I
Sal I
HinD III
XhoI
CCATGTCTGGTCTGAACGACATCTTCGAAGCTCAGAAAATCGAATGGCACGAA catcaccatcaccatcacgaaaacctgtattttcagagccatATGGCGAATTCTGGGGGATCCTCTGAGCTCGGTCGACTGGAAGCTTCTCTCGAG
Avi-tag
6XHis tag
TEV Protease
M S G L N D I F E A Q K I E W H E H H H H H E N L Y F Q S
Avi-tag
6xHis tag
TEV Protease
RNA transcript
ggggaauuguguggaggcggauaacaauucccccuucuagaaauaauuuuguuuaacuuuaagaaggagauauacaCCAUGUCUGGUCUCUGAACGACAUCUUCGAAGCUCAGAAAAUCGAAUGGCACGAA caucaccaucaccaucacgaaaacctgtattttcagagccaauAUG...
5' UTR
Avi-tag
6xHIS tag
TEV Protease
pNESG_Nano6HT
Init
Nco I
Nde I
EcoR I
BamH I
Sac II
Sal I
HinD III
XhoI
CATGGACGTTGAAGCTTGGCTGGAGGAAGGTGTTCCGCTGGTTGAAACC catcaccatcaccatcacgaaaacctgtattttcagagccatATGGCGAATTCTGGGGATCCTCTGCGAGGCTCTGTCGACGGCAAAGCTTCTCGAG
Nano-tag
6xHis tag
TEV Protease
M D V E A W L D E R V P L V E T H H H H H H E N L Y F Q S
Nano-tag
6xHIS
TEV Protease
RNA transcript
ggggaauuguguggaggcggauaacaauucccccuucuagaaauaauuuuguuuaacuuuaagaaggagauauacaCCAUGGACGTTGAAGCTTGGCTGGAGGAAGGTGTTCCGCTGGTTGAAACC caucaccaucaccaucacgaaaacctgtauuuucagagccaauAUG...
5' UTR
Nano-tag
6xHIS
TEV Protease

FIG. 1

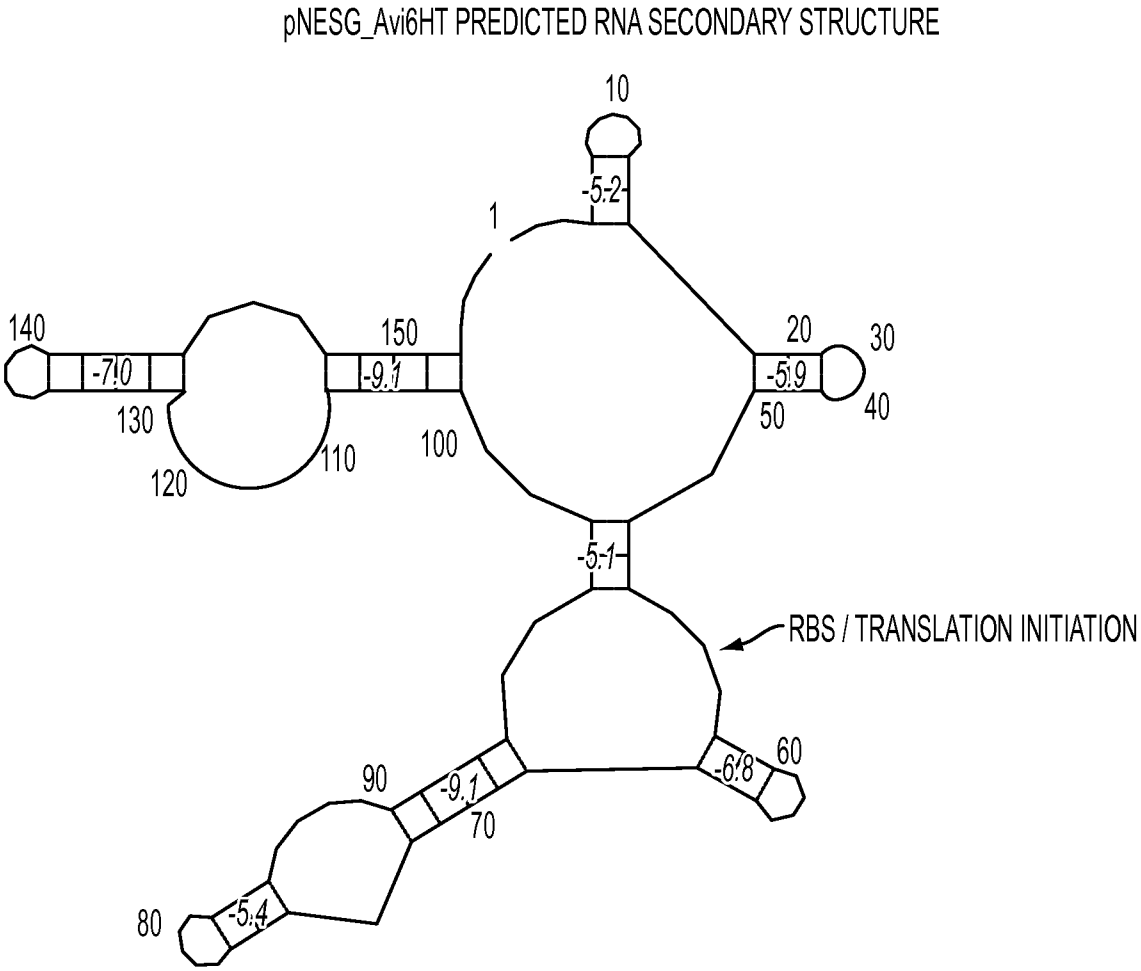

pNESG_Avi6HT PREDICTED RNA SECONDARY STRUCTURE
10
-5.2-
1
140
150
20
30
-7.0
-9.1
-5.9
130
50
40
120
110
100
-5.1-
RBS / TRANSLATION INITIATION
60
-6.8
90
-9.1
70
80
-5.4

FIG. 2

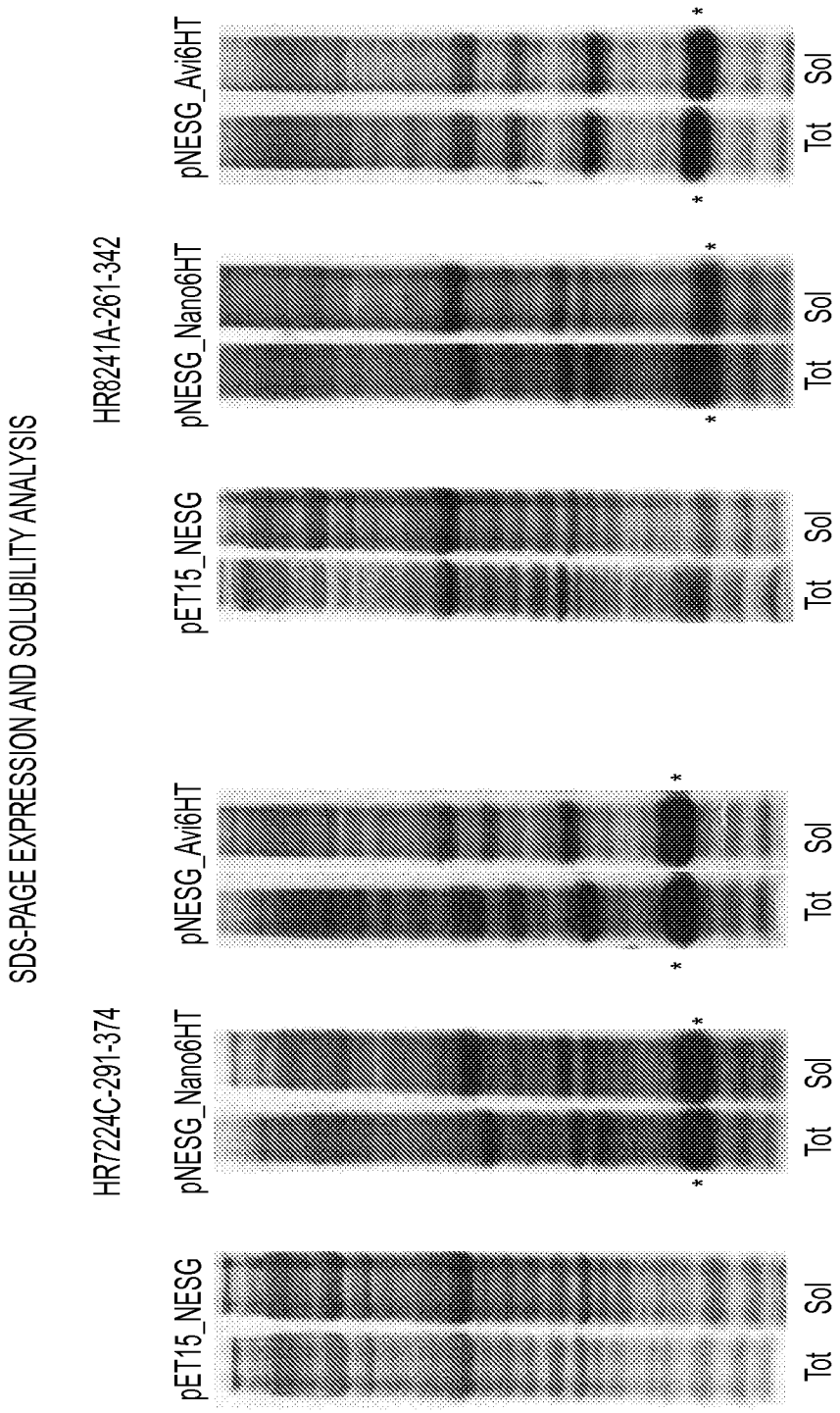

SDS-PAGE EXPRESSION AND SOLUBILITY ANALYSIS
HR7224C-291-374
HR8241A-261-342
pET15_NESG
pNESG_Nano6HT
pNESG_Avi6HT
Tot
Sol
*

FIG. 3

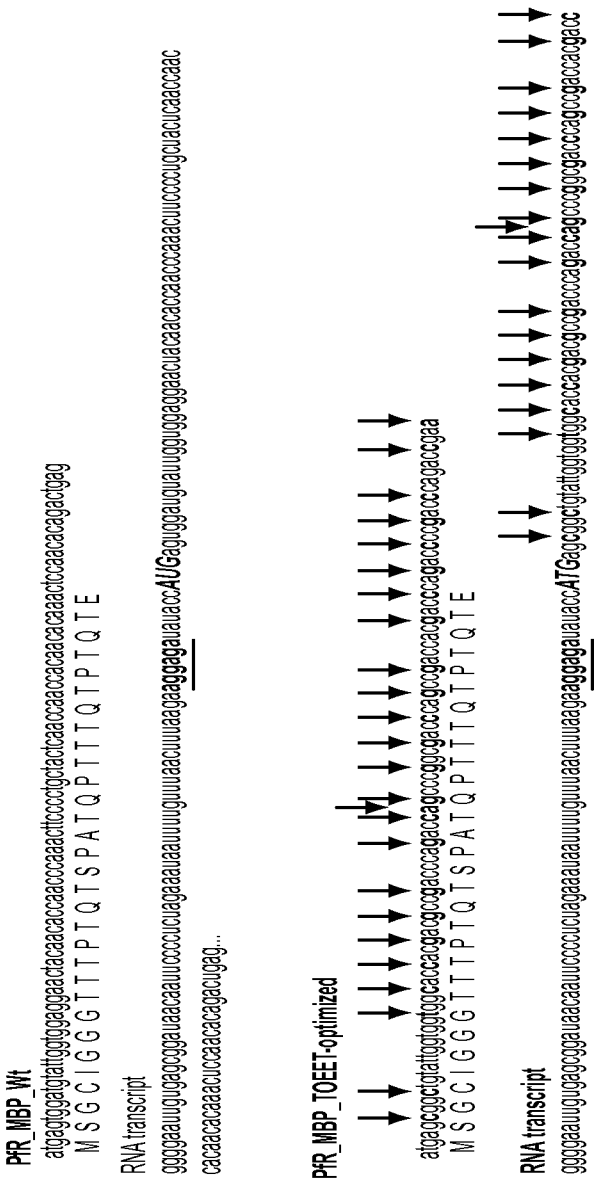

PKR_MBP_Wt
atgagtggatgtattggtgggaggaatacacaacacaaccaaccaaacttcccctgctactcaacaacacaacacaacaactccaacacagactgag
M S G C I G G G T T T P T Q T S P A T Q P T T T Q T P T Q T E
RNA transcript
gggagaauuguggagcggaauaacaauucccccuagaaauaauuuuguuuaacuuuaagaaggagauauaccAUGaguggauguauuggguggaggaacuacaacaccaacccaaacuuccccuguacucaacaac
cacaacacaaacucaacacagacuggag...
PKR_MBP_TOEET-optimized
atgagcggctgtattggtggtggcaccacggacggcccagcccggcgaccagcccagcgaccacgcccggacccagcccagcgaccacggaa
M S G C I G G G T T T P T Q T S P A T Q P T T T Q T P T Q T E
RNA transcript
gggagauuguggagcggaauaacaauucccccuagaaauaauuuuguuuaacuuuaagaaggagauauaccATGagcggctgtattggtggtggcaccacggacggcccagcccggcgaccagcccagcgaccacggacc
cagacccccgacccagacccgaa...

FIG. 4

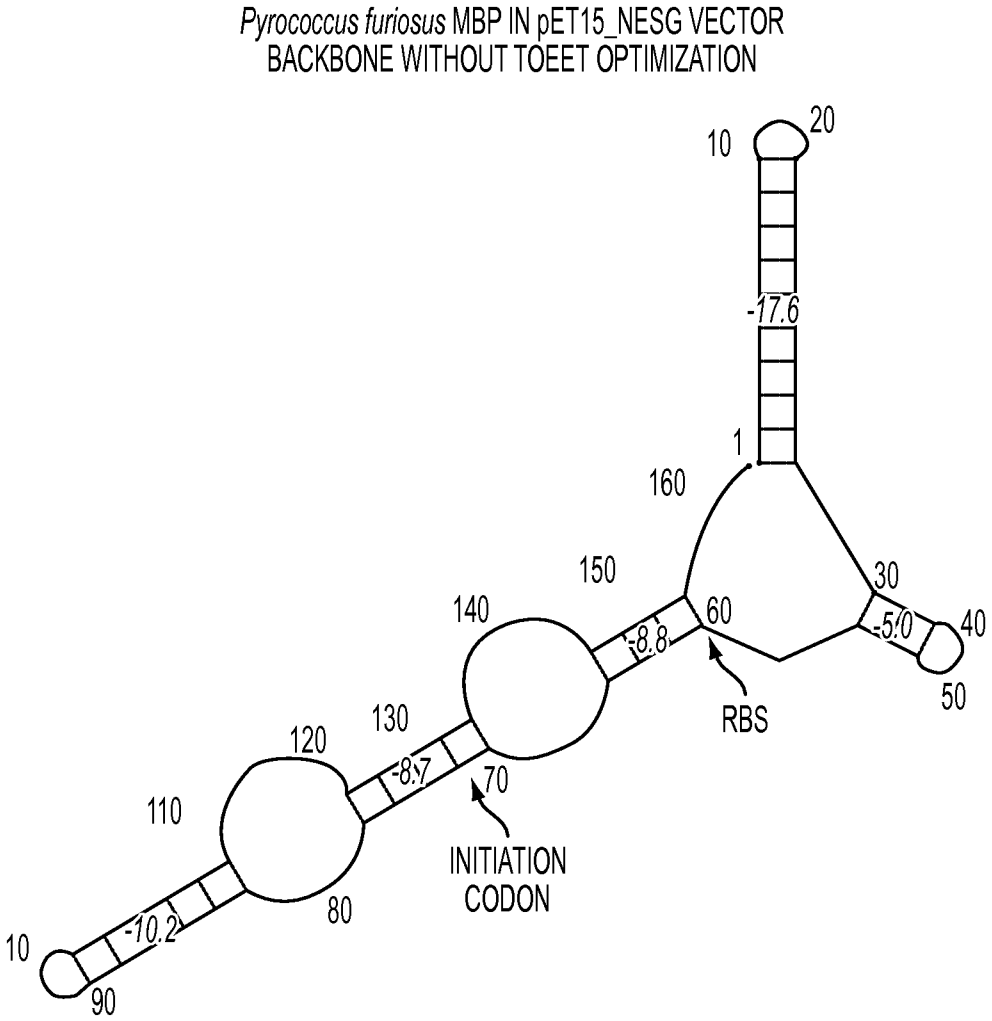

Pyrococcus furiosus MBP IN pET15_NESG VECTOR
BACKBONE WITHOUT TOEET OPTIMIZATION
10
20
-17.6
1
160
150
140
30
60
-8.8
-5.0
40
50
RBS
130
120
-8.7
70
110
INITIATION
CODON
80
-10.2
10
90

FIG. 5

Pyrococcus furiosus MBP IN pET15_NESG VECTOR
BACKBONE WITH TOEET OPTIMIZATION

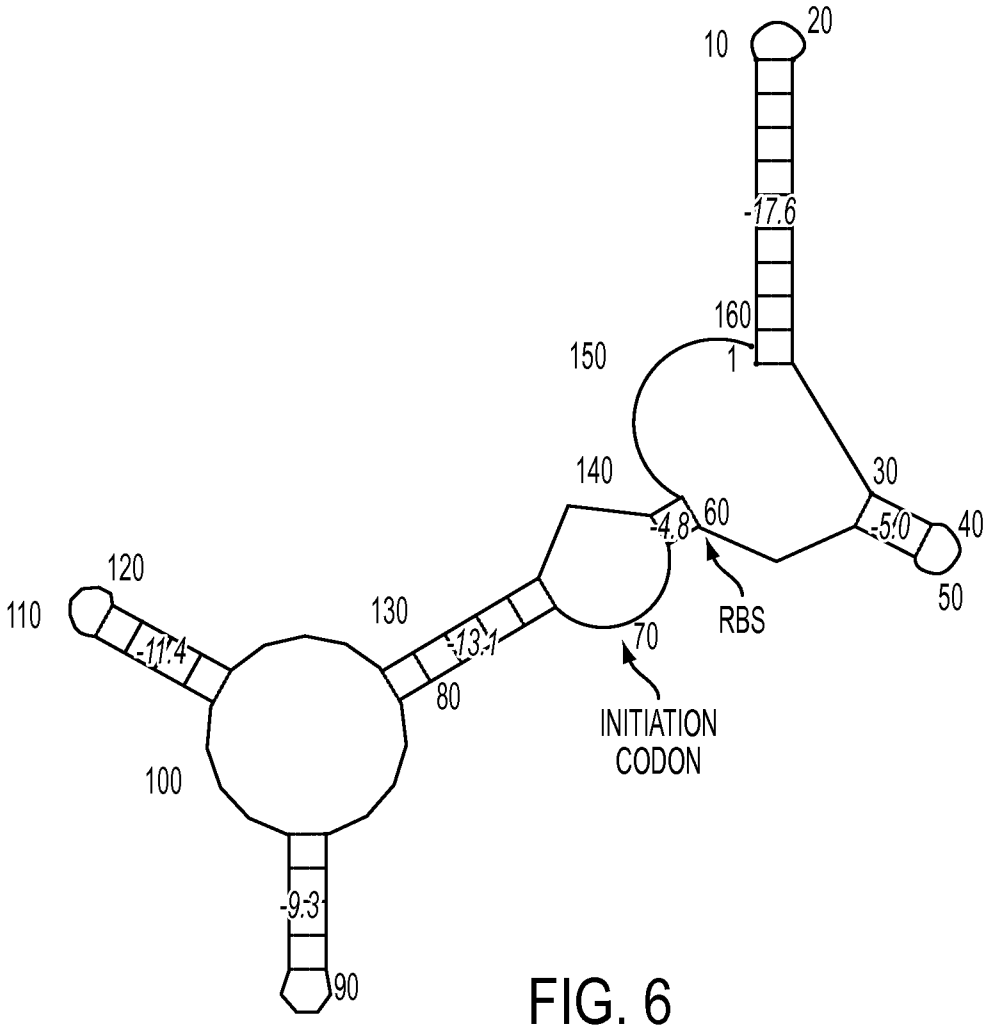


FIG. 6

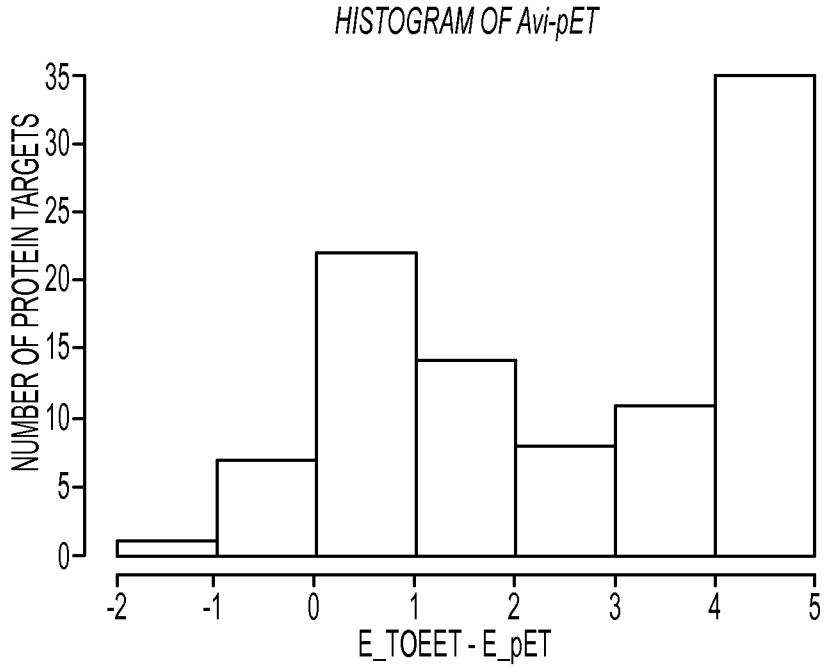

HISTOGRAM OF Avi-pET
NUMBER OF PROTEIN TARGETS
0
5
10
15
20
25
30
35
-2
-1
0
1
2
3
4
5
E_TOEET - E_pET

FIG. 7a

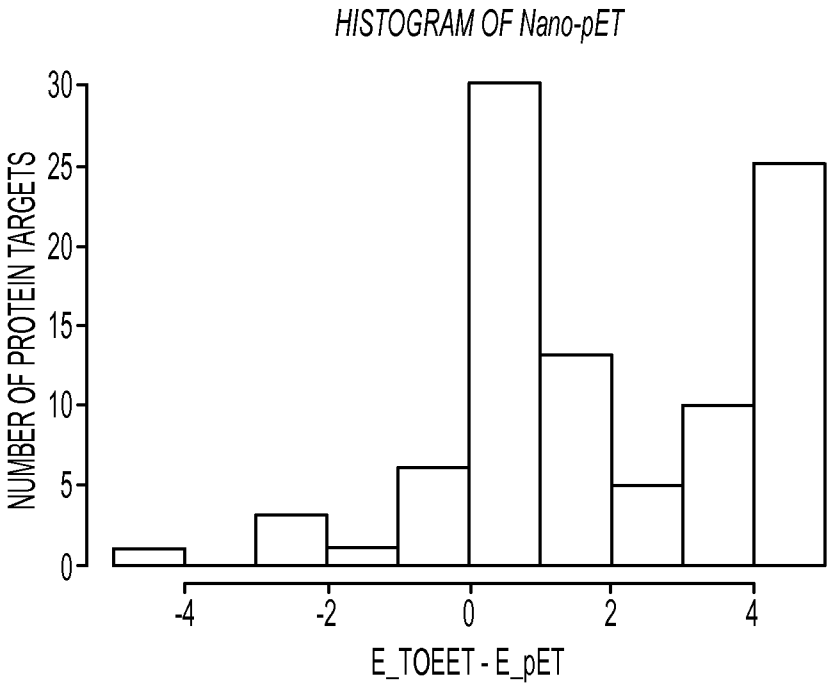

HISTOGRAM OF Nano-pET
NUMBER OF PROTEIN TARGETS
0
5
10
15
20
25
30
-4
-2
0
2
4
E_TOEET - E_pET

FIG. 7b