

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 November 2006 (23.11.2006)

PCT

(10) International Publication Number  
**WO 2006/123097 A2**

(51) International Patent Classification:  
C12N 15/85 (2006.01)

(74) Agent: **GODDARD, David, John**; Harrison Goddard  
Foote, Orlando House, 11c Compstall Road, Marple  
Bridge, Stockport SK6 5HH (GB).

(21) International Application Number:  
PCT/GB2006/001656

(81) Designated States (unless otherwise indicated, for every  
kind of national protection available): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV,  
LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI,  
NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG,  
SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US,  
UZ, VC, VN, YU, ZA, ZM, ZW.

(22) International Filing Date: 9 May 2006 (09.05.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
0509965.0 17 May 2005 (17.05.2005) GB  
60/682,277 18 May 2005 (18.05.2005) US

(84) Designated States (unless otherwise indicated, for every  
kind of regional protection available): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT,  
RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,  
GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (for all designated States except US): **SERO-  
LOGICALS INVESTMENT COMPANY INC.**  
[US/US]; 5655 Spalding Drive, Norcross, GA 30092  
(US).

(72) Inventors; and

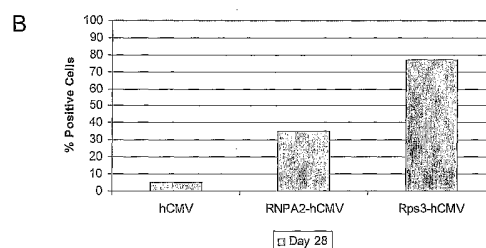
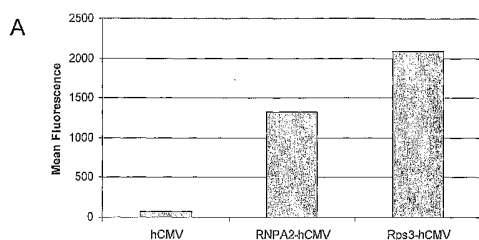
(75) Inventors/Applicants (for US only): **SIMPSON,  
David, John** [GB/GB]; 4 Picasso Drive, Meir  
Park, Stoke-on-Trent, Staffordshire ST3 7GF (GB).  
**WILLIAMS, Steven, Geraint** [GB/GB]; 17 Chidlow  
Close, Hough, Near Crewe, Cheshire CW2 5RE (GB).  
**IRVINE, Alistair, Simpson** [GB/GB]; 4 Maple Close,  
Doveridge, Derbyshire DE6 5LU (GB).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: IMPROVED EXPRESSION ELEMENTS



(57) Abstract: The invention relates to genetic elements capable of improving the levels of expression of operably-linked tran-  
scription units. In particular, said genetic elements are derived from the 5' untranslated regions of ribosomal protein genes and may  
comprise a CpG island. Also provided are vectors and host cells comprising said genetic elements and methods of use to obtain high  
levels of recombinant gene expression.

WO 2006/123097 A2

## Improved expression elements

### BACKGROUND

The present invention relates to polynucleotides comprising elements conferring improved expression on operably-linked transcription units. These elements are naturally associated with the promoter regions of ribosomal protein genes and, in recombinant DNA constructs; confer high and reproducible levels of gene expression. The present invention also relates to vectors comprising such polynucleotide sequences, host cells comprising such vectors and use of such polynucleotides, vectors or host cells in therapy, for production of recombinant proteins in cell culture and other biotechnological applications.

The current model of chromatin structure in higher eukaryotes postulates that genes are organised in "domains" (Dillon, N. & Grosveld, F. Chromatin domains as potential units of eukaryotic gene function. *Curr. Opin. Genet. Dev.* 4, 260-264 (1994); Higgs, D.R. Do LCRs open chromatin domains? *Cell* 95, 299-302 (1998)) Chromatin domains are envisaged to exist in either a condensed, "closed", transcriptionally silent state, or in a de-condensed, "open" and transcriptionally competent configuration. The establishment of an open chromatin structure characterised by increased DNaseI sensitivity, DNA hypomethylation and histone hyperacetylation, is considered a pre-requisite to the commencement of gene expression.

The open and closed nature of chromatin regions is reflected in the behaviour of transgenes that are randomly integrated into the host cell genome. Identical constructs give different patterns of tissue-specific and development stage-specific expression when integrated at different locations in the mouse genome (Palmiter, R.D. & Brinster, R.L. *Ann. Ref. Genet.* 20, 465-499 (1986); Allen, N.D. *et al. Nature* 333, 852-855 (1988); Bonnerot, C., Grimber, G., Briand, P. & Nicolas, J.F. *Proc. Natl. Acad. Sci. USA* 87:6331-6335 (1990)).

The chromatin domain model of gene organisation suggests that genetic control elements that are able to establish and maintain a transcriptionally competent open chromatin structure should be associated with active regions of the genome.

Locus Control Regions (LCRs) are a class of transcriptional regulatory elements with long-range chromatin remodelling capability. LCRs are functionally defined in transgenic

mice by their ability to confer site-of-integration independent, transgene copy number-dependent, physiological levels of expression on a gene linked in *cis*, especially single copy transgenes Fraser, P. & Grosveld, F. *Curr. Opin. Cell Biol.* 10, 361-365 (1998); Li, Q., Harju, S. & Peterson, K.R. *Trends Genet.* 15: 403-408 (1999). Crucially, such expression is tissue-specific. LCRs are able to obstruct the spread of heterochromatin, prevent PEV (Kioussis, D. & Festenstein, R. *Curr. Opin. Genet. Dev.* 7, 614-619 (1997)) and consist of a series of DNase I hypersensitive (HS) sites which can be located either 5' or 3' of the genes that they regulate (Li, Q., Harju, S. & Peterson, K.R. *Trends Genet.* 15: 403-408 (1999)).

The generation of cultured mammalian cell lines producing high levels of a therapeutic protein product is a major developing industry. Chromatin position effects make it a difficult, time consuming and expensive process. The most commonly used approach to the production of such mammalian "cell factories" relies on gene amplification induced by a combination of a drug resistance gene (e.g., DHFR, glutamine synthetase (Kaufman R.J. *Methods Enzymol* 185, 537-566 (1990)). and the maintenance of stringent selective pressure. The use of vectors containing LCRs from highly expressed gene domains, using cells derived from the appropriate tissue, greatly simplifies the procedure, giving a large proportion of clonal cell lines showing stable high levels of expression (Needham M, Gooding C, Hudson K, Antoniou M, Grosveld F and Hollis M. *Nucleic Acids Res* 20, 997-1003 (1992); Needham M, Egerton M, Millest A, Evans S, Popplewell M, Cerillo G, McPheat J, Monk A, Jack A, Johnstone D and Hollis M. *Protein Expr Purif* 6, 124-131 (1995).

However, the tissue-specificity of LCRs, although useful in some circumstances, is also a major limitation for many applications, for instance where no LCR is known for the tissue in which expression is required, or where expression in many, or all, tissues is required.

US 6,689,606 and co-pending patent application WO 00/0539, incorporated by reference herein, describe elements that are responsible, in their natural chromosomal context, for establishing an open chromatin structure across a locus that consists exclusively of ubiquitously expressed, housekeeping genes. These elements are not derived from an LCR and comprise extended methylation-free CpG islands.

In mammalian DNA, the dinucleotide CpG is recognised by a DNA methyltransferase enzyme that methylates cytosine to 5-methylcytosine. However, 5-methylcytosine is unstable and is converted to thymine. As a result, CpG dinucleotides occur far less frequently than one would expect by chance. Some sections of genomic DNA nevertheless do have a frequency of CpG that is closer to that expected, and these sequences are known as "CpG islands". As used herein a "CpG island" is defined as a sequence of DNA, of at least 200bp, that has a GC content of at least 50% and an observed / expected CpG content ratio of at least 0.6 (i.e. a CpG dinucleotide content of at least 60% of that which would be expected by chance) (Gardiner-Green M and Frommer M. *J Mol Biol* 196, 261–282 (1987); Rice P, Longden I and Bleasby A *Trends Genet* 16, 276–277 (2000).

Methylation-free CpG islands are well-known in the art (Bird *et al* (1985) *Cell* 40: 91-99, Tazi and Bird (1990) *Cell* 60: 909-920) and may be defined as CpG islands where a substantial proportion of the cytosine residues are not methylated and which usually extend over the 5' ends of two closely spaced (0.1-3 kb) divergently transcribed genes. These regions of DNA are reported to remain hypomethylated in all tissues throughout development (Wise and Pravtcheva (1999) *Genomics* 60: 258-271). They are often associated with the 5' ends of ubiquitously expressed genes, as well as an estimated 40% of genes showing a tissue-restricted expression profile (Antequera, F. & Bird, A. *Proc. Natl. Acad. Sci. USA* 90, 1195-11999 (1993); Cross, S.H. & Bird, A.P. *Curr. Opin, Genet. Dev.* 5, 309-314 (1995) and are known to be localised regions of active chromatin (Tazi, J. & Bird, A. *Cell* 60, 909-920 (1990).

An 'extended' methylation-free CpG island is a methylation-free CpG island that extends across a region encompassing more than one transcriptional start site and/or extends for more than 300bp and preferably more than 500bp. The borders of the extended methylation-free CpG island are functionally defined through the use of PCR over the region in combination with restriction endonuclease enzymes whose ability to digest (cut) DNA at their recognition sequence is sensitive to the methylation status of any CpG residues that are present. One such enzyme is *HpaII*, which recognises and digests at the site CCGG, which is commonly found within CpG islands, but only if the central CG residues are not methylated. Therefore, PCR conducted with *HpaII*-digested DNA and over a region harbouring *HpaII* sites, does *not* give an amplification product due to *HpaII* digestion if the DNA is unmethylated. The PCR will *only* give an amplified product if the DNA is methylated. Therefore, beyond the methylation-free region *HpaII* will not digest

the DNA a PCR amplified product will be observed thereby defining the boundaries of the "extended methylation-free CpG island".

It has been shown (WO 00/05393) that regions spanning methylation-free CpG islands encompassing dual, divergently transcribed promoters from the human TATA binding protein (*TBP*)/proteasome component-B1 (*PSMB1*) and heterogeneous nuclear ribonucleoprotein A2/B1 (*hnRNPA2*)/heterochromatin protein 1H<sub>sy</sub> (*HP1<sup>Hsy</sup>*) gene loci give reproducible, physiological levels of gene expression and that they are able to prevent a variegated expression pattern and silencing that normally occurs with transgene integration within centromeric heterochromatin.

It is known that methylation-free CpG islands associated with actively transcribing promoters possess the ability to remodel chromatin and are thus thought to be a prime determinant in establishing and maintaining an open domain at housekeeping gene loci (WO 00/05393) and that such elements confer an increased proportion of productive gene delivery events with improvements in the level and stability of transgene expression.

Ribosomes are large RNA and protein complexes responsible for the translation of mRNA into polypeptides. Each ribosome is comprised of 4 ribosomal RNA (rRNA) molecules and large number of ribosomal proteins (currently thought to be 79 in mammalian cells). Ribosomal proteins have functions including facilitation of rRNA folding, protection from cellular ribonucleases, and co-ordinating protein synthesis. Some ribosomal proteins have additional extraribosomal functions (Wool, 1996, TIBS 21: 164–165). Given the structural and functional similarities of ribosomes across species, it is unsurprising that the amino acid sequence conservation of ribosomal proteins is high, and among mammals the sequences of most ribosomal proteins are almost identical (Wool *et al*, 1995, Biochem Cell Biol 73: 933–947).

Two ribosomal proteins appear atypical in that they are expressed in the form of propeptides (carboxy-extension proteins) fused to ubiquitin. Ubiquitin is a highly conserved 76-residue polypeptide involved in a variety of cellular functions, including the regulation of intracellular protein breakdown, cell cycle regulation and stress response (Hershko & Ciechanover, 1992, Annu Rev Biochem 61: 761–807; Coux *et al*, 1996, Annu Rev Biochem 65: 801–847).

Ubiquitin is encoded by two distinct classes of gene. One is a poly-ubiquitin gene encoding a linear polymer of ubiquitin repeats. The other comprises genes encoding natural fusion proteins in which a single ubiquitin molecule is linked to the ribosomal protein rps27A or rpL40 (Finley *et al*, 1989, Nature 338: 394–401; Chan *et al*, 1995, Biochem Biophys Res Commun 215: 682–690; Redman & Burris, 1996, Biochem J 315: 315–321).

The common structural features of ribosomal protein promoters are discussed by Perry (2005, BMC Evolutionary Biology 5: 15). The promoters may be classified according to the nature of the TATA box motifs, number and type of transcription factor binding sites and location of AUG start codons. However, such classification does not appear to predict promoter strength and evidence suggests that several such promoters tested have equivalent transcriptional activity as measured by expression of a linked reporter gene (Hariharan *et al*, 1989, Genes Dev 3: 1789–800).

US patent 6,063,598 discloses the hamster-ubiquitin/S27a promoter its use to drive high level production of recombinant proteins. However, there is no suggestion of its use to enhance the expression of a gene primarily transcribed from a further promoter (i.e one other than hamster-ubiquitin/S27a promoter).

US application US 2004/0148647 discloses a reporter assay using an expression vector comprising a hamster ubiquitin /S27A promoter functionally linked to a gene for a product of interest and a fluorescent protein reporter. Again, the application only discloses constructs in which transcription of gene of interest is from the hamster-ubiquitin/ S27a promoter itself.

It remains an objective in the field of recombinant gene expression to obtain higher and more reliable levels of expression, particularly for *in vivo* and *ex vivo* therapeutic applications and for *in vitro* recombinant protein production.

## **SUMMARY OF THE DISCLOSURE**

Throughout the description and claims of this specification, the words “comprise” and “contain” and variations of the words, for example “comprising” and “comprises”, means

"including but not limited to", and is not intended to (and does not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

### **Definitions**

As used herein, a promoter region is defined as being a genomic nucleotide sequence consisting of a promoter and transcriptional start site together with 5kb of 5' sequence upstream of the transcriptional start site and 500bp 3' sequence downstream of the distal end of the first exon.

A 5' untranslated region means a region 5' of the translational start encoded in the genomic or cDNA sequence. It is taken to include all upstream regulatory elements. A 5' upstream sequence is used to mean sequence 5' to the transcriptional start encoded in the genomic sequence.

As used herein, 'transcribable nucleic acid' means a nucleic acid which, when operably linked to a functional promoter and other regulatory sequences, is capable of being transcribed to give a functional RNA molecule, such as mRNA. Such sequences may comprise open reading frames, which encode translatable polypeptide sequences. Alternatively, the functional RNA may have another function, such as ribosomal RNA, ribozymes or antisense RNA.

'Gene' is commonly taken to mean the combination of the coding region of the transcribable nucleic acid, together with the promoter from which it is transcribed and other regulatory sequences such as enhancers and 3' polyadenylation signals. In genomic DNA genes also contain introns. 'Transcription unit' is sometimes used to describe a functional combination of at least a promoter and minimal regulatory

sequences with a transcribable nucleic acid, often derived from cDNA from which the introns have been spliced out. 'Cistron' is defined as a nucleic acid encoding a single polypeptide together with functional initiation and termination signals. 'Transgene' implies a gene that has been transferred from one genome to another, although the term may be more loosely applied to any gene or even transcribable nucleic acid comprised in a recombinant DNA construct such as a vector.

Promoter and enhancer are terms well known in the art and include the following features which are provided by example only, and not by way of limitation. Promoters are 5', *cis*-acting regulatory sequences directly linked to the initiation of transcription. Promoter elements include so-called TATA box and RNA polymerase initiation selection (RIS) sequences which function to select a site of transcription initiation. These sequences also bind polypeptides which function, *inter alia*, to facilitate transcription initiation selection by RNA polymerase.

In simple terms, promoters are directional elements that act to initiate transcription of sequences situated less than 100 (and usually less than 50) nucleotide base pairs (bp) downstream. They contain a number of short consensus nucleotide sequences that act as binding sites for various proteins that participate in the initiation of transcription and the assembly of a multi-subunit complex known as the pre-initiation complex (McKnight and Tjian, 1987, Cell 46: 795–805). In most genes, this occurs at a very widely conserved sequence known as the TATA box (TATAAA) to which the TATA box-binding protein (TBP, a subunit of the general transcription factor TFIID) binds. There follows an ordered assembly of more than ten further transcription factors to finally form the Pol II holoenzyme complex. RNA transcription actually starts at an initiator site about 25–30 bases downstream (Breathnach and Chambon, 1981, Annu Rev Biochem 50: 349–393) to which TBP also binds.

Most functional promoters contain further upstream promoter elements (UPEs), of which the most highly conserved are the CAAT box (CCAAT, the binding site for the transcription factors CBF, C/EBP and NF-1), about 70–200bp upstream, and the GC box (GGGCGG, binding site for the general transcription factor Sp-1) a similar distance upstream. Although basal levels of transcription occur from the TATA box alone, for most promoters at least the CAAT and GC boxes are required for optimal levels of transcription.

Enhancers are sequences that act non-directionally to increase transcription from promoters situated locally but not necessarily immediately adjacent (up to several kilobases away (Kadonaga (2004) *Cell* 116: 247–257). Enhancers contain short (8–12bp) consensus sequences representing the binding sites for a wide range of transcriptional activator proteins (Ondek *et al*, 1988, *Science* 236: 1237–1244) including some, such as NF-1 and SP-1 that are also associated with promoter elements. These sequences are often duplicated in tandem or inverted repeats.

In some natural transcription units, including the very active immediate/early gene transcription units of many DNA viruses such as cytomegalovirus, enhancer and promoter elements may be functionally combined into what is effectively one extended upstream element.

Promoters may be regulated, being responsive to cell type, temperature, metal ions or other factors; or constitutive, giving transcription that is unresponsive to such factors. For many purposes a strong, constitutive promoter giving consistent, high, levels of transcription in many, if not all, cell types is highly advantageous. For many years the enhancer/promoter element driving immediate/early gene expression in human cytomegalovirus has been very widely used for driving such expression of heterologous genes in eukaryotic expression vectors (Foecking & Hoffstetter, 1986, *Gene* 45: 101–105).

It was hypothesised that promoter regions of ribosomal protein genes might have useful activity in boosting and stabilising expression of linked transgenes and that the regulatory regions from highly expressed genes might be more likely to contain elements that are very effective at maintaining chromatin in a transcriptionally active conformation. The linking of such elements to a heterologous promoter might then generate a more open chromatin environment surrounding that promoter, resulting in increased expression. It will be understood by one of skill in the art that promoters derived from ribosomal protein genes are to be distinguished from ribosomal promoters, which are RNA polymerase Type I dependent promoters from which rRNA is transcribed.

To test the hypothesis, RNA was obtained from exponentially growing CHO-K1 and NS0 cell lines and a microarray analysis against 13,443 murine genes was performed. We limited our analysis to elements having a high CpG island content and a likelihood of bi-directional promoters. Using criteria based on the minimal effective sequence from the

hnRNPA2 regulatory region, approximately 3kb of DNA from a selection of these genes was amplified by PCR from NS0 genomic DNA. These sequences were then cloned into EGFP expression vectors and transfected into CHO-K1 along with hnRNPA2 control versions of the same vector.

It was found that sequences derived from the promoter regions of two ribosomal proteins gave consistently high levels of expression of the heterologous reporter sequence in the assay used. In each case, the promoter region comprised a GC-rich sequence extending from the 5' region upstream of the actual promoter elements well into the first exon and, indeed into the first intron. This GC-rich sequence fulfilled the criteria of being an extended CpG island, as defined herein as extending over 300bp.

Accordingly, the invention provides an isolated polynucleotide comprising

- a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
- b) a heterologous promoter
- c) a transcribable nucleic acid sequence adjacent said heterologous promoter

wherein the transcribable nucleic acid sequence is transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.

an element. Preferably said element comprises more than 1kb and most preferably more than 3kb 5' untranslated sequence from a ribosomal protein gene.

The contiguous nucleotides are selected from the promoter area, which extends from a point 5kb upstream (5' relative to the sense strand) of the transcriptional start site to a point 500bp downstream (3' relative to the sense strand) of the distal (3') end of the first exon.

Preferably said ribosomal protein gene is selected from the list consisting of RPSA, RPS2, RPS3, RPS3A, RPS4, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27A, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL27, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37A, RPL38, RPL39,

RPL41, RPLP0, RPLP1 and RPLP2 and their orthologues. More preferably it is RPS3 or RPS11.

In one preferred embodiment the element comprises a CpG island, preferably an extended CpG island of at least 300bp and more preferably 500bp. Preferably the CpG island is unmethylated. It is also preferred that said element comprises the promoter from a ribosomal protein gene and from which transcription of the ribosomal protein gene is naturally initiated. Such a promoter is often referred to as the endogenous promoter. In a preferred embodiment, the element further comprises or more exons of said ribosomal protein gene.

It is preferred that the ribosomal protein gene is a mammalian gene, although such genes and their promoters and 5' upstream sequences are highly conserved across species and might alternatively be an insect, nematode or yeast gene. Preferably, however, it is a human or rodent gene and most preferably it is a mouse gene.

In a highly preferred embodiment, the isolated polynucleotide of the invention comprises nucleotides 38 to 3154 of the mouse rps3 nucleotide sequence depicted in SEQ ID NO:1. Alternatively it comprises nucleotides 12 to 3032 of the mouse rps11 nucleotide sequence as listed in SEQ ID NO:2.

In one aspect, in addition to the element described, the polynucleotide further comprises a promoter not naturally associated with said element from a ribosomal protein gene. In this embodiment a heterologous promoter (distinct from the endogenous promoter which may or may not be present in the first element) is situated in an adjacent and operably-linked position downstream from the element containing ribosomal protein gene-derived 5' sequence. In this arrangement, expression directed by this heterologous promoter is enhanced by the effect of the ribosomal protein gene element.

In one embodiment, said promoter is a constitutive promoter, more preferably selected from the list consisting of the cytomegalovirus early/immediate promoter, SV40, EF-1 $\alpha$ , Rous sarcoma virus (RSV) LTR, or HIV2 LTR or combinations of sequences derived therefrom. More preferably the promoter is a CMV immediate/early promoter. Most preferably it is the mouse or guinea pig CMV immediate/early promoter.

Alternatively, said promoter may be a tissue-specific promoter, which directs expression in a limited range of tissues. Such promoters are well-known in the art and include those from  $\beta$ -globin, the  $\kappa$  and  $\lambda$  immunoglobulin light chains, immunoglobulin heavy chain, desmin, tyrosinase, CD2, IL-3, myosin light chain, human melanoma inhibitory activity gene promoter and keratins. In a particularly preferred embodiment the promoter is a tumour-selective promoter, which directs expression preferentially one or more tumour types. Examples of such promoters include those from carcino-embryonic antigen (CEA), prostate-specific antigen (PSA), cyclooxygenase-2 (COX-2), alpha-fetoprotein (AFP), tyrosinase, and T-cell Factors 1–4 (TCF) -based promoters.

The transcribable nucleic acid may encode any useful polypeptide for *in vitro* expression and is preferably selected from the list consisting of an antibody, a functional epitope-binding fragment of an antibody, growth factor, cytokine, protein kinase, soluble receptor, membrane-bound receptor, or blood clotting factor. Alternatively, the transcribable nucleic acid may encode a therapeutic gene of use for *in vivo* or *ex vivo* gene therapy. Such a therapeutic nucleic acid may act by replacing or supplementing the function of a defective gene causing a disease such as cystic fibrosis, thalassaemia, sickle anaemia, Fanconi's anaemia, haemophilia, severe combined immunodeficiency (SCID), phenylketonuria (PKU), alpha-1 antitrypsin deficiency, Duchenne muscular dystrophy, ornithine transcarbamylase deficiency or osteogenesis imperfecta. Alternatively, it may encode a cytotoxic agent or prodrug-converting enzyme selectively expressed in a target cell, such as a malignant cancer cell, in order to kill it. Such applications, and many others, are well-known to those of skill in the art and the relevance of the current invention in enhancing the expression of therapeutic nucleic acids will be clear to such skilled practitioners.

In another aspect, the invention provides a vector comprising the polynucleotide of invention as disclosed above. Preferably said vector is an expression vector adapted for eukaryotic gene expression.

Typically said adaptation includes, by example and not by way of limitation, the provision of transcription control sequences (promoter sequences) which mediate cell/tissue specific expression. Adaptations also include the provision of selectable markers and autonomous replication sequences which both facilitate the maintenance of said vector in either the eukaryotic cell or prokaryotic host. Vectors which are maintained autonomously are referred to as episomal vectors. Episomal vectors are desirable since

they are self-replicating and so persist without the need for integration. Episomal vectors of this type are described in WO98/07876.

Adaptations which facilitate the expression of vector encoded genes include the provision of transcription termination/polyadenylation sequences. This also includes the provision of internal ribosome entry sites (IRES) which function to maximise expression of vector encoded genes arranged in bicistronic or multi-cistronic expression cassettes.

These adaptations are well-known in the art. There is a significant amount of published literature with respect to expression vector construction and recombinant DNA techniques in general. Please see, Sambrook et al (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY and references therein; Marston, F (1987) *DNA Cloning Techniques: A Practical Approach Vol III* IRL Press, Oxford UK; *DNA Cloning: F M Ausubel et al, Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.(1994).

The vector may be an episomal vector or an integrating vector. Preferably, the vector is a plasmid. Alternatively, the vector may be a virus, such as an adenovirus, adeno-associated virus, a herpesvirus, vaccinia virus, lentivirus or other retrovirus.

Alternatively, such a vector may comprise

- a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
- b) a heterologous promoter
- c) a multiple cloning site

wherein a transcribable nucleic acid sequence inserted into said multiple cloning site is capable of being transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.

In a further aspect the invention provides a host cell comprising the isolated polynucleotide or vector as herein described. Preferably said host cell is a mammalian cell, more preferably selected from the list consisting of CHO, NS0, BHK, HeLa, HepG2.

Also provided by the invention is a method of expressing a polypeptide comprising inserting expression vector comprising the polynucleotide of the invention into an appropriate host cell as herein described and culturing said host cell in suitable

conditions to allow expression. Preferably said polypeptide is a therapeutically useful polypeptide.

In a further aspect the invention provides a pharmaceutical preparation comprising the polynucleotide, vector or host cell as herein described and a pharmaceutically acceptable carrier, excipient, buffer or medium.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** shows a plasmid map of vector rps3- 1005-EGFP (see Example 1)

**Figure 2** shows a plasmid map of vector rps11-1005-EGFP (see Example 2)

**Figure 3** shows the expression of the EGFP reporter as expressed by various rps3 constructs in CHO-K1 cells analysed by FACS analysis, 8 days post transfection. Figure A shows mean fluorescence, and Figure B indicates the percentage of cells expressing the reporter gene to a detectable level (% positive cells). See Example 1.

**Figure 4** shows reporter gene expression of rps11 constructs in CHO-K1 cells 7 days post-transfection (analysed by FACS). A and C are total counts, B to E are results based on just the expressing cells within the population. Figure A shows the mean fluorescence of cells in the stably selected pool, and Figure B shows percentage (%) positive cells. See Example 2.

**Figure 5** shows the expression level of a reporter gene in stably transfected NS0 cells when it is driven by the hCMV promoter with no additional elements present or when either an 8kb hnRNPA2 or a 3kb RPS3 element are placed immediately 5' to the hCMV promoter. Figure A shows the mean fluorescence intensity of the stable pools at 28 days and Figure B shows the percentage (%) positive cells.

**Figure 6** shows similar data to Figure 5 for rps11 constructs. Figure A shows the mean fluorescence intensity values for stable pools generated with either hCMV driven constructs or identical constructs with a 8kb hnRNPA2 or 3kb RPS3 elements immediately 5' to the promoter. Figure B shows the percentage of cells within the pool expressing the reporter gene (percentage (%) positive cells).

## DETAILED DESCRIPTION

## Materials and Methods

### **Microarray analysis.**

Total RNA was extracted from ~80% confluent CHO-K1 cells using RNeasy RNA extraction kit (Qiagen, Crawley, UK) according to the manufacturers protocol. Total RNA (2µg/µl) was subjected to microarray expression analysis using the mouse 70-mer oligonucleotide library (Operon V.1) representing 13,443 known transcripts. The University of Cincinnati, Genomics and Microarray Laboratory undertook Microarray analysis according to referenced protocols (<http://microarray.uc.edu>).

Gene transcript sequences were ranked according to increasing fluorescence. Since our previous study detailed the *HNRPA2B1/CBX3* loci as a chromatin-remodelling element and conferring benefit to *hCMV* the *HNRPA2* transcript was identified as the baseline expression level. However, using the available microarray analysis the *HNRPA2* transcript was barely detectable. Since the expression level of the *HNRPA2* transcript was minimal, using *HNRPA2* as our reference would have identified 3829 sequences for potential analysis. Therefore, 7 sequences from the top 2% (76 sequences) of the ordered, expressed transcripts were identified according to the criteria of containing a CpG island and one or more putative/known transcriptional start sites (see Table 1). CpG islands position, size and GC: CG ratios were verified using GrailEXP (<http://compbio.ornl.gov>). Putative/known transcriptional starts sites were identified from NIX blast analysis (<http://www.hgmp.mrc.ac.uk>) and Ensembl databases (<http://www.ensembl.org>).

### **PCR amplification of CpG island-containing fragments.**

PCR oligonucleotides were designed to amplify approximately 3kb fragments encompassing the complete CpG island embedded promoter region whilst including approximately 500bp of coding sequence according to known or predicted coding sequence structure (see Table 2).

PCR reactions contained oligonucleotide sets specific each genomic fragment (2pmol of each primer; Table 2). PCR amplification was achieved using the Failsafe™

PCR premixes A-F (Cambio, UK), 1unit *Taq* DNA polymerase (Promega, UK) and 200ng of template DNA. Initial denaturation was 96°C for 2 min, whilst PCR amplification

was carried out for 35 cycles (94°C for 1min, 55-60°C for 1 min, 72°C for 5min). A final extension step (72°C for 10min) was included.

PCR products were gel purified, using GFX DNA purification columns (Amersham, UK) according to the manufacturers protocol, and subjected to TOPO TA cloning® according to the manufacturers protocol (TOPO; Invitrogen, UK). Sense and anti-sense orientations were obtained for each CpG island-containing fragment cloned into TOPO vectors (Invitrogen, UK).

#### **Expression vector construction.**

A control expression vector (designated CET1005EGFP, SEQ ID NO:20) was constructed by the insertion of an *hCMV/EGFP/sv40* pA (NheI/AgeI deleted multiple cloning site) from pEGFP-N1 into CET 900 followed by the insertion of the AscI cassette from this vector into the AscI site of CET 1005.

All CpG island fragments were removed from TOPO2.1 (Invitrogen, UK) unless otherwise stated. *Terf2ip* Acc65I/EcoRV fragment was inserted into Acc65I/SwaI of 1005. *GAPDH* SpeI/SnaBI was inserted into PmeI/XbaI of 1005. *RPS3* XbaI/SpeI fragment was inserted into XbaI of 1005. *RPS11* and *TUBA1* EcoRI blunt fragments was removed from TOPO4.0 and TOPO2.1 respectively (Invitrogen, UK) and inserted into PmeI of 1005. Finally, *A430106P18Rik* (EcoRV) and *2510006D16Rik* (BstXI) fragments were also inserted into PmeI of 1005. All CpG island containing fragments were inserted in both sense and anti-sense orientations immediately upstream of the *hCMV* promoter.

#### **Cell lines and transfections.**

CHO-K1 cells were grown in HAMS F12 (Invitrogen, Paisley, UK) plus 4500 mg/l L-ananyl-L-glutamine, 10 µg/ml each of penicillin and streptomycin, and 10% (v/v) heat inactivated foetal calf serum (FCS; Invitrogen, Paisley, UK). Transfection was carried out by electroporation using approximately 10<sup>7</sup> cells from 80% confluent cultures and a BioRad Gene Pulser II™ set to deliver a single pulse of 975µF at 250V. Transfections used 2µg of linearised CET1005EGFP plasmid and equivalent molar quantities for expression vectors of different size. Stably transfected cells were selected and maintained in growth medium containing 12.5µg/ml puromycin sulphate (Sigma, UK).

#### **Quantification of transgene expression**

Analysis of cells transfected with *EGFP* reporter constructs was with a Becton-Dickinson

FACScan using the parental CHO-K1 cell line as a background, autofluorescence control.

Table 1. Sequences analysed

Locus	Acc. # <sup>a</sup>	Description <sup>b</sup>	CpG island <sup>c</sup>		
			bp	% CG	GC/CG
<i>Terf2ip</i>	AB041557	Telomeric repeat binding factor 2 interacting protein 1 ( <i>TRF2</i> - interacting telomeric protein Rap1).	968	64.56	0.90
<i>Gapd</i> <sup>d</sup>	M32599	Glyceraldehyde-3-phosphate dehydrogenase	1187	60.50	0.84
<i>RPS3</i>	NM012052	<i>RPS3</i> - 40S ribosomal protein S3.	419	60.39	0.87
<i>TUBA1</i>	M13445	Tubulin alpha-1 chain (Alpha-tubulin 1).	850	66.30	0.91
<i>RPS11</i>	AK011207	<i>RPS11</i> - 40S ribosomal protein S11.	957	59.71	0.95
<i>A430106P18Rik</i>	AK020778	Expressed sequence tag	982	63.62	0.93
<i>2510006D16Rik</i>	AK010915	Expressed sequence tag	679	67.69	0.75

<sup>a</sup>Genbank Accession

<sup>b</sup>Ensembl description (<http://www.ensembl.org/>)

<sup>c</sup>Grailexp (<http://compbio.ornl.gov/grailexp>)

<sup>d</sup>Gapd – derived from human sequence

Table 2. PCR oligonucleotides and amplicon sizes

Locus	Sense	Antisense	Amplicon
<i>Terf2ip</i>	gtagttctgacttgaaatgt (SEQ ID NO: 3)	aactgacctgccatgccattc (SEQ ID NO: 4)	2995 bp
<i>Gapd</i>	gagcagtcgggtgtcacta (SEQ ID NO: 5)	gcagagaagcagacagttatg (SEQ ID NO: 6)	3096 bp
<i>RPS3</i>	cagagcatcaagtacctgtga (SEQ ID NO: 7)	taaccactaagccatctctcc (SEQ ID NO: 8)	3056 bp
<i>TUBA1</i>	caagaacaaggaagctggcc (SEQ ID NO: 9)	taaaacccacagcactgtaggg (SEQ ID NO: 10)	3049 bp
<i>RPS11</i>	aagactgfttgccatgcc (SEQ ID NO: 11)	ggatgacaatggcctctgc (SEQ ID NO: 12)	3020 bp
<i>A430106P18Rik</i>	atggttaggttcacgtcc (SEQ ID NO: 13)	atccctcacattgccaagcc (SEQ ID NO: 14)	3128 bp
<i>2510006D16Rik</i>	acttaagacctgatgcctcc (SEQ ID NO: 15)	gctagcttacataggcagcc (SEQ ID NO: 16)	2997 bp

**Example 1: rps3 element driven expression**

SEQ ID NO:1 shows the RPS3 cloned sequence (Nucleotides 38 to 3154) ; SEQ ID NO:17 shows the complete plasmid sequence of pRPS3-1005-EGFP; SEQ ID NO:18 shows the complete plasmid sequence of pCET1015-EGFP.

EGFP expression levels, 8 days post-transfection, were investigated, within CHO-K1 pools containing hCMV alone (control construct; plasmid pCET1005-EGFP, linearised with *PmeI* prior to transfection), constructs containing an 8kb RNPA2 fragment (plasmid pCET1015-EGFP, linearised with *PmeI* prior to transfection) and Rps3 (plasmid pRPS3-1005-EGFP; linearised with *PmeI* prior to transfection).

Pools generated with Rps3 containing constructs show a significant increase in EGFP expression levels compared to control constructs. Addition of the Rps3 sequence upstream of the hCMV promoter resulted in a 5.5- or 1.5-fold increase in mean fluorescence intensity relative to the control or hnRNPA2 element containing constructs respectively (Figure 3A).

The activity of the constructs was investigated in NS0 cells. The increase in mean fluorescence intensity in stable pools when RPS3 element or hnRNPA2 elements are

included in the constructs, compared to the hCMV promoter alone, was 28-fold or 18-fold respectively (Figure 5A).

In both CHO-K1 and NS0 cells, the percentage positive cells was significantly increased with the hnRNPA2 element but this increase was greater with the RPS3 element (Figures 3B and 5B)

#### **Example 2: rps11 element-driven expression**

SEQ ID NO:2 shows the RPS11 cloned sequence (nucleotides 12 to 3032); SEQ ID NO:19 shows the complete sequence of pRPS11-1005-EGFP.

Rps11 containing- and control vectors (*PmeI* linearised) were transfected into CHO-K1 and NS0 cell lines and stable pools were generated by puromycin selection. Mean EGFP expression levels were assessed by FACscan analysis.

The addition of the Rps11 element upstream of the hCMV resulted in a 1.2-fold increase in mean EGFP expression levels, in CHO-K1 pools, compared to a construct containing the previously described RNPA2 fragment (Fig 4A).

NS0 cell lines stably transfected with Rps11 containing constructs demonstrated a 1.8 and 1.5 fold, (respectively) increase in mean EGFP expression levels compared to hCMV and RNPA2 constructs (Fig 6A).

An increase in percentage positive cells was observed for CHO-K1 cell lines transfected with Rps11 constructs compared to RNPA2 constructs (Fig 4B). Furthermore, an increase in percentage positive cells was observed in NS0 pools transfected with Rps11 constructs compared to both hCMV and RNPA2 (Fig 6B)

## Claims

1. An isolated polynucleotide comprising
  - a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
  - b) a heterologous promoter
  - c) a transcribable nucleic acid sequence adjacent said heterologous promoterwherein the transcribable nucleic acid sequence is transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.
2. A polynucleotide according to claim 1, wherein said ribosomal protein gene is selected from the list consisting of RPSA, RPS2, RPS3, RPS3A, RPS4, RPS5, RPS6, RPS7, RPS8, RPS9, RPS10, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS28, RPS29, RPS30, RPL3, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL23, RPL23A, RPL24, RPL26, RPL27, RPL27A, RPL28, RPL29, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL37, RPL37A, RPL38, RPL39, RPL41, RPLP0, RPLP1, RPLP2.
3. A polynucleotide according to claim 1 or 2, wherein said ribosomal protein gene is RPS3 or RPS11.
4. A polynucleotide according to claims 1 to 3, wherein said element comprises at least 1000 contiguous nucleotides from the promoter region from a ribosomal protein gene.
5. A polynucleotide according to any preceding claim, wherein said element comprises one or more exons of said ribosomal protein gene.
6. A polynucleotide according to any preceding claim, wherein said element comprises a CpG island.
7. A polynucleotide according to claim 5, wherein said element comprises an extended CpG island.

8. A polynucleotide according to claim 6, wherein said element comprises an extended, methylation-free CpG island.
9. A polynucleotide according to any preceding claim, wherein the ribosomal protein gene is a mammalian gene.
10. A polynucleotide according to claim 9, wherein the ribosomal protein gene is a murine gene.
11. A polynucleotide according to claim 10 comprising the nucleotide sequence of SEQ ID NO:1.
12. A polynucleotide according to claim 10 comprising the nucleotide sequence of SEQ ID NO:2.
13. A polynucleotide according to claim 9, wherein the ribosomal protein gene is a human gene.
14. A polynucleotide according to any preceding claim, wherein said heterologous promoter is a constitutive promoter.
15. A polynucleotide according to claim 14, wherein said constitutive promoter is selected from the list consisting of cytomegalovirus early/immediate promoter, SV40, EF-1 $\alpha$ , Rous sarcoma virus (RSV) LTR and HIV2 LTR
16. A polynucleotide according to any of claims 1 to 13, wherein said heterologous promoter is a tissue-specific promoter.
17. A polynucleotide according to claim 16, wherein said heterologous promoter is a tumour-selective promoter.
18. A polynucleotide according to claim 17, wherein said promoter is selected from the list consisting of carcino-embryonic antigen (CEA), prostate-specific antigen (PSA), cyclooxygenase-2 (COX-2), alpha-fetoprotein (AFP), tyrosinase, and T-cell Factors 1–4 (TCF) -based promoters.

19. A polynucleotide according to any preceding claim, wherein said transcribable nucleic acid encodes a polypeptide selected from the list consisting of an antibody, a functional epitope-binding fragment of an antibody, growth factor, cytokine, protein kinase, soluble receptor, membrane-bound receptor, blood clotting factor [etc...]
20. A vector comprising the polynucleotide of any preceding claim.
21. A eukaryotic expression vector according to claim 20.
22. A eukaryotic expression vector comprising
- a) an element comprising at least 500 contiguous nucleotides from the promoter region of a ribosomal protein gene
  - b) a heterologous promoter
  - c) a multiple cloning site
- wherein a transcribable nucleic acid sequence inserted into said multiple cloning site is capable of being transcribed from said heterologous promoter and the level of said transcription is enhanced by said element.
23. A host cell comprising an isolated polynucleotide according to any of claims 1 to 19, or the vector of any of claims 20 to 22.
24. A host cell according to claim 23, wherein said cell is selected from the list consisting of CHO, NS0, BHK, HeLa, HepG2.
25. A method of expressing a polypeptide comprising inserting expression vector according to either claim 21 or claim 22 into an appropriate host cell and culturing said host cell in suitable conditions to allow expression.
26. Method according to claim 25, wherein said polypeptide is a therapeutically useful polypeptide.
27. A pharmaceutical preparation comprising a polynucleotide according to any of claims 1 to 19, a vector according to any of claims 20 to 22, or a host cell according to either claim 23 or claim 24, and a pharmaceutically acceptable carrier, excipient, buffer or medium.

Figure 1

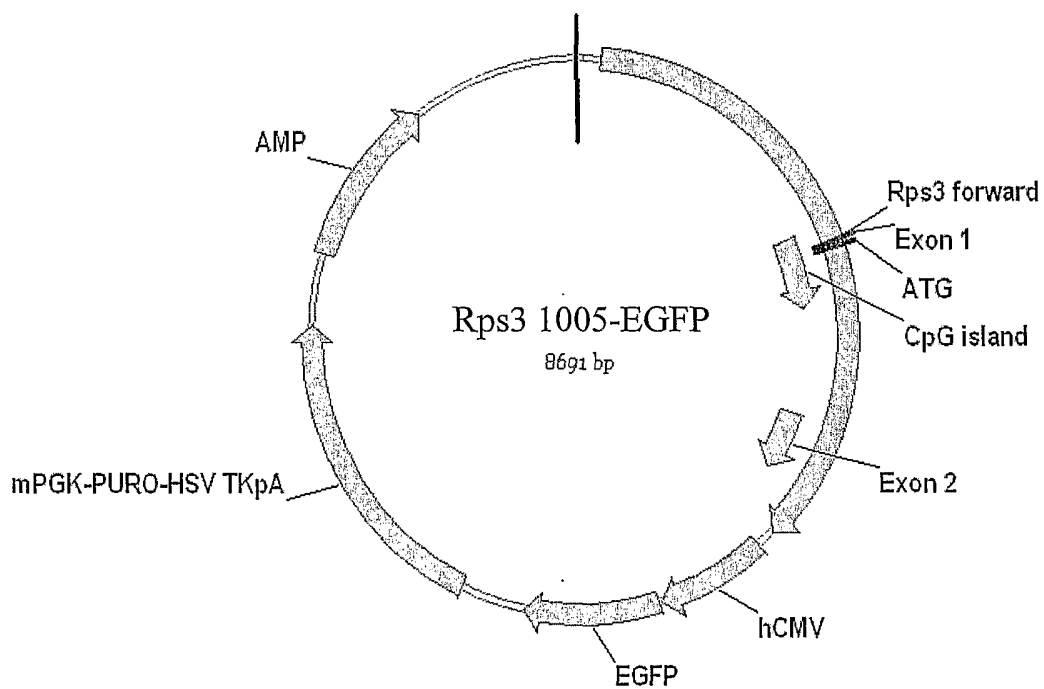


Figure 2

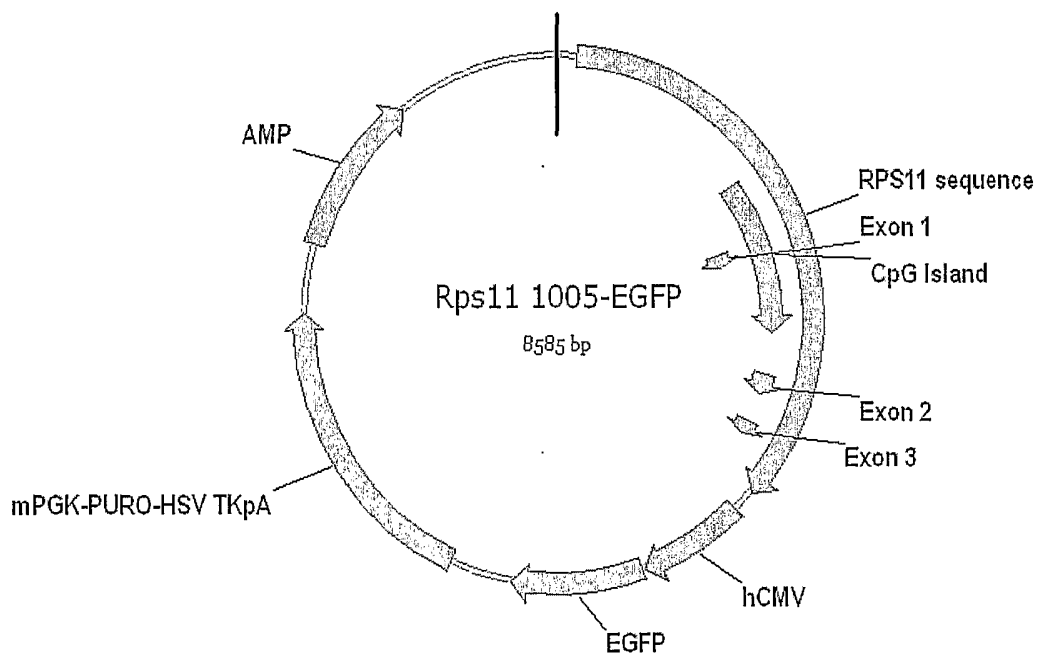
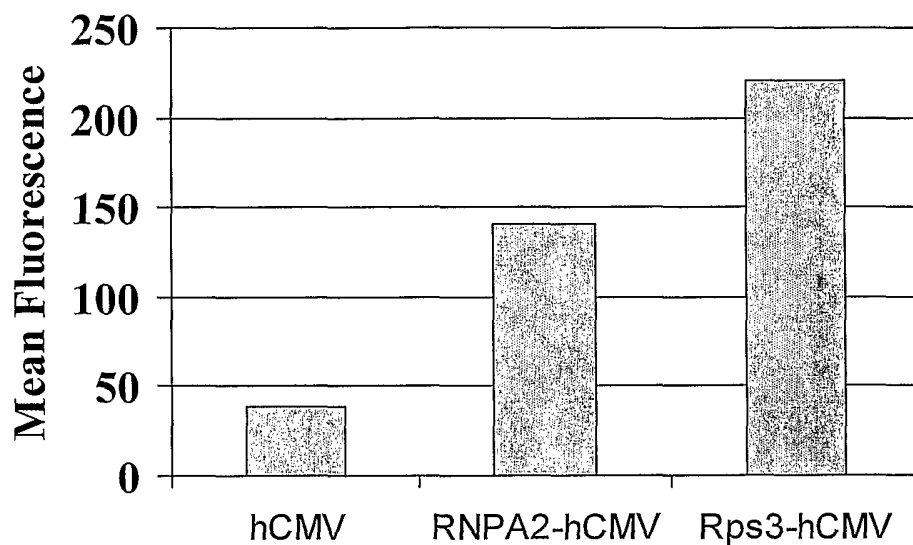
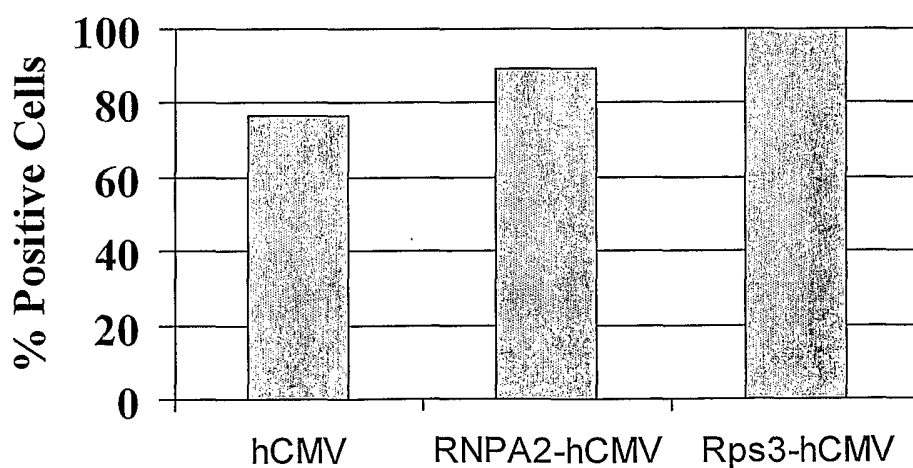


Figure 3.

A



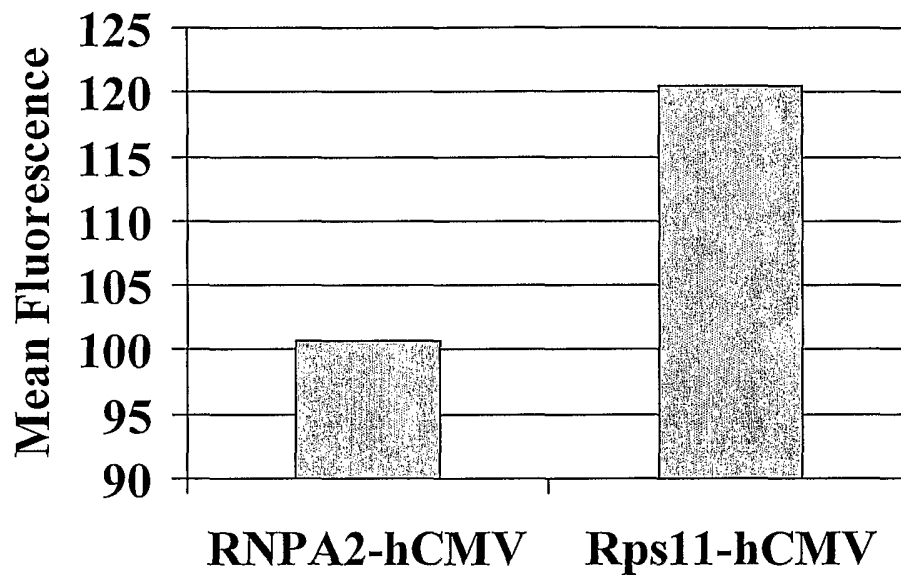
B



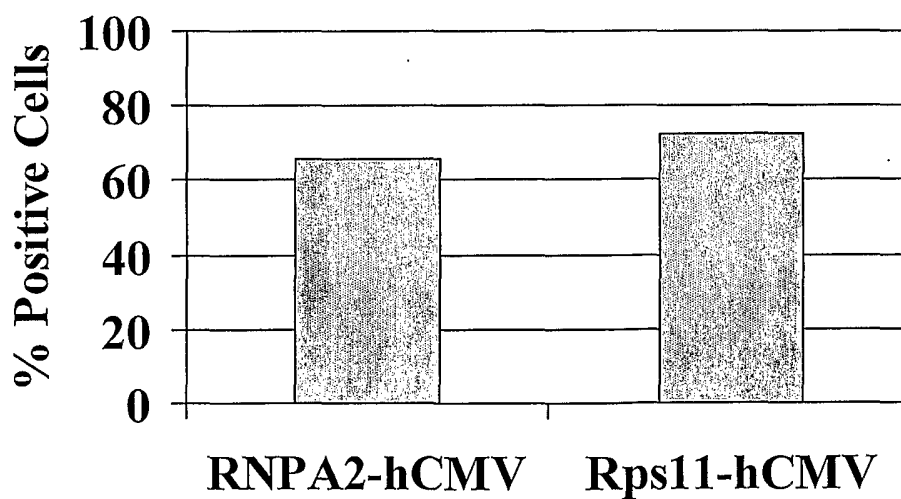
Day 8

Figure 4.

A



B



Day 7

Figure 5.

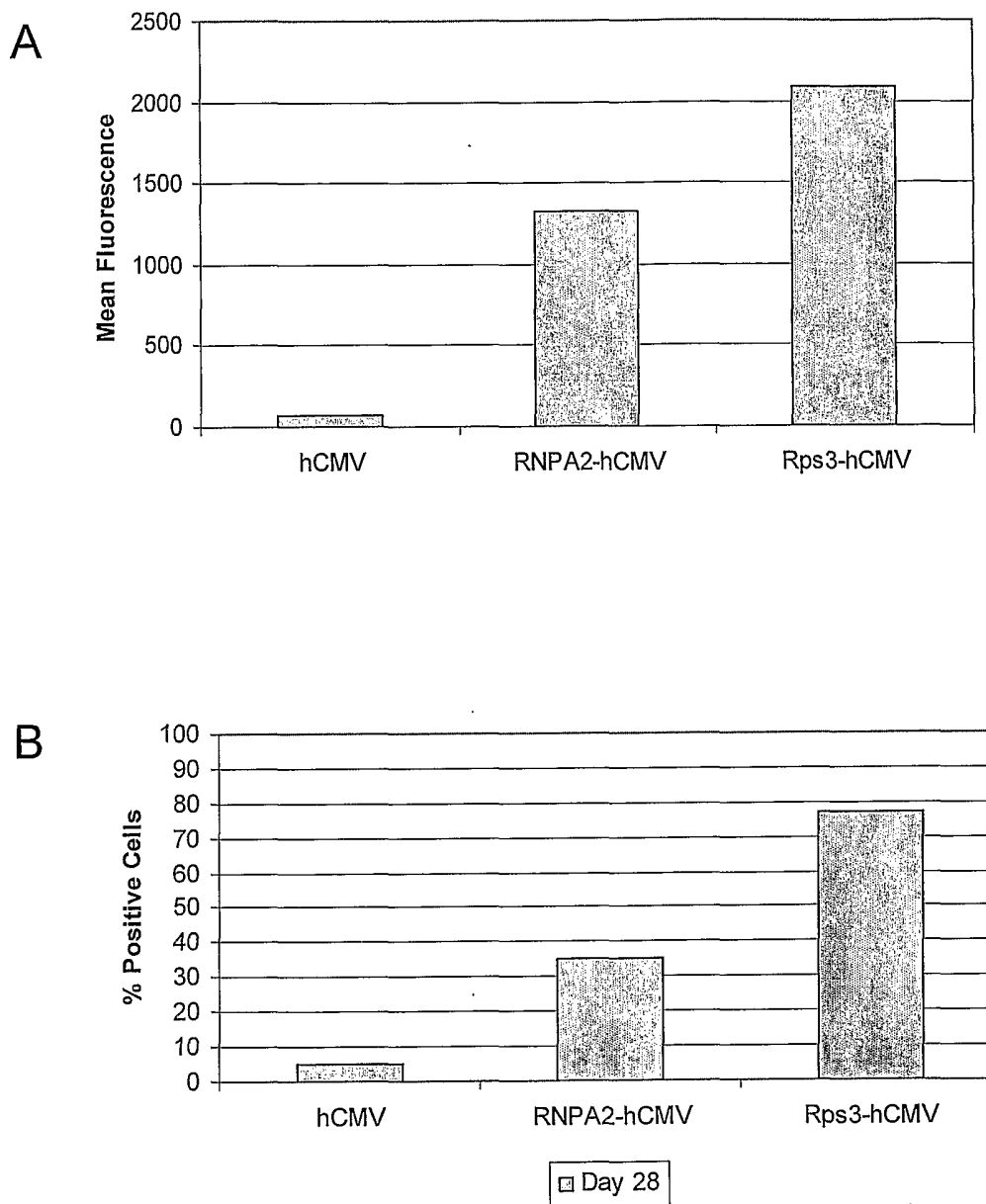
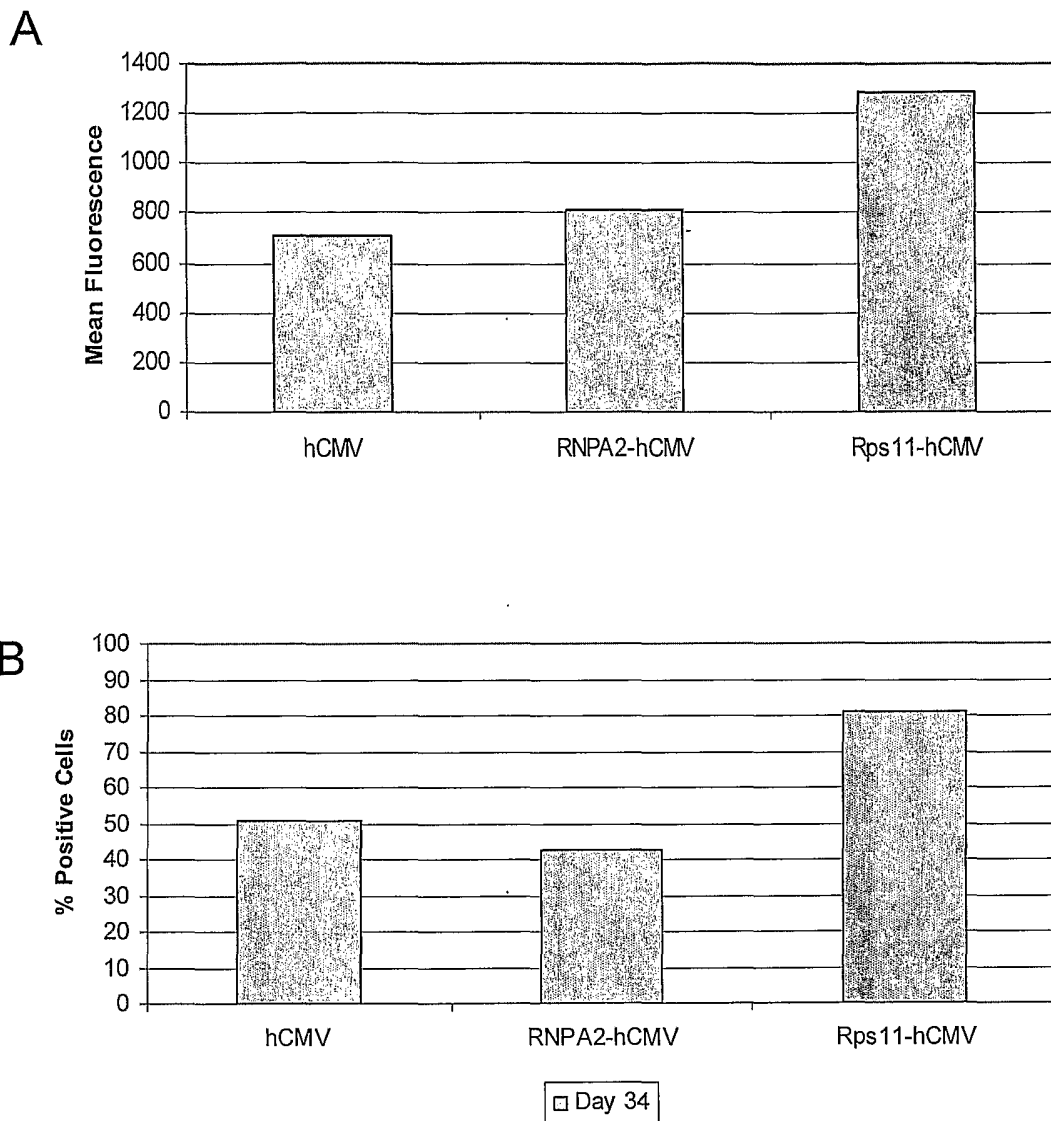


Figure 6.



P108872WO rps UCOEs.ST25  
SEQUENCE LISTING

<110> Serologicals Investment Company  
 <120> Improved expression elements  
 <130> P108872WO  
 <160> 20  
 <170> PatentIn version 3.2  
 <210> 1  
 <211> 3145  
 <212> DNA  
 <213> Mus musculus rps3  
 <400> 1  
 ctagtaacgg ccgccagtgt gctggaattc gcccttataa cactgagcc atctctccag 60  
 ccctgagtca tgattttagt gtgagaggca tcattgaatt ttctgagcac ggccatcagg 120  
 gtagctggca caggtcttca gatacaagga gatagttata agaaggcagc catggctgtg 180  
 gtgcactaga aatggagaaa cagcttcatc aggtgacaga ccagtctgac tctgtcccat 240  
 gattagaagc catcttgta caaggtcaaa ataagttcat tcctgttttc tgtaacactt 300  
 gggtttgatc ctgtcgtcaa cccattttct ggaatttgac atgttccata ctccattata 360  
 ccctgacttc caccctgata agatgttctg ccaagttcct gtgtagccaa cattcccctg 420  
 gaaatctctc ttcccttgga aaccacctag tcttagaaat ttgagttat ataaattcca 480  
 cttctatggt tgatgctatt ctttaaaact ccactttagg gagatagccc tgtctgatag 540  
 aaaataaaac ttgcttaatt tgtctaaaag agtttaagta atagttttta cttttgttcc 600  
 gtgggattaa tacaggggtga aacagactcc cgtgtttcca gtgtgaagtg agccacacac 660  
 tgcagtacaa gttatatcag caggttctgc ctctgggcaa tgaacttttg cttgtgtgga 720  
 catcagggtc tgtgtgaagg gaaggtccta tggcctagat ttatactatt caacagtctg 780  
 tccccgaagc cctgggtgctt tattattttg acaagcccct gctgctggta ttccaccctg 840  
 ctgcbagtc aaaaagttcc tgtctcggaa aaacaaaaca aaacaaaaca accaaaaaat 900  
 aaaatTTTTT tttccacag gttctagtgg aggtgctcac taccagaaat cctacaaata 960  
 agcccatctc atggatcagg gtttacctt gtaataatat taaatctgtg tgcattgtgcg 1020  
 cacgcatgtg ttttatgctt gcatatatgt atacgcagcc atggttttct actgtcccac 1080  
 tcaactctgta acttactgag ccatccagct ggtcctctaa atacatttca atgaaagttt 1140  
 tcattagcgt gaacgtgaag gtggtaaaat ctgttagtgt gtgcttatgc ctgtggtttg 1200  
 cacctctagt ctgaagggtg ctcttttcaa attttttatt tatttacgtt tttacttttg 1260  
 agtcagaaac tcataaaggc caagctggcc tcgaattcgc tatgtagtca atgatgacct 1320  
 taaacttgtg accctctact tcgtagtgc tggaaaccca agcttgctga gtacagagca 1380  
 ctttcagacc ggaactagat gtctacttcc tgttccgcct acattacagg ttgctagggt 1440  
 acaccccccc tacgccgttt tagacgcaa acttcatttc ccatgcaaaa cttcatttcc 1500

P108872WO rps UCOEs.ST25

catgaacact tgcaagggtc gccgcgctgc gcggcgctcat tgctcccgcc ctatatacct 1560  
 acttccgccc gcgagccact tcctttcctt tcagcggcgc gcggctgcaa gatggcggtg 1620  
 cagatttcca agaagaggaa ggtaagcgtc tgggcccggt tcgggagtc gcccggggtt 1680  
 ctacaagtgc cagggaggcc tgtggctccg tgatcagtcc tgtggagcgt ctggggccgc 1740  
 ctgccgtctc ttcgagcctc ggatggccgt agattgtgta ttgggccgga gccgggagcag 1800  
 tgctgtgtgc ctgggcaagg gagggacaaa ctctcgagt tctggaccga ctcgaacacc 1860  
 gggcgcctcc agttccggac tagacacctt tgagcgtttc ttggtctcca taatagtaat 1920  
 cctgtggcac agttagaggc cgtgtgccat cagatctagt ccagtttctt tagtaagtga 1980  
 agtttagcag tcccttctct tagtcgctg atcctgcaag tggccatagt tgaaagccta 2040  
 cttactgact gctgccgtgt tctctcggga cccggagctg cagcgtccct gtggttatca 2100  
 tttcatgggg gaaaagtgtg caggttgcca ggtttagaaa tagatggtct gtcgtttgtg 2160  
 cttatgcaca cagatgataa acctgttttg agtcaggatt cctctcctat ccgaggtaca 2220  
 acttacagtc ccagctgtac atgtgctact tggagacaga tttttctttg tctcttgggg 2280  
 gtagattatg ccgtagagcc cttcgatgaa gaggtgatga cgagtctgag taggaagtgt 2340  
 tgtctttgtc caagatgcct cactatgctg cgttctgtgg cacagctgaa agcactgtgg 2400  
 tcaaaagaaa cttcctaaag atgaccaaga ggcatttgtc tgagaagggg tgctgctttt 2460  
 ctgtagggcc attgggcttg ctctgactaa ccctgtcttc acctcagagg taacttgttt 2520  
 cctttggttc agtttgtagc tgatggcatc ttcaaagctg agctgaatga atttctcact 2580  
 cgggagctgg ctgaagatgg ctactctgga gttgaagtcc gagttacacc aaccaggaca 2640  
 gaaatcatta ttttagccac caggtagaaa taccattgat tgtcacctgt aaatactgtg 2700  
 tgtactgaga tgctgtgtaa acttgggcca accaagcagt aaatctggcc tcagtgggtg 2760  
 taactgcttt gttagaactg catttgggaa gaacttacct tccatttaac gtgtgtgctg 2820  
 gcgttgtggt gggcggcagg tgggatcttg agtaaaggt tgcgcttccc ctctacagga 2880  
 cacagaatgt tcttggggag aagggtcgtc ggatcagaga gttgaccgca gttgtccaga 2940  
 agcgcttttg cttccctgaa ggcagcgtag aggtgagttc ctctgcttta tctcccgggg 3000  
 gtttttagact gagttgggat gtggcttctg ctatagaatt gtacttctga aaacctgaca 3060  
 tggccagtga cagtcacagg tacttgatgc tctgagggcg aattctgcag atatccatca 3120  
 cactggcggc cgctcgagca tgcatt 3145

<210> 2  
 <211> 3039  
 <212> DNA  
 <213> Mus musculus rps11

<400> 2  
 aattcgccct taagactggt tgccctatgc ctgcctggcc tgcccttctt ccgccgcaa 60  
 ctagggaagt ggggacacaa ggttccttag gcactgctcc tgtgggtaga ggggacatta 120  
 gagagctgac agcgcaccac ctgcatgagt ttttattaaa gtgcaaacca tgggatgaat 180

## P108872WO rps UCOEs.ST25

cagttgagct	tcagtgttga	aatgagtag	cagggctgcc	ccacccacct	gaccaagtac	240
cctattctgc	agctatgaaa	atgagatctg	cacatgagct	ggggttcaca	agtgcacact	300
tggagcactg	ccttgctcct	tcccagcaga	ccacaaagca	gtatTTTTct	ggaggatttt	360
atgtgcta	aaattatttg	acttaagtgt	gtacgatgtg	tgctgtgcag	agaggggag	420
agggcaccag	caggtcatct	gcatgggggg	ccccttggg	tgaatccttg	ctcacgggat	480
aggctttggt	gctcaaaagt	tgcatatata	catcttgggt	cctgtcctag	atgggtttac	540
tgtaagtcag	caccaagata	caagagctgg	tacctggact	gtaggaggtc	aggccatgac	600
acaaaggctg	ggactaaagg	catttaccac	gcctgagtct	tctggttcct	taaacaatcaa	660
atccttccgg	gggctggcga	gatggctcag	tggttaagag	cacagactgc	tcttacgaag	720
gatccgagtt	caaatccag	caaccaaag	gtgcctaaca	actatccata	atgaaatctg	780
atgccctctt	ctggagtatc	tgagaacagc	tacagtgtac	ttacatataa	tcttaaaaaat	840
gcttcccatg	ttaaccacca	ctagagtttt	tattacagct	agctgacctg	gaagccaagt	900
ccttatgcct	ccgtgagtgc	tggggttaaa	aagatccagc	accactcaaa	atgtcaatct	960
atTTTgaaaa	tatgctttat	actgttctag	cccatctgtg	cagggctaga	acgggtgaata	1020
cgagaaactg	acacaagctt	ttgccacctg	gctaaatggt	tcctctatta	cctgggggtgg	1080
tcacctaagg	ttagacactc	atccacgagt	agtcaggaca	taaaccatc	aaagtgtggg	1140
tagacgcgca	gcctgagata	ctgtcaacaa	aggacatgcg	accttgggtga	cgtcggcctt	1200
taataaaagg	aagaaagggt	gactattcgg	tcgacgctgg	ctgctcctga	catcgtatgg	1260
cagatactct	gctgtaaagc	ggttcacccc	tttcttgaga	cccgtctgctg	acggccgctt	1320
ctctctggaa	actgaatccc	agcacgtggt	tccaacccg	tacggcacgc	cttctccgcc	1380
ctaagcctcg	ccgtaccaca	tgatgcacgt	ttcctccaca	tcgtgctcct	gaaatctcgc	1440
gagatgatag	gatcttcccg	ccccttagtc	ctccccgctc	atggcggcgt	acggacagtc	1500
ccaggaacgc	gggctctcgc	cggaagtacc	tcccacctcc	gtgaggataa	ccccgcgtca	1560
cttccgcccc	gacctcgcgt	ggtgaataag	gaagccggga	gcgggccctgc	ctctcccttt	1620
ctccggcggc	cggaagatg	gcgacattc	aggttcgagc	gtttagttgc	tttccccga	1680
cgcttcggtg	tggagcgtat	cccttggcgt	cctcgttgtc	ttacgatta	gctgaagcga	1740
ggatgcctgc	gaatgccttc	gtctcaggcg	gctcggaaat	ccgggctcta	cgagtaatg	1800
gggtccctgg	cgcttcggga	gttggttcct	aaagctcaga	gcttaacggg	tgaggattg	1860
tggcgggagg	agggcatcct	gcggcgagg	agtcctgagg	cggcagagcc	ggggacactg	1920
ggtaaagcag	gttttttccc	cttgatggag	actgaggccc	ggacctcgtg	cgctctacgg	1980
cagggctgag	gtcccacact	cgctgtagtt	ttcagtgtga	gcgagctct	ggcctcagatg	2040
agcttaggct	tgtcttaaac	ttgccatcct	gcctcaacct	caaccgggat	gacagatccg	2100
gcccaccagg	ctcggctacg	tggacataag	cttgaatccc	gaatgagtgg	atTTTgtatgt	2160
tttggaggtc	cagtctggct	gaaaagctct	ttttgatctc	agccgtgagt	tctgcaggct	2220

P108872WO rps UCOEs.ST25

gtggagggtgt tagatgggac gcagtggtg agctaaacta gacttgggggt ggttggagag 2280  
 ccctgaccag ccggttttgg cgattggggc aaataagggt gaaggtagga aggaagaaat 2340  
 attgtctctg atttccttga actttacctg caacctcacc aaattctcat ccctacagac 2400  
 ggagcgtgct taccaaaagc agcctacgat ctttcaaac aagaagcggg ttctgctggg 2460  
 agaaaccggc aaggaaaaac tccctcggta ctacaagaat atcgggtctag gcttcaagac 2520  
 gcctaaagag gtacaggacc ctccagcaga tgagatccct gctgccctgc acgtgtggga 2580  
 gcacagccac cccgccccct tcacagtggc ttcccatggg cccctgggaa ttgtagtatg 2640  
 ggccctgagg cgtcatcctt ggttctgttt aggaagtggg aatctaaacc ccactttctt 2700  
 aactttgcag gctattgagg gtacctacat agacaagaaa tgccccttca ctggtaacgt 2760  
 ctccatccga ggtcggatcc tgtctggtga gtgggatggt ggaaggggtg ttctagggtc 2820  
 ctgcgctccag gggcgctggc aagtgatgtc tgttctcacg atggcttca gatgtcctct 2880  
 agggcactgc tgagacagcc agttgacaaa gctgatgcca taaatggagc ttcttgggag 2940  
 ccccgttcaa ctgactccta cctgctaaca cttttctggt actctcccag gtgtcgtgac 3000  
 gaagatgaag atgcagagga ccattgtcat ccaagggcg 3039

<210> 3  
 <211> 22  
 <212> DNA  
 <213> Artificial

<220>  
 <223> synthetic PCR oligonucleotide

<400> 3  
 gtagtttctg acttgaaat gt 22

<210> 4  
 <211> 21  
 <212> DNA  
 <213> Artificial

<220>  
 <223> synthetic PCR oligonucleotide

<400> 4  
 aactgacctg ccatgccatt c 21

<210> 5  
 <211> 19  
 <212> DNA  
 <213> Artificial

<220>  
 <223> synthetic PCR oligonucleotide

<400> 5  
 gagcagtccg gtgtcacta 19

<210> 6  
 <211> 21

P108872WO rps UCOEs.ST25

<212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 6  
 gcagagaagc agacagttat g 21

<210> 7  
 <211> 21  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 7  
 cagagcatca agtacctgtg a 21

<210> 8  
 <211> 21  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 8  
 taaccactaa gccatctctc c 21

<210> 9  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 9  
 caagaacaag gaagctggcc 20

<210> 10  
 <211> 22  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 10  
 taaaaccac agcactgtag gg 22

<210> 11  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 11  
 aagactgttt gcctcatgcc 20

P108872WO rps UCOEs.ST25

<210> 12  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 12  
 ggatgacaat ggtcctctgc 20

<210> 13  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 13  
 gtggtttag gttcacgtcc 20

<210> 14  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 14  
 atccctcaca ttgccaagcc 20

<210> 15  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 15  
 acttaagacc tgatgcctcc 20

<210> 16  
 <211> 20  
 <212> DNA  
 <213> Artificial  
  
 <220>  
 <223> Synthetic PCR oligonucleotide  
  
 <400> 16  
 gctagcttac ataggcagcc 20

<210> 17  
 <211> 8691  
 <212> DNA  
 <213> Artificial  
  
 <220>

P108872WO rps UCOEs.ST25

&lt;223&gt; Vector pRPS3 1005 EGFP

&lt;400&gt; 17

cgttgtaaaa	cgacggccag	tgaattgtaa	tacgactcac	tatagggcga	attgggtacc	60
gggccccccc	tcgaagttta	aacattttaa	tctagtaacg	gccgccagtg	tgctggaatt	120
cgcccttata	accactgagc	catctctcca	gccctgagtc	atgattttag	tgtgagaggc	180
atcattgaat	tttctgagca	cggccatcag	ggtagctggc	acaggtcttc	agatacaagg	240
agatagttat	aagaaggcag	ccatggctgt	ggtgcactag	aaatggagaa	acagcttcat	300
caggtgacag	accagtctga	ctctgtccca	tgattagaag	ccatcttggt	acaaggtaa	360
aataagttca	ttcctgtttt	ctgtaacact	tgggtttgat	cctgtcgtca	accatttttc	420
tggaatttga	catgttccat	actccattat	accctgactt	ccaccctgat	aagatgttct	480
gccaagttcc	tgtgtagcca	acattcccct	ggaaatctct	cttcccctgg	aaaccaccta	540
gtcttagaaa	ttttgagtta	tataaattcc	acttctatgt	ttgatgctat	tctttaaacc	600
tccactttag	ggagatagcc	ctgtctgata	gaaaataaaa	cttgcttaat	ttgtctaaaa	660
gagtttaagt	aatagttttt	acttttgttc	cgtgggatta	atacagggtg	aaacagactc	720
ccgtgtttcc	agtgtgaagt	gagccacaca	ctgcagtaca	agttatatca	gcaggttctg	780
cctctgggca	atgaactttt	gcttgtgtgg	acatcagggg	ctgtgtgaag	ggaaggctct	840
atggcctaga	tttatactat	tcaacagtct	gtccccgaag	ccctggtgct	ttattatttt	900
gacaagcccc	tgctgctggt	attccaccct	gctgcgagtc	aaaaaagttc	ctgtctcgga	960
aaaacaaaaac	aaaacaaaaac	aaccaaaaaa	taaaattttt	ttttcccaca	ggttctagtg	1020
gaggtgctca	ctaccagaaa	tcctacaaat	aagcccatct	catggatcag	ggtttacctt	1080
tgtaataata	ttaaatctgt	gtgcatgtgc	gcacgcatgt	gttttatgct	tgcatatatg	1140
tatacgcagc	catggttttc	tactgtccca	ctcactctgt	aacttactga	gccatccagc	1200
tggctctcta	aatacatttc	aatgaaagtt	ttcattagcg	tgaacgtgaa	ggtggtaaaa	1260
tctgttagtg	tgtgcttatg	cctgtggttt	gcacctctag	tctgaagggt	gctcttttca	1320
aattttttat	ttatttacgt	ttttactttt	gagtcagaaa	ctcataaagg	ccaagctggc	1380
ctcgaattcg	ctatgtagtc	aatgatgacc	ttaaacttgt	gaccctctac	ttcgtttagtg	1440
ctggaacccc	aagcttgctg	agtacagagc	actttcagac	cggaactaga	tgtctacttc	1500
ctgttccgcc	tacattacag	gttgctaggt	tacaccccc	ctacgccgtt	ttagacgcaa	1560
aacttcattt	cccatgcaaa	acttcatttc	ccatgaacac	ttgcaagggt	cgccgcgctg	1620
cgcggcgtca	ttgctcccgc	cctatatacc	tacttccgcc	cgcgagccac	ttcctttcct	1680
ttcagcggcg	cgcggtctga	agatggcggg	gcagatttcc	aagaagagga	aggtaagcgt	1740
ctgggcccgg	ttcgggagtc	cgccgcgggt	tctacaagtg	ccagggaggc	ctgtggctcc	1800
gtgatcagtc	ctgtggagcg	tctggggccg	cctgccgtct	cttcgagcct	cggatggccg	1860
tagattgtgt	attgggccgg	agccgggcga	gtgctgtgtg	cctgggcaag	ggagggacaa	1920
actcctcgag	ttctggaccg	actcgaacac	cgggcgcctc	cagttccgga	ctagacacct	1980

## P108872WO rps UCOEs.ST25

ttgagcgttt	cttggctctcc	ataatagtaa	tcctgtggca	cagttagagg	gcgtgtgcca	2040
tcagatctag	tccagtttct	ttagtaagtg	aagtttagca	gtcccttctc	ttagtcgcgt	2100
gatcctgcaa	gtggccatag	ttgaaagcct	acttactgac	tgctgccgtg	ttcactcggg	2160
acccggagct	gcagcgtccc	tgtggttatc	atttcatggg	ggaaaagtgt	gcaggttgcc	2220
aggtttagaa	atagatggtc	tgtcgtttgt	gcttatgcac	acagatgata	aacctgtttt	2280
gagtcaggat	tcctctccta	tccgaggtac	aacttacagt	cccagctgta	catgtgctac	2340
ttggagacag	atcttctctt	gtctcttggg	tgtagattat	gccgtagagc	ccttcgatga	2400
agaggtgatg	acgagtctga	gtaggaagtg	ttgtctttgt	ccaagatgcc	tcactatgct	2460
gcgttctgtg	gcacagctga	aagcactgtg	gtcaaaagaa	acttcctaaa	gatgaccaag	2520
aggcatttgt	ctgagaaggg	ttgctgcttt	tctgtagggc	cattgggctt	gctctgacta	2580
accctgtctt	cacctcagag	gtaacttggt	tcctttgggt	cagttttag	ctgatggcat	2640
cttcaaagct	gagctgaatg	aatttctcac	tcgggagctg	gctgaagatg	gctactctgg	2700
agttgaagtc	cgagttacac	caaccaggac	agaaatcatt	atcttagcca	ccaggtagaa	2760
ataccattga	ttgtcacctg	taaatactgt	gtgtactgag	atgctgtgta	aaactgggccc	2820
aaccaagcag	taaatctggc	ctcagtgggg	gtaactgctt	tgttagaact	gcatttggga	2880
agaacttacc	ttccatttaa	cgtgtgtgct	ggcgttgtgg	tgggcggcag	gtgggatctt	2940
gagtaaagtg	ttgcgcttcc	cctctacagg	acacagaatg	ttcttgggga	gaagggtcgt	3000
cggatcagag	agttgaccgc	agttgtccag	aagcgccttg	gcttccctga	aggcagcgtg	3060
gaggtgagtt	cctctgcttt	atctcccggg	ggtttttagac	tgagttggga	tgtggcttct	3120
gctatagaat	tgtacttctg	aaaacctgac	atggccagtg	acagtcacag	gtacttgatg	3180
ctctgagggc	gaattctgca	gatatccatc	acactggcgg	ccgctcgagc	atgcatctag	3240
aagcttatcg	ataccggtgg	cgcgccaatt	gaattaagat	ctggcccaat	gggccgtacg	3300
aattcgagct	cggtacccgg	ggatcctgat	ctaatagtaa	tcaattacgg	ggtcattagt	3360
tcatagccca	tatatggagt	tccgcgttac	ataacttacg	gtaaattggc	cgcttggctg	3420
accgccaac	gacccccgcc	cattgacgtc	aataatgacg	tatgttccca	tagtaacgcc	3480
aatagggact	ttccattgac	gtcaatgggt	ggagtattta	cggtaaactg	cccacttggc	3540
agtacatcaa	gtgtatcata	tgccaagtac	gccccctatt	gacgtcaatg	acggtaaagt	3600
gccccctgg	cattatgccc	agtacatgac	cttatgggac	tttctactt	ggcagtacat	3660
ctacgtatta	gtcatcgcta	ttaccatggg	gatgcggttt	tggcagtaca	tcaatgggcg	3720
tggatagcgg	tttgactcac	ggggatttcc	aagtctccac	cccattgacg	tcaatgggag	3780
tttgttttgg	caccaaatac	aacgggactt	tccaaaatgt	cgtaacaact	ccgccccatt	3840
gacgcaaagt	ggcggtaggc	gtgtacgggtg	ggaggtctat	ataagcagag	ctggtttagt	3900
gaaccgtcag	atccgtcgcc	accatgggtga	gcaagggcga	ggagctgttc	accgggggtgg	3960
tgcccatcct	ggtcgagctg	gacggcgacg	taaacggcca	caagttcagc	gtgtccggcg	4020

## P108872WO rps UCOEs.ST25

agggcgaggg	cgatgccacc	tacggcaagc	tgaccctgaa	gttcatctgc	accaccggca	4080
agctgcccgt	gccctggccc	accctcgtga	ccaccctgac	ctacggcgtg	cagtgcctca	4140
gccgctaccc	cgaccacatg	aagcagcacg	acttcttcaa	gtccgcatg	cccgaaggct	4200
acgtccagga	gcgcaccatc	ttcttcaagg	acgacggcaa	ctacaagacc	cgcgccgagg	4260
tgaagttcga	gggcgacacc	ctggtgaacc	gcatcgagct	gaagggcatc	gacttcaagg	4320
aggacggcaa	catcctgggg	cacaagctgg	agtacaacta	caacagccac	aacgtctata	4380
tcatggccga	caagcagaag	aacggcatca	aggtgaactt	caagatccgc	cacaacatcg	4440
aggacggcag	cgtgcagctc	gccgaccact	accagcagaa	cacccccatc	ggcgacggcc	4500
ccgtgctgct	gcccgacaac	cactacctga	gcaccagtc	cgccctgagc	aaagacccca	4560
acgagaagcg	cgatcacatg	gtcctgctgg	agttcgtgac	cgccgcccgg	atcactctcg	4620
gcatggacga	gctgtacaag	taaagcggcc	gcgactctag	atcataatca	gccataccac	4680
atltgtagag	gttttacttg	ctttaaaaaa	cctcccacac	ctccccctga	acctgaaaca	4740
taaaatgaat	gcaattgttg	ttgttaactt	gtttattgca	gcttataatg	gttacaataa	4800
aagcaatagc	atcacaaatt	tcacaaataa	agcatttttt	tcactgcatt	ctagttgtgg	4860
tttgtccaaa	ctcatcaatg	tatcttaact	agagtcgacc	tgaggcatg	caagcttacc	4920
ggtggcgcg	gcgccaattg	ttaattaaga	tctggcccaa	tgggcccgtac	gaattcctta	4980
ggctaccggg	taggggaggg	gcttttccca	aggcagtctg	gagcatgcg	tttagcagcc	5040
ccgctgggca	cttggcgcta	cacaagtggc	ctctggcctc	gcacacattc	cacatccacc	5100
ggccggtagg	cgccaaccgg	ctccgttctt	tggtggcccc	ttcgcgccac	cttctactcc	5160
tcccctagtc	aggaagttcc	ccccgcccc	gcagctcgcg	tcgtgcagga	cgtgacaaat	5220
ggaagtagca	cgtctcacta	gtctcgtgca	gatggacagc	accgctgagc	aatggaagcg	5280
ggtaggcctt	tggggcagcg	gccaatagca	gctttgctcc	ttcgctttct	gggctcagag	5340
gctgggaagg	ggtgggtccg	ggggcgggct	caggggcggg	ctcaggggcg	gggcggg	5400
ccgaaggctc	tccggaggcc	cggcattctg	cacgcttcaa	aagcgacgt	ctgccgcgct	5460
gttctcctct	tcctcatctc	cgggcctttc	gaccagctta	ccatgaccga	gtacaagccc	5520
acggtgcgcc	tcgccaccg	cgacgacgtc	cccagggccg	tacgcaccct	cgccgccg	5580
ttcgccgact	accccgccac	gcgccacacc	gtcgatccgg	accgccacat	cgagcggg	5640
accgagctgc	aagaactctt	cctcacgcgc	gtcgggctcg	acatcggcaa	ggtgtggg	5700
gcgacgacg	gcgccg	ggcggctctg	accacgccg	agagcgtcga	agcggggg	5760
gtgttcg	agatcggccc	gcgcatggcc	gagttgagcg	gttcccggct	ggccgcg	5820
caacagatgg	aaggcctcct	ggcgccgcac	cggcccaagg	agcccgcgtg	gttcttggcc	5880
accgtcggcg	tctcgcccga	ccaccagggc	aagggctctg	gcagcggcgt	cgtgctcccc	5940
ggagtggagg	cggccgagcg	cgccggggtg	cccgccttcc	tgagacctc	cgccccgc	6000
aacctcccct	tctacgagcg	gctcggcttc	accgtcaccg	ccgacgtcga	ggtgcccga	6060

## P108872WO rps UCOEs.ST25

ggaccgcgca	cctggtgcat	gacccgcaag	cccgggtgct	gacgcccgcc	ccacgacccg	6120
cagcgcgccga	ccgaaaggag	cgcacgaccc	catgcatcgt	agacgaaatg	accgaccaag	6180
cgacgcccga	cctgcatca	cgagatttcg	attccaccgc	cgcttctat	gaaagggttg	6240
gcttcggaat	cgttttccgg	gacgcccggct	ggatgatcct	ccagcgcggg	gatctcatgc	6300
tggagttctt	cgcccaccct	agggggaggc	taactgaaac	acggaaggag	acaataccgg	6360
aaggaacccg	cgctatgacg	gcaataaaaa	gacagaataa	aacgcacggg	gttgggtcgt	6420
ttgttcataa	acgcgggggt	cggtcccagg	gctggcactc	tgtcgatacc	ccaccgagac	6480
cccattgggg	ccaatacgcc	cgcgtttctt	ccttttcccc	acccacccc	ccaagtccgg	6540
gtgaaggccc	agggctcgca	gccaacgtcg	gggcggcagg	ccccagctt	ttgttccctt	6600
tagtgagggt	taatttcgag	cttggcgtaa	tcatggatc	agctgtttcc	tgtgtgaaat	6660
tgttatccgc	tcacaattcc	acacaacata	cgagccggaa	gcataaagt	taaagcctgg	6720
ggtgccta	gagtgagcta	actcacatta	attgcgttgc	gctcactgcc	cgctttccag	6780
tcgggaaacc	tgtcgtgcca	gcatcgcgag	cacttttcgg	ggaaatgtgc	gcggaacccc	6840
tatttgttta	tttttctaaa	tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	6900
ataaatgctt	caataatatt	gaaaaaggaa	gagtatgagt	attcaacatt	tccgtgtcgc	6960
ccttattccc	ttttttgcgg	cattttgcct	tcctgttttt	gctcaccag	aaacgctggt	7020
gaaagtaaaa	gatgctgaag	atcagttggg	tgcacgagtg	ggttacatcg	aactggatct	7080
caacagcggg	aagatccttg	agagttttcg	ccccgaagaa	cgttttccaa	tgatgagcac	7140
ttttaaagtt	ctgctatgtg	gcgcggtatt	atcccgtatt	gacgccgggc	aagagcaact	7200
cggtcgcccgc	atacactatt	ctcagaatga	cttggttgag	tactcaccag	tcacagaaaa	7260
gcatcttacg	gatggcatga	cagtaagaga	attatgcagt	gctgccataa	ccatgagtga	7320
taacactgcg	gccaacttac	ttctgacaac	gatcggagga	ccgaaggagc	taaccgcttt	7380
tttgcacaa	atgggggatc	atgtaactcg	ccttgatcgt	tgggaaccgg	agctgaaatga	7440
agccatacca	aacgacgagc	gtgacaccac	gatgcctgta	gcaatggcaa	caacgttgcg	7500
caaaactatta	actggcgaac	tacttactct	agcttcccgg	caacaattaa	tagactggat	7560
ggaggcggat	aaagttgcag	gaccacttct	gcgctcggcc	cttccggctg	gctggtttat	7620
tgctgataaa	tctggagccg	gtgagcgtgg	gtctcgcggg	atcattgcag	cactggggcc	7680
agatggtaag	ccctcccgta	tcgtagttat	ctacacgacg	gggagtcagg	caactatgga	7740
tgaacgaaat	agacagatcg	ctgagatagg	tgcctcactg	attaagcatt	ggtaactgtc	7800
agactcgcga	cactgcatta	atgaatcggc	caacgcgcgg	ggagaggcgg	tttgcgtatt	7860
gggcgctctt	ccgcttcctc	gctcactgac	tcgctgcgct	cggtcgttcg	gctgcggcga	7920
gcggtatcag	ctcactcaaa	ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	7980
ggaaagaaca	tgtgagcaaa	aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	8040
ctggcgtttt	tccataggct	ccgccccct	gacgagcatc	acaaaaatcg	acgctcaagt	8100

## P108872WO rps UCOEs.ST25

cagaggtggc gaaacccgac aggactataa agataccagg cgtttcccc tggagctcc 8160  
 ctctgtcgct ctctgtttcc gaccctgccg cttaccggat acctgtccgc ctttctccct 8220  
 tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc ggtgtaggtc 8280  
 gttcgctcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg ctgcgctta 8340  
 tccggtaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc actggcagca 8400  
 gccactggta acaggattag cagagcgagg tatgtaggag gtgctacaga gttcttgaag 8460  
 tgggtggccta actacggcta cactagaagg acagtatttg gtatctgctc tctgctgaag 8520  
 ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaac caccgctggt 8580  
 agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaagg atctcaagaa 8640  
 gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc a 8691

<210> 18  
 <211> 13827  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Vector pCET 1015 EGFP

<400> 18  
 cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggcga attgggtacc 60  
 gggccccccc tcgaagttta aacatttaaa tctagaagct tcaatgtttt tagcacctc 120  
 tgtgtggagg aaaataatgc agattattct aattagtgtata atatctaacc acattaaaat 180  
 atattacata gtaaactaca ctccataatt ttataaattt gactccccag ggtaataaac 240  
 tagtctctag tctgctcacc ttcaactgta caataaagtc ttggttcttt tgaaatagac 300  
 ctcaaatgag acacctaaaa ttcaaagtgt ctttacattt aaagacacct acaggaaagc 360  
 aggtaaaaga gccagggtta aaacaaattc taaaaccact tagctgcagt taaacatata 420  
 gtaaagatgc actaaagtgt cttactctgt aaatcccttc cacttcagga aatattccac 480  
 tttcccattc actacacgct gatctagtac tttttccacg acaaattctt caggctctgc 540  
 ctcttcaact tttttactct ttccattctg ttttttccc attttttgct aaaataaaaac 600  
 aaaagagaaa ttaagaaata ttctcttga attttgagca cttttcaag gctcaattgc 660  
 ttatattatt atcacattcg acataaattt ttacttctat atcccagggc agacaccttc 720  
 tggaaagatt aaaagtcaac agacaataaa ataaaagaat gctttatctt gttcatttag 780  
 ttcaaactta caaccacca ccaaataat acaataaaaa aacactatct ggaaacagtt 840  
 atttttttcc agtctttttt tttgagacag ggtctcacac tcttgtcgcc caggctggag 900  
 tgcaagtggc tgatctcagc tcaactgcaac ctccgcctcc ccaggttcaa gcagttctca 960  
 tgcctcagcc tccagagtag ctgggattat aggcggatgc caccatgccg ggctaatttt 1020  
 ttttgtgttt ttattagaaa cagggtttca ccatgttgac caggctggtc tcaaactcct 1080  
 gacctgaagt gattcaccag cctgggcctc ccaaagtgtc ggcattacag gcgtgagcca 1140

## P108872WO rps UCOEs.ST25

ctgcgccccg	ccctgtagtc	ttaaaagacc	aagtttacta	atthttcactc	atthttaacaa	1200
cactgcaaca	aacaactatg	caggaagtac	ctaaaggggtg	atccagagaa	gcaagtagta	1260
gtgacaggtc	ttaggtgaac	ctatgacaga	ccttgatcc	acccccagat	ggtaaaagcc	1320
ccagccccct	tctcaattca	aatattaatg	tcaaaagcat	caatgatata	gagaaaagat	1380
aaatgcagaa	tgaaaacatg	gttcaaaatc	ctgatacca	ctgcagggtc	aactatagag	1440
accactagga	ggttcaatta	aaggacaaga	ttatthttcc	ataatctctg	tagataatat	1500
ttcttaccac	ttagaacaaa	actataaagc	tatcacttca	agagaccaac	attacaaatt	1560
tatthttaatt	ccctaagggtg	aaaaaaatcc	ttccttcctg	gthttctcaag	agaaagtcta	1620
tactggtaac	caaattcact	ttaaacaggc	atthttcttg	gtatgacact	atthtaagaga	1680
agcaggaaac	caacgtgaac	cagctctttc	caatggctca	agatthtcta	tgagaggact	1740
aaaaatgggg	aaaatthtta	tgagaggatt	aaaaatgggg	gaaaaaaaac	cctgaaatgg	1800
ttaatcagaa	gatcctatgg	gctgagaagg	aatccatctt	aacatthcat	cttaaagcaa	1860
atgctattgc	cgggggcagt	ggctcatgcc	tgtaatccca	gcactthggg	aggccgaggt	1920
gggcagatca	tctgagggtca	ggagthtgag	accagcctga	ccaacatgga	gaaaccccg	1980
ttctactaaa	aatacaaaat	tagccaggca	tagtggtgca	tgctgtaat	cccagctact	2040
tgggaggctg	aggcaggaga	actgcttgaa	cccaggaggc	ttaagthg	gtgagccaag	2100
atcacgccat	tgactcttag	cctggacaac	aagagaaaa	ctctgtctca	aaaaaacaca	2160
aaaacaaaaa	acccaatac	tatthaaaaa	agataaacct	taatthgctca	atcattaag	2220
ccatcccaca	agtaaagcag	caagcagaaa	aaagthtaaga	acacctcaag	gctacagaag	2280
gacatthcaa	gctatgcagg	catatgaagt	gtgcagacag	atatgtaaga	aaggcctcaa	2340
gactgcaaaa	gggcatttca	agctatgcaa	gcatataggt	aacacataca	cacacacaaa	2400
ataaaatccc	ctgaaatata	aaaacatgca	gcaaacacct	gacgthtttg	gataccatth	2460
ctaagtcagg	tgthtatgatt	ctcattagtc	aagatacttg	agtactgggc	ccaaacagct	2520
ttctgccact	gtacagtata	agaaggtagg	aataatgggtg	ggaggagcaa	agacaaactg	2580
taatagacag	aagtgtatca	gataacctata	ctacatgaaa	aacaaaacag	ctactgccac	2640
aaaggagaa	ggctaacaaa	ataaagtcaa	caataaatac	agaaaatgaa	aaggatacac	2700
actaaggtht	acaaaaaaa	aaaggcagac	aaaatgccat	acagthattca	ttcactacta	2760
tggcattcat	aagctagtht	caaatgctca	ctatthttct	ttatagthata	tatthgctt	2820
aaccagcac	thttthtcaa	aagthgatga	gtcaaaaata	atthtccatt	atthtaagthga	2880
aatthaacagc	acacatact	cacaacacta	atgaatthtt	aaaatggaaa	gthtaagaact	2940
thtaaagthg	ccaacctgthg	atctthtca	aaataaacta	aatacaata	cagaccccaa	3000
aggctatcaa	thgctgcaa	aaacaacttc	gthttthccag	ggtaaacaga	atctaatgca	3060
gaatctaatg	caggthaaac	agactthaatg	cagaatctaa	thgatggcaca	aatthaaaaat	3120
cactaacgthg	ccctthtttag	thtgaaacc	agagagagca	catacaagcc	aaaaacaaat	3180

## P108872WO rps UCOEs.ST25

gctttatfff	acctaggaga	cattaacatt	cacctttacg	tgtttaagat	taatgcaatg	3240
ttaaataattg	tgaaaactgt	aactttgaat	ttcatgattt	ttatgtgaat	attccagggt	3300
ttaaaaaaac	ttgtaacatg	acatggctga	ataagataaa	aaaaaaatct	agccttttct	3360
cccttctggc	tcataattgc	gatttcgatc	atfttgftta	aaaaacaaaa	cactgcaatg	3420
aattaaactt	aatattcttc	tatgttttag	agtaagftaa	aacaagataa	agtgacccaaa	3480
gtaatttgaa	agattcaatg	acttttgctc	caacctagggt	gcacaaggta	ccttgttctt	3540
taaattgggc	tttaatgaaa	atacttctcc	agaattctgg	ggatttaaga	aaaattatgc	3600
caaccaacaa	gggctttacc	atftttatgta	acatftttca	acgctgcaaa	aatgtgtgta	3660
tttctatftg	aagataaaaa	tcctcagcaa	aatccacatt	gcactgtcct	tcaaagatta	3720
gccttctftg	aactagftaa	gacactatta	agccaagcca	gtatctccct	gtaatgaatt	3780
cgfttttctc	ttaattfttc	cctgtaattt	acactgggag	agctgggaaa	tatgtggatg	3840
taaatttctc	agccacagag	atgcaaagtt	atactgtggg	gaaaaaaaaac	ttgagttaaa	3900
tccttacata	ttfttagftt	tcattaactt	accaatgtag	ttftgttgga	ggccatfttt	3960
tttattgcag	acttgaagag	ctattactag	aaaaatgcat	gacagttaag	gtaagfttgc	4020
atgacacaaa	aaaggtaact	aaatacaaat	tctgfttgga	ttccaacccc	caagtagaga	4080
gcgcacactt	tcaaacgtga	atacaaatcc	agagtagatc	tgcgctccta	cctacattgc	4140
ttatgatgta	cttaagtacg	tgtcctaacc	atgtgagtct	agaaagactt	tactggggat	4200
cctggtacct	aaaacagctt	cacatggctt	aaaatagggg	accaatgtct	tttccaatct	4260
aagtcccatt	tataataaag	tccatgttcc	atftttaaag	gacaatcctt	tcggtttaaa	4320
accaggcacg	attacccaaa	caactcacia	cggtaaagca	ctgtgaaatct	tctctgttct	4380
gcaatcccaa	cttggfttct	gctcagaaac	cctccctctt	tccaatcggg	aattaaataa	4440
caaaaggaaa	aaacttaaga	tgcttcaacc	ccgfttctgt	acactfttgaa	aaaagaatca	4500
cctcttgcaa	acacccgctc	ccgacccccg	ccgctgaagc	ccggcgtcca	gaggcctaag	4560
cgcggttgcc	cgccccacc	cgggagcgcg	ggcctctgtg	tcagcgcatc	cgcggggaga	4620
aacaaaggcc	gcggcacggg	ggctcaaggg	cactgcgcca	caccgcacgc	gcctaccccc	4680
gcgcgccac	gttaactggc	ggtcgccgca	gcctcgggac	agccggccgc	gcgcccag	4740
gctcgcgac	gcgggaccac	gcgccgccct	ccgggaggcc	caagtctcga	cccagccccg	4800
cgtggcgctg	ggggagggg	cgctccgcc	ggaacgcggg	tgggggagg	gagggggaaa	4860
tgcgctftgt	ctcgaaatgg	ggcaaccgtc	gccacagctc	cctacccccct	cgagggcaga	4920
gcagtcccc	cactaactac	cgggctggcc	gcgcgccagg	ccagccgcga	ggccaccgcc	4980
cgaccctcca	ctcttccc	cagctcccgg	cgcggggtcc	ggcgagaagg	ggaggggagg	5040
ggagcggaga	accgggcccc	cgggacgcgt	gtggcatctg	aagcaccacc	agcgagcgag	5100
agctagagag	aaggaaagcc	accgacttca	ccgcctccga	gctgctccgg	gtcgcgggtc	5160
tgacgcgtct	ccggccctcc	gcgcctacag	ctcaagccac	atccgaagg	ggagggagcc	5220

## P108872WO rps UCOEs.ST25

gggagctgcg cgcggggccc ccgggggggag ggggtggcacc gcccacgccg ggcggccacg 5280  
 aagggcgggg cagcgggcgc gcgcgcggcg gggggagggg ccggcgccgc gcccgctggg 5340  
 aattggggcc ctaggggggag ggcggaggcg ccgacgaccg cggcacttac cgttcgcggc 5400  
 gtggcgcccg gtggtcccca aggggagggg agggggaggc ggggcgagga cagtgaccgg 5460  
 agtctcctca gcggtggctt ttctgcttgg cagcctcagc ggctggcgcc aaaaccggac 5520  
 tccgcccact tcctcgcccg ccggtgagag ggtgtggaat cctccagacg ctgggggagg 5580  
 gggagtggg agcttaaaaa ctagtacccc tttgggacca ctttcagcag cgaactctcc 5640  
 tgtacaccag gggtcagttc cacagacgcg ggccaggggt gggtcattgc ggcgtgaaca 5700  
 ataatttgac tagaagttga ttcgggtgtt tccggaagg gccgagtcaa tccgccgagt 5760  
 tggggcacgg aaaacaaaaa ggggaaggcta ctaagattht tctggcgggg gttatcattg 5820  
 gcgtaactgc agggaccacc tcccgggttg agggggctgg atctccaggc tgcggattaa 5880  
 gccctcccg tcggcgtaa tttcaaactg cgcgacgttt ctacactgcc ttcgccaagg 5940  
 caggggcccg gaccctattc caagaggtag taactagcag gactctagcc ttccgcaatt 6000  
 cattgagcgc atttacggaa gtaacgctcg gtactgtctc tggccgcaag ggtgggagga 6060  
 gtacgcattt ggcgtaaggt gggcgtaga gccttcccgc cattggcggc ggatagggcg 6120  
 tttacgcgac ggcctgacgt agcgggaagac gcgtagtggt gggggaagg tctagaaaag 6180  
 cggcggcagc ggctctagcg gcagtagcag cagcgccggg tcccgtgagg aggtgctcct 6240  
 cgagagttg tttctcgagc agcggcagtt ctactacag cgcaggacg agtccggttc 6300  
 gtgttcgtcc gcggagatct ctctcatctc gctcggctgc gggaaatcgg gctgaagcga 6360  
 ctgagtccgc gatggaggta acgggtttga aatcaatgag ttattgaaaa gggcatggcg 6420  
 aggccgttg cgctcagtg gaagtcggcc agccgcctcc gtgggagaga ggaggaat 6480  
 cggaccaatt cagtagcagt ggggcttaag gtttatgaac ggggtcttga gcggaggcct 6540  
 gagcgtacaa acagcttccc caccctcagc ctcccggcg catttccct cactgggggt 6600  
 gggggatggg gagctttcac atggcggacg ctgccccgt ggggtgaaag tggggcgcg 6660  
 aggcgggaat tcttattccc tttctaaagc acgctgctc gggggccacg gcgtctcctc 6720  
 ggcgagcgtt tcggcgggca gcaggtctc gtgagcgagg ctgaggagct tcccctcccc 6780  
 ctctctccc ggaaccgatt tggcggccgc cattttcatg gctcgcctc ctctcagcgt 6840  
 tttccttata actcttttat tttcttagtg tgctttctc atcaagaagt agaagtgggt 6900  
 aactatthtt ttttcttct cgggctgttt tcatatcgtt tcgaggtgga tttggagtgt 6960  
 tttgtgagct tggatcttta gagtctgcy cacctcatta aaggcgtca gccttcccct 7020  
 cgatgaaatg gcgccattgc gttcggagc cacaccgaag agcggggagg gggggtgctc 7080  
 cgggtttgcy ggccccgttt cagagaagat atcaccacc agggcgtcgg gccgggttca 7140  
 atgagagcc taggacaaag aaaccattht atgtthttcc tgtctthttt ttcctttgag 7200  
 taacggthtt atctgggtct gcagtcagta aaacgacaga tgaaccgcy caaaataaac 7260

## P108872WO rps UCOEs.ST25

ataaattgga	agccatcggc	cacgaggggc	agggacgaag	gtggttttct	gggcggggga	7320
gggatattcg	cgtcagaatc	ctttactggt	cttaaggatt	ccgtttaagt	tgtagagctg	7380
actcatttta	agtaatgttg	ttactgagaa	gtttaaccct	tacgggacag	atccatggac	7440
ctttatagat	gattacgagg	aaagtgaaat	aacgattttg	tccttagtta	tacttcgatt	7500
aaaacatggc	ttcagaggct	ccttcctgta	atgcgtatgg	attgatgtgc	aaaactgttt	7560
tgggcctggg	ccgctctgta	tttgaacttt	gttacttttc	tcattttggt	tgcaatcttg	7620
gttgaacatt	acattgataa	gcataaggtc	tcaagcgaag	ggggtctacc	tggttatfff	7680
tctttgaccc	taagcacggt	tataaaataa	cattgtttaa	aatcgatagt	ggacatcggg	7740
taagtttggg	taaattgtga	ggtaagtaat	gagtttttgc	tttttgtag	tgatttgtaa	7800
aacttgttat	aatgtacat	tatccgtaat	ttcagtttag	agataaccta	tgtgctgacg	7860
acaattaaga	ataaaaacta	gctgaaaaaa	tgaaaataac	tatcgtgaca	agtaaccatt	7920
tcaaaagact	gctttgtgtc	tcataggagc	tagtttgatc	atctcagtta	atfttttctt	7980
taatftttac	gagtcatgaa	aactacagga	aaaaaaatct	gaactggggt	ttaccactac	8040
tttttaggag	ttgggagcat	gcgaatggag	ggagagctcc	gtagaactgg	gatgagagca	8100
gcaattaatg	ctgcttgcta	ggaacaaaaa	ataattgatt	gaaaattacg	tgtgactfff	8160
tagtttgcat	tatgcgtttg	tagcagttgg	tcctggatat	cactttctct	cgtttgaggt	8220
tttttaacct	agttaacttt	taagacaggt	ttccttaaca	ttcataagtg	cccagaatac	8280
agctgtgtag	tacagcatat	aaagatttca	gctctgaggt	ttttcctatt	gacttggaaa	8340
attgftttgt	gcctgtcgct	tgccacatgg	ccaatcaagt	aagcttatcg	ataccggtgg	8400
cgcgccaatt	gaattaagat	ctggcccaat	gggccgtacg	aattcgagct	cggtaccggg	8460
ggatcctgat	ctaatagtaa	tcaattacgg	ggtcattagt	tcatagccca	tatatggagt	8520
tccgcgttac	ataacttacg	gtaaatggcc	cgctggctg	accgccaac	gacccccgcc	8580
cattgacgtc	aataatgacg	tatgttccca	tagtaacgcc	aatagggact	ttccattgac	8640
gtcaatgggt	ggagtattta	cggtaaactg	cccacttggc	agtacatcaa	gtgtatcata	8700
tgccaagtac	gccccctatt	gacgtcaatg	acggtaaatg	gccccctgg	cattatgccc	8760
agtacatgac	cttatgggac	tttcctactt	ggcagtacat	ctacgtatta	gtcatcgcta	8820
ttaccatggt	gatgcggttt	tggcagtaca	tcaatgggcg	tggatagcgg	tttgactcac	8880
ggggatttcc	aagtctccac	cccattgacg	tcaatgggag	tttgttttgg	caccaaatac	8940
aacgggactt	tccaaaatgt	cgtaacaact	ccgccccatt	gacgcaaatg	ggcggtaggc	9000
gtgtacggtg	ggaggtctat	ataagcagag	ctggtttagt	gaaccgtcag	atccgtcgcc	9060
accatggtga	gcaagggcga	ggagctgttc	accggggtgg	tgcccatcct	ggtcgagctg	9120
gacggcgacg	taaacggcca	caagttcagc	gtgtccggcg	agggcgaggg	cgatgccacc	9180
tacggcaagc	tgaccctgaa	gttcatctgc	accaccggca	agctgcccgt	gcccctggccc	9240
accctcgtga	ccaccctgac	ctacggcgtg	cagtgttca	gccgctaccc	cgaccacatg	9300

## P108872WO rps UCOEs.ST25

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgaccatc 9360  
ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 9420  
ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 9480  
cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 9540  
aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 9600  
gccgaccact accagcagaa caccacctc ggcgacggcc ccgtgctgct gcccgacaac 9660  
cactacctga gcaccagtc cgccctgagc aaagaccca acgagaagcg cgatcacatg 9720  
gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 9780  
taaagcggcc gcgactctag atcataatca gccataccac atttgtagag gttttacttg 9840  
ctttaaaaaa cctcccacac ctccccctga acctgaaaca taaaatgaat gcaattgttg 9900  
ttgttaactt gtttattgca gcttataatg gttacaaata aagcaatagc atcacaatt 9960  
tcacaaataa agcatttttt tcaactgcatt ctagtgtggg tttgtccaaa ctcatcaatg 10020  
tatcttaact agagtcgacc tgcaggcatg caagcttacc ggtggcgcgc caattgttaa 10080  
ttaagatctg gcccaatggg ccgtacgaat tccttaggct accgggtagg ggaggcgctt 10140  
ttccaaggc agtctggagc atgcgcttta gcagccccgc tgggcacttg gcgctacaca 10200  
agtggcctct ggcctcgac acattccaca tccaccggcc ggtaggcgcc aaccggctcc 10260  
gttctttggt ggccccctcg cgccacctt tactcctccc ctagtcagga agttcccccc 10320  
cgccccgcag ctcgcgctgt gcaggacgtg acaaatggaa gtagcacgtc tcaactgtct 10380  
cgtgcagatg gacagcaccg ctgagcaatg gaagcgggta ggcctttggg gcagcggcca 10440  
atagcagctt tgctccttcg ctttctgggc tcagaggctg ggaaggggtg ggtccggggg 10500  
cgggctcagg ggcgggctca ggggcggggc gggcgcccga aggtcctccg gaggcccggc 10560  
attctgcacg cttcaaaagc gcacgtctgc cgcgctgttc tcctcttct catctccggg 10620  
cctttcgacc agcttaccat gaccgagtac aagcccacgg tgcgcctcgc caccgcgac 10680  
gacgtccca gggccgtacg caccctcgcc gccgcgttcg ccgactacc cgccacgcgc 10740  
cacaccgtcg atccggaccg ccacatcgag cgggtcaccg agctgcaaga actcttcctc 10800  
acgcgcgtcg ggctcgacat cggcaaggtg tgggtcgcgg acgacggcgc cgcggtggcg 10860  
gtctggacca cgccggagag cgtcgaagcg ggggcgggtg tcgccgagat cggcccgcgc 10920  
atggccgagt tgagcggttc ccggctggcc gcgcagcaac agatggaagg cctcctggcg 10980  
ccgcaccggc ccaaggagcc cgcgtggttc ctggccaccg tcggcgtctc gcccgaccac 11040  
cagggcaagg gtctgggcag cgccgtcgtg ctccccggag tggaggcggc cgagcgcgc 11100  
ggggtgcccg ctttctgga gacctccgc ccccgcaacc tccccttcta cgagcggctc 11160  
ggcttcaccg tcaccgccga cgtcgaagtg cccgaaggac cgcgcacctg gtgcatgacc 11220  
cgcaagcccg gtgcctgacg cccgccccac gaccgcagc gcccgaccga aaggagcgca 11280  
cgaccccatg catcgtagac gaaatgaccg accaagcgac gcccaacctg ccatcacgag 11340

## P108872WO rps UCOEs.ST25

atttcgattc caccgccgcc ttctatgaaa ggttgggctt cggaatcgtt ttccgggacg 11400  
 ccggctggat gatcctccag cgcggggatc tcatgctgga gttcttcgcc caccctaggg 11460  
 ggaggctaac tgaaacacgg aaggagacaa taccggaagg aaccgcgct atgacggcaa 11520  
 taaaaagaca gaataaacg cacggtggtg ggtcgtttgt tcataaacgc ggggttcggt 11580  
 cccagggctg gactctgtc gatacccac cgagaccca ttggggccaa tacgcccgcg 11640  
 tttcttcctt ttccccacc ccccccaa gttcgggtga aggccaggg ctgcagcca 11700  
 acgtcggggc ggaggcccc cagcttttgt tcccttagt gagggttaat ttcgagcttg 11760  
 gcgtaatcat ggtcatagct gtttctgtg tgaattggt atccgctcac aattccacac 11820  
 aacatacgag ccggaagcat aaagtgtaaa gcctggggtg cctaatgagt gagctaactc 11880  
 acattaattg cgttgcgctc actgcccgc tccagtcgg gaaacctgct gtgccagcat 11940  
 cgcgagcact ttcggggaa atgtgcgcg aaccctatt tgtttat tctaaataca 12000  
 ttcaaatag tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 12060  
 aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 12120  
 ttgccttctt gttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 12180  
 gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 12240  
 ttttcgcccc gaagaacggt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 12300  
 ggtattatcc cgtattgacg ccgggcaaga gcaactcggc cgccgcatac actatttctca 12360  
 gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 12420  
 aagagaatta tgcagtgctg ccataacat gatgataac actgcggcca acttacttct 12480  
 gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt 12540  
 aactcgctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 12600  
 caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 12660  
 tactctagct tcccggcaac aattaataga ctggatggag gcgataaag ttgcaggacc 12720  
 acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga 12780  
 gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 12840  
 agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga 12900  
 gataggtgcc tcaactgatta agcattggta actgtcagac tcgacgact gcattaatga 12960  
 atcgccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgtc 13020  
 actgactcgc tgcgctcggc cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg 13080  
 gtaatacggc tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 13140  
 cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgttttcca taggctccgc 13200  
 cccctgacg agcatcaca aatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 13260  
 ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 13320  
 ctgcccgtta ccggatacct gtccgcctt ctcccttcgg gaagcgtggc gctttctcat 13380

## P108872WO rps UCOEs.ST25

agctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg 13440  
 cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 13500  
 aaccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 13560  
 gcgaggatg taggcgggtg tacagagttc ttgaagtggg ggcctaacta cggctacact 13620  
 agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 13680  
 ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtgggttttt tgtttgcaag 13740  
 cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 13800  
 tctgacgctc agtggaacga aaactca 13827

<210> 19  
 <211> 8585  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Vector pRPS11 1005 EGFP

<400> 19  
 cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggcga attgggtacc 60  
 gggccccccc tcgaagttta attcgcctt aagactgttt gcctcatgcc tgccctggcct 120  
 gcccttcctc cgccgccaac taggggaagtg gggaccaaag gttccttagg cactgctcct 180  
 gtgggtagag gggacattag agagctgaca gcgcaccacc tgcatgagtt ttatttaaag 240  
 tgcaaaccat gggatgaatc agttgagctt cagtgttgaa aatgagtagc agggctgccc 300  
 caccacctg accaagtacc ctattctgca gctatgaaaa tgagatctgc acatgagctg 360  
 gggttcacia gtgcacactt ggagcactgc cttgctcctt cccagcagac caciaagcag 420  
 tatttttctg gaggatttta tgtgctaata aattatttga cttagtgtg tacgatgtgt 480  
 gctgtgcaga gaggggcaga gggcaccagc aggtcatctg catggggggc cccttggggt 540  
 gaatccttgc tcacgggata ggctttgttg ctcaaaagtt gcagatatac atcttgggtc 600  
 ctgtcctaga tgggtgttact gtaagtcagc accaagatac aagagctggg acctggactg 660  
 taggaggtca ggccatgaca caaaggctgg gactaaaggc atttaccacg cctgagtctt 720  
 ctggttcttt aaacatcaaa tccttccggg ggctggcgag atggctcagt ggttaagagc 780  
 acagactgct cttacgaagg atccgagttc aaatcccagc aaccaaattg tgccatacaa 840  
 ctatccataa tgaaatctga tgccctcttc tggagtatct gagaacagct acagtgtact 900  
 tacatataat cttaaaaatg cttcccatgt taaccaccac tagagttttt attacagcta 960  
 gctgacctgg aagccaagtc cttatgcctc cgtgagtgtc ggggttaaaa agatccagca 1020  
 ccaactcaaaa tgtcaatcta ttttgaaaat atgctttata ctgttctagc ccatctgtgc 1080  
 agggctagaa cgggtgaatac gagaaactga cacaagcttt tgccacctgg ctaaattggtt 1140  
 cctctattac ctgggggtggg cacctaaggt tagacactca tccacgagta gtcaggacat 1200  
 aaacccatca aagtgtgggt agacgcgcag cctgagatac tgtcaacaaa ggacatgcca 1260

## P108872WO rps UCOEs.ST25

ccttggtgac gtcggccttt aataaaagga agaaagggtg actattcggc cgacgctggc	1320
tgctcctgac atcgtatggc agatactctg ctgtaaagcg gttcaccctt ttcttgagac	1380
ccgctctgca cggccgcttc tctctggaaa ctgaatccca gcacgtgttt cccaaccctg	1440
acggcacgcc ttctccgccc taagcctcgc cgtaccacat gatgcacggt tcctccacat	1500
cgtgctcctg aaatctcgcg agatgatagg atcttcccgc cccttagtcc tccccgca	1560
tggcggcgta cggacagtcc caggaacgcg ggctctcgcg ggaagtacct cccacctccg	1620
tgaggataac cccgcgtcac ttccgccccg acctcgcgtg gtgaataagg aagccgggag	1680
cggccctgcc tctccctttc tccggcggcc ggggaagatgg cggacattca ggttcgagcg	1740
tttagttgct tcccccgac gcttcgggtg ggagcgtatc ccttggcgtc ctcgttgtct	1800
tacgcattag ctgaagcgag gatgcctcgc aatgccttcg tctcaggcgg ctcggaaatc	1860
cgggctctac gcagtaatgg ggtccctggc gcttcgggag ttggttctta aagctcagag	1920
cttaacgggt gagggattgt ggcgggagga gggcatcctg cggcgcggga gtcctgcggc	1980
ggcagagccg gggacactgg gtaaagcagg ttttttccc ttgatggaga ctgaggcccg	2040
gacctcgtgc gctctacggc agggctcgcg tcccacctc gctgtagttt tcagtgtgag	2100
cgcagctctg gcctcgatga gcttaggctt gtcttaaact tgccatcctg cctcaacctc	2160
aaccgggatg acagatccgg cccaccaggc tcggctacgt ggacataagc ttgaatcccg	2220
aatgagtgga tttgtatggt ttggaggctc agtctggctg aaaagctctt tttgatctca	2280
gccgtgagtt ctgcaggctg tggagggtgt agatgggacg cagtgtgtga gctaaactag	2340
acttgggggtg gttggagagc cctgaccagc cggttttggc gattggggca aataagggtg	2400
aaggtaggaa ggaagaaata ttgtctctga tttccttgaa ctttacctgc aacctcacca	2460
aattctcatc cctacagacg gagcgtgctt accaaaagca gcctacgatc tttcaaaaca	2520
agaagcgggt tctgctggga gaaaccggca aggaaaaact ccctcggtag tacaagaata	2580
tcggcttagg cttcaagacg cctaaagagg tacaggacc tccagcagat gagatccctg	2640
ctgccctgca cgtgtgggag cacagccacc ccgcccctt cacagtggct tcccatgggc	2700
ccctgggaat tgtagtatgg gccctgaggc gtcatccttg gttctgttta ggaagtggta	2760
atctaaacc cactttctta actttgcagg ctattgaggg tacctacata gacaagaaat	2820
gccccttcac tggtaacgct tccatccgag gtcggatcct gtctgggtgag tgggatgtg	2880
gaaggggtgt tctaggttcc tgcgtccagg ggcgctggca agtgatgtct gttctcacga	2940
tggctctcag atgtcctcta gggcactgct gagacagcca gttgacaaag ctgatgccat	3000
aaatggagct tcttgggagc cccgttcaac tgactcctac ctgctaacac ctttctgtta	3060
ctctcccagg tgtcgtgacg aagatgaaga tgcagaggac cattgtcatc caagggcgaa	3120
acatttaaat ctagaagctt atcgataccg gtggcgcgcc aattgaatta agatctggcc	3180
caatgggccg tacgaattcg agctcggtag ccggggatcc tgatctaata gtaatcaatt	3240
acggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat	3300

## P108872WO rps UCOEs.ST25

ggccccgctg	gctgaccgcc	caacgacccc	cgcccattga	cgtaataat	gacgtatggt	3360
cccatagtaa	cgccaatagg	gactttccat	tgacgtcaat	gggtggagta	tttacggtaa	3420
actgcccact	tggcagtaca	tcaagtgtat	catatgccaa	gtacgcccc	tattgacgtc	3480
aatgacggta	aatggccccg	ctggcattat	gcccagtaca	tgaccttatg	ggactttcct	3540
acttggcagt	acatctacgt	attagtcatc	gctattacca	tggtgatgcg	gttttggcag	3600
tacatcaatg	ggcgtggata	gcggtttgac	tcacggggat	ttccaagtct	ccaccccatt	3660
gacgtcaatg	ggagtttggt	ttggcaccaa	aatcaacggg	actttccaaa	atgtcgtaac	3720
aactccgccc	cattgacgca	aatgggcggt	aggcgtgtac	gggtgggaggt	ctatataagc	3780
agagctgggt	tagtgaaccg	tcagatccgt	cgccaccatg	gtgagcaagg	gcgaggagct	3840
gttcaccggg	gtggtgccca	tcctggtcga	gctggacggc	gacgtaaacg	gccacaagtt	3900
cagcgtgtcc	ggcgagggcg	agggcgatgc	cacctacggc	aagctgacct	tgaagttcat	3960
ctgcaccacc	ggcaagctgc	ccgtgccctg	gcccaccctc	gtgaccacc	tgacctacgg	4020
cgtgcagtgc	ttcagccgct	accccgacca	catgaagcag	cacgacttct	tcaagtccgc	4080
catgcccga	ggctacgtcc	aggagcgcac	catcttcttc	aaggacgacg	gcaactacaa	4140
gacccgccc	gaggtgaagt	tcgagggcga	cacctgggtg	aaccgcatcg	agctgaaggg	4200
catcgacttc	aaggaggacg	gcaacatcct	ggggcacaag	ctggagtaca	actacaacag	4260
ccacaacgtc	tatatcatgg	ccgacaagca	gaagaacggc	atcaaggtga	acttcaagat	4320
ccgccacaac	atcgaggacg	gcagcgtgca	gctcgccgac	cactaccagc	agaacacccc	4380
catcggcgac	ggccccgtgc	tgctgcccga	caaccactac	ctgagcacct	agtccgccct	4440
gagcaaagac	cccaacgaga	agcgcgatca	catggtcctg	ctggagtctg	tgaccgccgc	4500
cgggatcact	ctcggcatgg	acgagctgta	caagtaaagc	ggccgcgact	ctagatcata	4560
atcagccata	ccacatttgt	agaggtttta	cttgctttta	aaaacctccc	acacctcccc	4620
ctgaacctga	aacataaaat	gaatgcaatt	gttgttgtta	acttgtttat	tgcagcttat	4680
aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttactg	4740
cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	aactagagtc	gacctgcagg	4800
catgcaagct	taccggtggc	gcgcgcgcca	attgttaatt	aagatctggc	ccaatgggcc	4860
gtacgaattc	cttaggctac	cgggtagggg	aggcgctttt	ccaaggcag	tctggagcat	4920
gcgcttttagc	agccccgctg	ggcacttggc	gctacacaag	tggcctctgg	cctcgcacac	4980
attccacatc	caccggccgg	taggcgcaa	ccggctccgt	tctttggtgg	ccccttcgcg	5040
ccaccttcta	ctcctcccct	agtcaggaag	ttccccccg	ccccgcagct	cgcgctgctg	5100
aggacgtgac	aaatggaagt	agcacgtctc	actagtctcg	tgagatgga	cagcaccgct	5160
gagcaatgga	agcgggtagg	cctttggggc	agcggccaat	agcagctttg	ctccttcgct	5220
ttctgggctc	agaggctggg	aaggggtggg	tccggggcg	ggctcagggg	cgggctcagg	5280
ggcggggcgg	gcgcccgaag	gtcctccgga	ggcccggcat	tctgcacgct	tcaaaagcgc	5340

## P108872WO rps UCOEs.ST25

acgtctgccg cgctgttctc ctcttcctca tctccgggcc tttcgaccag cttaccatga 5400  
ccgagtacaa gcccacggtg cgctcgcca cccgcgacga cgtccccagg gccgtacgca 5460  
ccctcgccgc cgcgttcgcc gactaccccg ccacgcgcca caccgtcgat cgggaccgcc 5520  
acatcgagcg ggtcaccgag ctgcaagaac tcttcctcac gcgcgtcggg ctcgacatcg 5580  
gcaagggtgtg ggtcgcggac gacggcgccg cgggtggcggg ctggaccacg ccggagagcg 5640  
tcgaagcggg ggcggtgttc gccgagatcg gcccgcgcat ggccgagttg agcggttccc 5700  
ggctggccgc gcagcaacag atggaaggcc tcctggcgcc gcaccggccc aaggagcccg 5760  
cgtggttccct ggccaccgtc ggcgtctcgc ccgaccacca gggcaagggt ctgggcagcg 5820  
ccgtcgtgct ccccgagtg gaggcggccg agcgcgccgg ggtgcccgcc ttcctggaga 5880  
cctccgcgcc ccgcaacctc cccttctacg agcggctcgg cttcaccgtc accgccgacg 5940  
tcgagtgcc cgaaggaccg cgcacctggt gcatgaccgg caagcccggg gcctgacgcc 6000  
cgccccacga cccgcagcgc ccgaccgaaa ggagcgcacg accccatgca tcgtagacga 6060  
aatgaccgac caagcgacgc ccaacctgcc atcacgagat ttcgattcca ccgccgcctt 6120  
ctatgaaagg ttgggcttcg gaatcgtttt ccgggacgcc ggctggatga tcctccagcg 6180  
cgggatctc atgctggagt tcttcgcca ccctaggggg aggctaactg aaacacggaa 6240  
ggagacaata ccggaaggaa cccgcgctat gacggcaata aaaagacaga ataaaacga 6300  
cgggtgtggg tcgtttgttc ataaacgcgg ggttcgggtcc cagggtggc actctgtcga 6360  
taccaccg agacccatt ggggccaata cgcccgcgtt tcttcctttt cccacccca 6420  
cccccaagt tcgggtgaag gcccagggt cgcagccaac gtcggggcgg caggcccca 6480  
gctttgttc ctttagtga gggtaattt cgagcttggc gtaatcatgg tcatagctgt 6540  
ttcctgtgtg aaattgttat ccgctcacia ttccacacia catacgagcc ggaagcataa 6600  
agtgtaaagc ctggggtgcc taatgagtga gtaactcac attaattgcg ttgcgctcac 6660  
tgcccgttt ccagtcggga aacctgtcgt gccagcatcg cgagcacttt tcggggaaat 6720  
gtgcgcggaa cccctatttg tttatTTTTT taaatacatt caaatatgta tccgctcatg 6780  
agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat gagtattcaa 6840  
catttccgtg tcgcccttat tccctTTTTT gcggcatttt gccttcctgt ttttgctcac 6900  
ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg agtgggttac 6960  
atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga agaagtttt 7020  
ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg tattgacgcc 7080  
gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggg tgagtactca 7140  
ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg cagtgtgcc 7200  
ataaccatga gtgataaac tgccggccaac ttacttctga caacgatcgg aggaccgaag 7260  
gagctaaccg cttttttgca caacatggg gatcatgtaa ctcgccttga tcgttgggaa 7320  
ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc tgtagcaatg 7380

## P108872WO rps UCOEs.ST25

gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc ccggcaacaa 7440  
ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc ggcccttccg 7500  
gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg cggtatcatt 7560  
gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac gacggggagt 7620  
caggcaacta tggatgaacg aatagacag atcgtgaga taggtgcctc actgattaag 7680  
cattggtaac tgtcagactc gcgacactgc attaataaat cggccaacgc gcggggagag 7740  
gcggtttgcg tattgggagc tcttccgctt cctcgtcac tgactcgtg cgctcggctc 7800  
ttcggctgcy gcgagcggta tcagctcact caaaggcggg aatacgggta tccacagaat 7860  
caggggataa cgaggaaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta 7920  
aaaaggccgc gttgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa 7980  
atcgacgctc aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc 8040  
cccctggaag ctccctcgtg cgctctcctg ttccgacctt gccgcttacc ggatacctgt 8100  
ccgcctttct cccttcggga agcgtggcgc tttctcatag ctacgctgt aggtatctca 8160  
gttcggtgta ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccc 8220  
accgctgcy cttatccggt aactatcgtc ttgagtcca cccggtaaga cacgacttat 8280  
cgccactggc agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta 8340  
cagagtctt gaagtgggtg cctaactacg gctacactag aaggacagta tttggtatct 8400  
gcgctctgct gaagccagtt accttcgaa aaagagttgg tagctcttga tccggcaaac 8460  
aaaccaccgc tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa 8520  
aaggatctca agaagatcct ttgatctttt ctacgggggtc tgacgctcag tggaacgaaa 8580  
actca 8585

<210> 20  
<211> 5546  
<212> DNA  
<213> Artificial

<220>  
<223> Vector pCET 1005 EGFP

<400> 20  
cgttgtaaaa cgacggccag tgaattgtaa tacgactcac tatagggcga attgggtacc 60  
gggccccccc tcgaagttta aacatttaaa tctagaagct tatcgatacc ggtggcgcgc 120  
caattgaatt aagatctggc ccaatgggcc gtacgaattc gagctcggta cccggggatc 180  
ctgatctaata agtaataat tacgggggtca ttagttcata gcccatatat ggagttccgc 240  
gttacataac ttacggtaaa tggccccgct ggctgaccgc ccaacgacc cgcgccattg 300  
acgtcaataa tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa 360  
tgggtggagt atttacggta aactgcccac ttggcagtac atcaagtgta tcatatgcca 420  
agtacgcccc ctattgacgt caatgacggt aaatggcccc cctggcatta tgcccagtac 480

## P108872WO rps UCOEs.ST25

atgaccttat	gggactttcc	tacttggcag	tacatctacg	tattagtcat	cgctattacc	540
atggtgatgc	ggttttggca	gtacatcaat	gggcgtggat	agcggtttga	ctcacgggga	600
tttccaagtc	tccaccccat	tgacgtcaat	gggagtttgt	tttggcacca	aatcaacgg	660
gactttccaa	aatgtcgtaa	caactccgcc	ccattgacgc	aatggggcgg	taggcgtgta	720
cggtgggagg	tctatataag	cagagctggg	ttagtgaacc	gtcagatccg	tcgccaccat	780
ggtgagcaag	ggcgaggagc	tgttcaccgg	ggtggtgcc	atcctggtcg	agctggacgg	840
cgacgtaaac	ggccacaagt	tcagcgtgtc	cggcgagggc	gagggcgatg	ccacctacgg	900
caagctgacc	ctgaagttca	tctgcaccac	cggcaagctg	cccgtgccct	ggcccacct	960
cgtgaccacc	ctgacctacg	gcgtgcagtg	cttcagccgc	taccccgacc	acatgaagca	1020
gcacgacttc	ttcaagtccg	ccatgcccga	aggctacgtc	caggagcgca	ccatcttctt	1080
caaggacgac	ggcaactaca	agacccgcgc	cgaggtgaag	ttcgagggcg	acacctggt	1140
gaaccgcatc	gagctgaagg	gcatcgactt	caaggaggac	ggcaacatcc	tggggcacia	1200
gctggagtag	aactacaaca	gccacaacgt	ctatatcatg	gccgacaagc	agaagaacgg	1260
catcaagggtg	aacttcaaga	tccgccacia	catcgaggac	ggcagcgtgc	agctcgccga	1320
ccactaccag	cagaacaccc	ccatcggcga	cggccccgtg	ctgctgcccg	acaaccacta	1380
cctgagcacc	cagtccgcc	tgagcaaaga	ccccaacgag	aagcgcgatc	acatggtcct	1440
gctggagttc	gtgaccgccg	ccgggatcac	tctcggcatg	gacgagctgt	acaagtaaag	1500
cggccgcgac	tctagatcat	aatcagccat	accacatttg	tagaggtttt	acttgcttta	1560
aaaaacctcc	cacacctccc	cctgaacctg	aaacataaaa	tgaatgcaat	tgttgttgtt	1620
aacttgttta	ttgcagctta	taatggttac	aaataaagca	atagcatcac	aaatttcaca	1680
aataaagcat	ttttttcact	gcattctagt	tgtggtttgt	ccaaactcat	caatgtatct	1740
taactagagt	cgacctgcag	gcatgcaagc	ttaccggtgg	cgcgcgcgcc	aattgttaat	1800
taagatctgg	cccaatgggc	cgtacgaatt	ccttaggcta	ccgggtaggg	gaggcgcttt	1860
tcccaaggca	gtctggagca	tgcgctttag	cagccccgct	gggcacttgg	cgctacacia	1920
gtggcctctg	gcctcgcaca	cattccacat	ccaccggccg	gtaggcgcca	accggctccg	1980
ttctttggtg	gccccctcgc	gccaccttct	actcctcccc	tagtcaggaa	gttccccccc	2040
gccccgcagc	tcgcgtcgtg	caggacgtga	caaatggaag	tagcacgtct	cactagtctc	2100
gtgcagatgg	acagcaccgc	tgagcaatgg	aagcgggtag	gcctttgggg	cagcggccaa	2160
tagcagcttt	gctccttcgc	tttctgggct	cagaggctgg	gaaggggtgg	gtccgggggc	2220
gggctcaggg	gcgggctcag	gggcggggcg	ggcggccgaa	ggtcctccgg	aggcccggca	2280
ttctgcacgc	ttcaaaagcg	cacgtctgcc	gcgctgttct	cctcttcctc	atctccgggc	2340
ctttcgacca	gcttaccatg	accgagtaca	agcccacggt	gcgcctcgcc	accgcgacg	2400
acgtccccag	ggccgtacgc	accctcgccg	ccgcgttcgc	cgactacccc	gccacgcgcc	2460
acaccgtcga	tccggaccgc	cacatcgagc	gggtcaccga	gctgcaagaa	ctcttctca	2520

## P108872WO rps UCOEs.ST25

cgcgcgctcg gctcgacatc ggcaaggtgt gggctcgcgga cgacggcgcc gcggtggcgg 2580  
 tctggaccac gccggagagc gtcgaagcgg gggcggtggt cgccgagatc ggcccgcgca 2640  
 tggccgagtt gagcggttcc cggctggccg cgagcaaca gatggaaggc ctccctggcgc 2700  
 cgcaccggcc caaggagccc gcgtggttcc tggccaccgt cggcgtctcg cccgaccacc 2760  
 agggcaaggg tctgggcagc gccgtcgtgc tccccggagt ggaggcggcc gagcgcgccc 2820  
 ggggtgcccgc cttcctggag acctccgcgc cccgcaacct ccccttctac gagcggctcg 2880  
 gcttcaccgt caccgccgac gtcgaggtgc ccgaaggacc gcgcacctgg tgcattgaccc 2940  
 gcaagcccgg tgcctgacgc ccgccccacg acccgagcgc cccgaccgaa aggagcgcac 3000  
 gaccccatgc atcgtagacg aatgaccga ccaagcgacg cccaacctgc catcacgaga 3060  
 tttcgattcc accgccgctt tctatgaaag gttgggcttc ggaatcgttt tccgggacgc 3120  
 cggctggatg atcctccagc gcggggatct catgctggag ttcttcgccc accctagggg 3180  
 gaggctaact gaaacacgga aggagacaat accggaagga acccgcgcta tgacggcaat 3240  
 aaaaagacag aataaacgc acggtggttg gtcgtttgtt cataaacgcg gggttcggtc 3300  
 ccagggtg cactctgtcg ataccccacc gagaccccat tggggccaat acgcccgcgt 3360  
 ttcttccttt tccccacccc acccccgaag ttcgggtgaa ggcccagggc tcgcagccaa 3420  
 cgtcggggcg gcaggcccc agcttttgtt cccttagtg agggttaatt tcgagcttg 3480  
 cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca 3540  
 acatacgagc cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca 3600  
 cattaattgc gttgcgctca ctgccgctt tccagtcggg aaacctgtcg tgccagcatc 3660  
 gcgagcactt ttcggggaaa tgtgcgcgga acccctattt gtttattttt ctaaatacat 3720  
 tcaaataatg atccgctcat gagacaataa ccctgataaa tgcttcaata atattgaaaa 3780  
 aggaagagta tgagtattca acatttccgt gtcgccctta ttcccttttt tgccgcat 3840  
 tgccctcctg tttttgctca ccagaaacg ctggtgaaag taaaagatgc tgaagatcag 3900  
 ttgggtgcac gagggggtta catcgaactg gatctcaaca gcggtgaagat ccttgagagt 3960  
 tttcgccccg aagaacgttt tccaatgatg agcactttta aagttctgct atgtggcgcg 4020  
 gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag 4080  
 aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta 4140  
 agagaattat gcagtgtgc cataaccatg agtgataaca ctgcggccaa cttacttctg 4200  
 acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta 4260  
 actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac 4320  
 accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt 4380  
 actctagctt cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca 4440  
 cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg agccggtgag 4500  
 cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta 4560

## P108872WO rps UCOEs.ST25

gttatctaca	cgacggggag	tcaggcaact	atggatgaac	gaaatagaca	gatcgctgag	4620
ataggtgcct	cactgattaa	gcattggtaa	ctgtcagact	cgcgacactg	cattaatgaa	4680
tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggcg	ctcttccgct	tcctcgctca	4740
ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggc	atcagctcac	tcaaaggcgg	4800
taatacggtt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaggcc	4860
agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	aggctccgcc	4920
cccctgacga	gcatcacaaa	aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	4980
tataaaagata	ccaggcgttt	ccccctggaa	gctccctcgt	gcgctctcct	gttccgaccc	5040
tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	ctttctcata	5100
gctcacgctg	taggtatctc	agttcgggtg	aggcgttcg	ctccaagctg	ggctgtgtgc	5160
acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	cttgagtcca	5220
acccggtaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	attagcagag	5280
cgaggtatgt	aggcgggtgct	acagagtctt	tgaagtggty	gcctaactac	ggctacacta	5340
gaaggacagt	atttggatc	tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	5400
gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	gtttgcaagc	5460
agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	tctacggggt	5520
ctgacgctca	gtggaacgaa	aactca				5546