ENHANCED EXPRESSION SYSTEM AND METHODS OF USE THEREOF

TECHNICAL FIELD

This disclosure relates to an efficient protein expression system that utilizes piggyBac transposons and/or regulatory elements.

BACKGROUND

Many expression systems for recombinant proteins are available and include e.g., bacteria, yeast, fungi, insect, plant and mammalian cells. The protein expression systems differ significantly and may affect both the product formation and subsequent isolation/purification of the protein. Expression systems such as bacteria are incapable of producing specific classes of proteins, which require post-translational modifications such as glycosylation for bioactivity. Furthermore, many therapeutic proteins require complex post-translational modifications such as glycosylation in order to be biologically active.

Mammalian expression systems for producing therapeutic recombinant proteins are commonly used by pharmaceutical companies. Mammalian cells have the ability to carry out proper protein folding and complex post-translational modifications, which are necessary for the therapeutic activity of many proteins. Many mammalian cell expression systems have been approved for use in the production of therapeutic proteins. However, creating a stable mammalian cell expression system for therapeutic recombinant proteins is time-consuming, and the expression level is often not optimized, thus is not ideal for large-scale production.

As such, there is a need for an improved expression system, which can be used as a flexible platform to generate stable cell lines that can express various therapeutic recombinant proteins at an increased level. The ability to produce stable cell lines for expressing recombinant proteins in the same host cells that would likely be used in the manufacturing process in a timely manner would be particularly useful during early stage drug development. There is also a need for an expression system having enhanced transcription and translation efficiencies.

SUMMARY

This disclosure relates to an efficient protein expression system that utilizes piggyBac transposons and/or regulatory elements. This expression system provides a versatile tool that can efficiently integrate a nucleic acid of interest into the genome of a cell and can dramatically increase protein expression in the cell (e.g., Chinese Hamster Ovary cells). In one aspect, the expression system can efficiently move the vector sequence (e.g., a sequence between the two piggyBac ITRs) in the transposon vector into a target genome. The system can be used to create stable cell lines for expressing various proteins (e.g., antibody heavy chains and light chains). In one aspect, the piggyBac expression systems can be used with Chinese Hamster Ovary (CHO) cells to produce recombinant proteins for both research and biopharmaceutical manufacturing purposes.

In one aspect, provided herein is a nucleic acid comprising a 5'- ITR (inverted terminal repeat) sequence; a 3'-ITR sequence; and a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and SEQ ID NOs: 35-58.

In some embodiments, the 5' ITR sequence comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 68, and the 3' ITR sequence comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 60.

In some embodiments, the 5' ITR sequence comprises SEQ ID NO: 68 and the 3' ITR sequence comprises SEQ ID NO: 60.

In some embodiments, the nucleic acid as described herein, further comprising a 5'-internal domain and a 3'-internal domain. In some embodiments, the 5'-internal domain comprise a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66. In some embodiments, the 3'-internal domain comprise a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67. In some embodiments, the 5'-internal domain is immediately adjacent to the 5'-ITR, and the 3'-internal domain is immediately adjacent to the 3'-ITR.

In some embodiments, the nucleic acid comprises one or more regulatory element sequences selected from the group consisting of SEQ ID NOs: 1-15 (e.g., SEQ ID NO: 2).

In some embodiments, the nucleic acid comprises one or more regulatory element sequences selected from the group consisting of SEQ ID NOs: 35-46 (e.g., SEQ ID NO: 36).

In some embodiments, the nucleic acid as described herein, further comprises a promoter and a sequence encoding a polypeptide. In some embodiments, the sequence encoding the polypeptide is operably linked to the promoter. In some embodiments, the sequence encoding a polypeptide is located between two regulatory element sequences.

In some embodiments, the nucleic acid as described herein, further comprises a promoter and a sequence encoding two or more polypeptides. In some embodiments, the sequence encoding the two or more polypeptides is operably linked to the promoter. In some embodiments, the sequence encodes an antibody heavy chain and an antibody light chain.

In some embodiments, the nucleic acid further comprises a WXRE sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46. In some embodiments, the nucleic acid comprises two or more expression cassettes. In some embodiments, the nucleic acid comprises a selection marker. In some embodiments, the selection marker is an antibiotic resistance gene, a sequence encoding a fluorescent protein, or lacZ.

In one aspect, provided herein is a vector comprising the nucleic acid as described herein.

In one aspect, provided herein is a transposon vector comprising from 5' to 3': a 5' ITR sequence consisting of a sequence that is at least 95% identical to SEQ ID NO: 68; a non-transposon heterologous DNA sequence; and a 3'-ITR sequence consisting of a sequence that is at least 95% identical to SEQ ID NO: 60.

In some embodiments, the 5' ITR sequence consists of SEQ ID NO: 68 and the 3' ITR sequence consists of SEQ ID NO: 60.

In some embodiments, the transposon vector as described herein further comprises a 5'-internal domain and a 3'-internal domain. In some embodiments, the 5'-internal domain comprises a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66. In some embodiments, the 3'-internal domain comprises a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67. In some embodiments, the 5'-internal domain is immediately adjacent to the 5'-ITR, and the 3'-internal domain is immediately adjacent to the 3'-ITR.

In some embodiments, the non-transposon heterologous DNA sequence comprises a promoter and a sequence encoding one or more polypeptides. In some embodiments, the sequence encoding one or more polypeptides is operably linked to the promoter. In some embodiments, the promoter is a cytomegalovirus (CMV) promoter.

In some embodiments, the non-transposon heterologous DNA sequence further comprises a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58.

In some embodiments, the non-transposon heterologous DNA sequence further comprises a WXRE sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

In some embodiments, the non-transposon heterologous DNA sequence comprises a multiple cloning site.

In one aspect, provided herein is an expression system comprising: (a) a first nucleic acid comprising a 5'- ITR sequence, a non-transposon heterologous DNA sequence comprising a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58, a 3'-ITR sequence; and (b) a second nucleic acid encoding a piggyBac transposase.

In some embodiments, the second nucleic acid encoding a piggyBac transposase having an amino acid sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 33.

In some embodiments, the non-transposon heterologous DNA sequence comprises a promoter and a sequence encoding one or more polypeptides. In some embodiments, the sequence encoding one or more polypeptides is operably linked to the promoter.

In some embodiments, the sequence encodes an antibody heavy chain and/or an antibody light chain. In some embodiments, the sequence encodes a monoclonal antibody, a bispecific antibody, a recombinant protein, or a fusion protein.

In some embodiments, the promoter is a CMV promoter. In some embodiments, the promoter is an inducible promoter (e.g., a heat shock promoter, a metallothionein promoter, or a glucocorticoid response element).

In some embodiments, the non-transposon heterologous DNA sequence further comprises a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58.

In some embodiments, the non-transposon heterologous DNA sequence further comprises a WXRE regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

In one aspect, provided herein is an isolated nucleic acid comprising a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58.

In some embodiments, the nucleic acid further comprises a promoter and a protein-coding sequence.

In some embodiments, the regulatory element sequence is located between the promoter and the protein-coding sequence.

In some embodiments, the regulatory element sequence is located at the 3' of the protein-coding sequence.

In some embodiments, the regulatory element sequence can be transcribed to a 5'-UTR or a 3'-UTR. In some embodiments, the isolated nucleic acid further comprises a 5'-ITR and a 3'-ITR, and the regulatory element sequence is located between the 5'-ITR and the 3'-ITR.

In one aspect, provided herein is a vector comprising the nucleic acid as described herein.

In one aspect, provided herein is a vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction: a 5'-ITR comprising a TR_L, a 5'-ITR spacer, a IR_L; a promoter; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; a protein-coding sequence; and a 3'-ITR comprising a IR_R, a 3'-ITR spacer, a TR_R.

In one aspect, provided herein is a vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction: a 5'-ITR comprising a TR_L, a 5'-ITR spacer, a IR_L; a promoter; a protein-coding sequence; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; and 3'-ITR comprising a IR_R, a 3'-ITR spacer, a TR_R.

In one aspect, provided herein is a vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction: a 3'-ITR comprising a TR_R, a 3'-ITR spacer, a IR_R; a promoter; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; a protein-coding sequence; and a 5'-ITR comprising a IR_L, a 5'-ITR spacer, a TR_L.

In one aspect, provided herein is a vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction: a 3'-ITR comprising a TR_R, a 3'-ITR spacer, a IR_R; a promoter; a protein-coding sequence; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; and a 5'-ITR comprising a IR_L, a 5'-ITR spacer, a TR_L.

In some embodiments, the 5'-ITR comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 68.

In some embodiments, the 3'-ITR comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 60.

In some embodiments, the TR_L comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 61.

In some embodiments, the 5'-ITR spacer comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 62.

In some embodiments, the IR_L comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 63.

In some embodiments, the TR_R comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 64.

In some embodiments, the IR_R comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 65.

In some embodiments, the vector comprises two or more regulatory element sequences, each of which is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58.

In some embodiments, the vector further comprises a 5'-internal domain sequence and a 3'-internal domain sequence.

In some embodiments, the 5'-internal domain sequence comprises or consists of a sequence that is at least or about 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66 and the 3'-internal domain sequence comprises or consists of a sequence that is at least or about 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67.

In some embodiments, the regulatory element sequence can be transcribed to a 5'-UTR or a 3'-UTR.

In some embodiments, the vector further comprises a sequence encoding a piggyBac transposase. In some embodiments, the sequence encoding the piggyBac transposase is outside a region between the 5'-ITR and the 3'-ITR.

In one aspect, provided herein is a method of generating a cell for expressing a polypeptide of interest, comprising: (a) introducing into a cell: a transposon vector comprising: a 5'- ITR (inverted terminal repeat) sequence; a 3'-ITR sequence; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; and a sequence encoding the polypeptide of interest; and (b) culturing the cell under an appropriate condition. In some embodiments, the piggyBac transposon is integrated into the genome of the cell, thereby generating a cell for expressing the polypeptide of interest.

In some embodiments, the method further comprises introducing a vector comprising a sequence encoding a piggyBac transposase to the cell.

In some embodiments, the transposon vector comprises a sequence encoding a piggyBac transposase.

In some embodiments, the transposon vector is introduced into the cell by microinjection, high velocity propulsion, permeabilization, fusion, or electroporation. In some embodiments, the

cell is a Chinese hamster ovary (CHO) cell. In some embodiments, the cell is a mammalian cell or an insect cell.

In one aspect, provided herein is a cell comprising the nucleic acid as described herein, the vector as described herein, or the expression system as described herein.

In some embodiments, the cell is a Chinese hamster ovary (CHO) cell.

In one aspect, provided herein is a method of expressing a protein, comprising: culturing the cell as described herein under conditions that allow the cell to express the protein; and collecting and purifying the protein.

In one aspect, provided herein is a protein expressed by the cell as described herein, or produced by the method as described herein.

In one aspect, provided herein is a pharmaceutical composition comprising the protein as described herein and a pharmaceutically acceptable carrier.

In one aspect, provided herein is a nucleic acid comprising a 5'- ITR sequence; a 3'-ITR sequence; and one or more regulatory element sequences derived from CHO.

In one aspect, provided herein is an expression system comprising: (a) a first nucleic acid comprising a piggyBac transposon comprising the following genetic elements in a 5' to 3' direction: a first TTAA sequence; a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L; a 5'-internal domain (ID); a sequence of interest; a 3'-ID; a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR_R; and a second TTAA sequence; and (b) a second nucleic acid encoding a piggyBac transposase.

In one aspect, provided herein is a method of generating a cell for expressing a polypeptide of interest, comprising: (a) introducing into a cell: a transposon vector comprising the following genetic elements in a 5' to 3' direction: a first TTAA sequence; a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L; a 5'-internal domain (ID); a sequence of interest; a 3'-ID; a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR_R; and a second TTAA sequence; and (b) culturing the cell under an appropriate condition.

In one aspect, provided herein is a cell line whose genome is stably integrated with a piggyBac transposon comprising the following genetic elements: a 5'- ITR sequence; a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-15; and a 3'-ITR sequence.

In one aspect, provided herein a cell line whose genome is stably integrated with a piggyBac transposon comprising the following genetic elements: a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L; a 5'-internal domain (ID); a sequence of interest; a 3'-ID; a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR.

In one aspect, the disclosure provides a transposon vector comprising a first PB transposase recognition site sequence comprising or consisting of a sequence that is at least 95% identical to SEQ ID NO: 31; a non-transposon heterologous DNA sequence; and a second PB transposase recognition site sequence comprising or consisting of a sequence that is at least 95% identical to SEQ ID NO: 32.

In one aspect, the disclosure is also related to a vector comprising the polynucleotide molecule as described herein. In some embodiments, the vector is a recombinant expression vector.

In some embodiments, provided herein is the vector that further comprises one or more genes encoding one or more proteins.

In some embodiments, the protein is an antibody, a fusion protein, an enzyme, a soluble protein, a membrane protein, a structural protein, a ribosome protein, a zymogen, a cell surface receptor protein, a transcriptional regulatory protein, a translational regulatory protein, a chromatin protein, a hormone, a cell cycle regulatory protein, a G protein, a neuroactive peptide, an immunomodulatory protein, a blood component protein, an ion gate protein, a heat shock protein, dihydrofolate reductase, an antibiotic resistance protein, a functional fragment of any one of the proteins, an epitope fragment of any one of the proteins, and any combination thereof.

In one aspect, the disclosure is related to a recombinant host cell comprising the vector as described herein.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials are described herein for use in the present invention; other, suitable methods and materials known in the art can also be used. The materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, sequences, database entries, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control.

Other features and advantages of the invention will be apparent from the following detailed description and figures, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1A is a schematic diagram showing a piggyBac transposon plasmid.

FIG. 1B is a schematic diagram showing a piggyBac transposase plasmid.

FIG. 2 is a bar graph showing protein expression levels of three antibodies (A, B, and C) in transfected host cells.

FIG. 3 is a bar graph showing protein expression levels among host cells with different regulatory elements.

FIG. 4 is a bar graph showing protein expression levels of three antibodies (D, E, and F) in host cells transfected with piggyBac transposon plasmids with or without regulatory elements.

FIG. 5 shows the sequence of (1) piggyBac 5'-ITR; (2) piggyBac 3'-ITR; and (3) piggyBac transposase amino acid.

FIG. 6 shows the sequence of WXRE IDs: A-L.

FIG. 7 shows the reverse complementary sequence of WXRE IDs: A-L.

FIG. 8 shows the sequence of human CMV promoter and human EF-1α gene intron 1.

FIG. 9 illustrates a schematic diagram of a GFP-expressing vector without WXRE inserted therein.

FIG. 10 illustrates a schematic diagram of a GFP-expressing vector with WXRE inserted therein.

FIG. 11 illustrates the influence on the expression level of the fusion protein after adding transcriptional regulatory elements A~K, wherein A1 and A2 illustrate the forward and reverse directions of transcriptional regulatory element A respectively, and so forth.

FIG. 12 illustrates the influence on the specific productivity of the expression of the fusion protein after adding transcriptional regulatory elements A~K, wherein A1 and A2 illustrate the forward and reverse directions of transcriptional regulatory element A respectively, and so forth.

FIG. 13 illustrates a schematic diagram of a vector which expresses the heavy chain of Adalimumab and has WXRE inserted therein, wherein HC means the heavy chain.

FIG. 14 illustrates a schematic diagram of a vector which expresses the light chain of Adalimumab and has WXRE inserted therein, wherein LC means the light chain.

FIG. 15 illustrates the comparison of the expression level of Adalimumab on 14th day under different combined conditions of the transcriptional regulatory elements, wherein in sample 1 to sample 12, the components of the transcriptional regulatory element in the upstream of the heavy chain and the transcriptional regulatory element in the upstream of the light chain are found in Table 8.

FIG. 16 illustrates the amino acid sequences of the A chain of PD-L1, the heavy chain (HC) of Adalimumab and the light chain (LC) of Adalimumab.

DETAILED DESCRIPTION

DNA transposon is a type of repeat sequence naturally exists in animal genomes. The piggyBac transposon was first discovered in insect genomes, and the sequence contains key elements including transposases and inverted terminal repeat sequences. The piggyBac transposon system can "cut and paste" between the genome and plasmids to replace DNA sequences. As compared to other transposon systems, the piggyBac transposase is capable of precisely replacing the terminal inversely repeat sequences and the sequence located between the repeat sequences into the genomic DNA comprising the "TTAA" nucleotide sequence, without any other modifications or loss of the genomic DNA. The transposon systems have been used in stem cell studies, gene modifications and many cell-engineering areas.

Faster and more efficient protein expression technology is one of the major interests in academia and industry. The transient expression technology can deliver gram-scale recombinant proteins in about two weeks, but with significant compromise in productivity as opposed to classical stable cell pool. On the other hand, the stable cell pool suffers from low possibilities of exogenous DNA integration into the genome, as the integration of exogenous DNA into the genome usually occurs in random fashion. This inefficient DNA integration issue can cause slow recovery during antibiotic selection. While the recombinant protein expression level is fairly high, the time-consuming preparation of stable pool generation limits the use of this classical technology to meet the demand of industry.

Some existing transposon systems can be used to generate stable cell line that can produce recombinant proteins. However, these systems (e.g., the Tol2 transposon) usually take

more than one month to establish an acceptable stable cell line, rendering its limited advantage over classical stable cell pool approach unappealing. Other transposon systems (e.g., Sleeping Beauty, and Mos1) are much less efficient as compared to the piggyBac system for CHO cell lines as described in the present disclosure.

In addition, exogenous protein expression can be regulated by regulatory elements (REs). Many factors that regulate transcription as well as translation can have an impact on protein expression level. This disclosure also provides many DNA regulatory elements that have been identified by RNA abundance analysis. Experiments were further performed to demonstrate that they can improve exogenous protein expression.

The present disclosure provides a highly efficient expression system that utilizes piggyBac transposons and/or regulatory elements. The expression system can create stable cell lines that can express recombinant proteins at an increase level in a relatively short time.

PiggyBac transposons and transposase

DNA-based transposon systems first emerged as efficient tools for genome engineering of mammalian cells after the Sleeping Beauty transposon system was resurrected from the genome of the medaka fish. Transposon DNA vectors can be engineered for a variety of purposes, including transgenesis, gene therapy, gene trapping, or insertion of other DNA elements into the genomes of cells. The piggyBac (PB) transposon system is naturally active and was first discovered in insect cells while propagating Baculo-virus in the TN-386 cell line, from the cabbage looper moth *Trichoplusia ni*.

PiggyBac-based gene transfer or mobilization is carried out through a 'cut and paste' mechanism. When the piggyBac transposase protein is expressed in mammalian cells, it binds to the inverted repeats of the transposon, nicking the DNA and freeing a 3' hydroxyl group at both ends of the transposon. This results in hydrophilic attack of the flanking TTAA sequence and hairpin formation, freeing the transposon from its plasmid backbone. The plasmid backbone is then repaired by host cell factors by ligation of the complementary TTAA overhangs. PiggyBac transposase locates TTAA sequences in the genomic DNA of the mammalian cells. Through hairpin resolution of the transposon and hydrophilic attack of the genomic DNA by 30 hydroxyl groups on the transposon, a staggered four base-pair (bp) cut in the genomic DNA is produced, creating a transient double strand (ds) break with TTAA overhangs on both sides of the break.

The transposon is then inserted into the genomic DNA at the TTAA site, resulting in a duplication of this TTAA, such that a TTAA is found on both sides of the transposon. The sequence will be inserted into the genome, and the sequence will be passed on to all of the progeny cells. Upon excision of the transposon by piggyBac transposase, which can be induced and selected for to rid the cells of the transgene, the single-stranded TTAAs are religated to reform a single TTAA. Thus, the unique mechanism of piggyBac transposition results in a unique advantage: seamless excision of the transposon sequence. After piggyBac excises the transposon from DNA, it seamlessly generates the original piggyBac target site. A detailed description of piggyBac can be found e.g., in Woodard, et al., "piggyBac-ing models and new therapeutic strategies." Trends in biotechnology 33.9 (2015): 525-533; Cary et al. "Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses." Virology 172.1 (1989): 156-169; both of which are incorporated herein by reference in the entirety.

The wildtype piggyBac is a 2472-bp transposon with inverted terminal repeats (ITRs) and a 594-amino acid transposase within ITRs. The PB transposase recognizes the PB 5'-ITR and the PB 3'-ITR. The wildtype 5'-ITR includes the left terminal repeat (TR_L), a 31-bp spacer (5'-ITR spacer), and left internal repeat (IR_L). The sequence of the wildtype 5'-ITR is CCCTAGAAAGATAATCATATTGTGACGTACGTTAAAGATAATCATGCGTAAAATTGA CGCATG (SEQ ID NO: 68). In the wildtype 5'-ITR, the sequence of the TR_L is CCCTAGAAAGATA (SEQ ID NO: 61); the sequence of the 31-bp spacer is ATCATATTGTGACGTACGTTAAAGATAATCA (SEQ ID NO: 62); and the sequence of the IR_L is TGCGTAAAATTGACGCATG (SEQ ID NO: 63).

Similarly, the wildtype 3'-ITR includes the right terminal repeat (TR_R), a short spacer (GAC; the 3'-ITR spacer), and the right internal repeat (IR_R). The sequence of the wildtype 3'-ITR is CATGCGTCAATTTTACGCAGACTATCTTTCTAGGG (SEQ ID NO: 60). In wildtype 3'-ITR, the sequence of the TR_R is TATCTTTCTAGGG (SEQ ID NO: 64); the sequence of the IR_R is CATGCGTCAATTTTACGCA (SEQ ID NO: 65). As shown in **FIG. 5**, the terminal repeats are indicated by single underline, and the internal repeats are indicated by double underlines.

As used herein, the term "5'-ITR" refers to a sequence that is recognized by the piggyBac transposase for the transposon activity, including TR_L, IR_L, and optionally with a spacer

therebetween (e.g., the 31-bp spacer). As used herein, the term "3'-ITR" refers to a sequence that is recognized by the piggyBac transposase for the transposon activity, including TR_R, IR_R, and optionally with a spacer therebetween.

The present disclosure provides a piggyBac expression system. In some embodiments, the piggyBac transposase and piggyBac transposon are carried on two separate plasmids (e.g., the transposon vector and the transposase vector) (FIGS. 1A-1B). It is also possible to deliver the transposase and transposon on the same plasmid (cis) with the transposase gene outside of the transposon inverted terminal repeat elements (ITRs).

In some embodiments, the transposon vector has piggyBac (PB) 5'- and 3'- inverted terminal repeats (ITRs). The 5'-ITR can comprise or consist of e.g., TR_L, and IR_L, and optionally a spacer sequence (e.g., the 31-bp spacer). The 3'-ITR can comprise or consist of e.g., TR_R, and IR_R, and optionally a spacer sequence (e.g., GAC). The sequence of interest (e.g., gene of interest or GOI) can be inserted into the vector between the 5'-ITR and the 3'-ITR or between PB transposase recognition site sequences. In some embodiments, the GOI is operably linked to a promoter. In some embodiments, the GOI is operably linked to one, two, three, four, five, or more regulatory elements. The regulatory element can be located at 5' of GOI or at 3' of GOI. In some embodiments, the regulatory element is between the promoter and GOI or between GOI and a polyA signal sequence. The polyA signal sequence provides the signal for polyadenylation on the transcribed mRNA. In some embodiments, the transposon vector can further comprise a selection marker. The selection marker can be operably linked to the same promoter or a different promoter. In some embodiments, the selection marker and GOI can be under the control of the same promoter. In some embodiments, the selection marker can have its own promoter. Usually, the host cell's genome does not itself provide a selection marker functionality. Thus, the cells with the correct modification can be screened for the selection marker.

In some embodiments, the sequence of interest comprises various genetic elements, e.g., restriction sites, loxP sites, regulatory elements, promoters, enhancers, expression cassettes, genetic operons, etc. In some embodiments, the sequence of interest does not have any protein coding sequences.

In some embodiments, a PB transposase vector is provided. The PB transposase vector is designed to express the PB transposase. During the experiments, host cells (e.g., CHO-K1 cells) are transfected with both vectors. The PB transposase expressed from the PB transposase vector

recognizes the PB 5'-ITR and the PB 3'-ITR located on the transposon vector and efficiently moves and integrates the nucleic acid sequence between the PB 5'- and 3'- ITRs (including both the PB 5'- and 3'- ITRs) into a chromosomal TTAA site in the cells. In some embodiments, the cells are then selected by the selection marker (*e.g.*, an antibiotic resistance gene) and protein expression activity.

In some embodiments, the 5'- and 3'-internal domains can be used in connection with 5'-ITR and 3'-ITR to increase the integration efficiency of the PB transposon (e.g., by at least 10%, 20%, 30%, 40%, 50% as compared a PB transposon without the 5'- or 3'-internal domain sequences). In some embodiments, the PB transposon recognition site comprises a 5'-internal domain sequence that is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the 5'-internal domain sequence as shown in FIG. 5. The 5'-internal domain sequence in FIG. 5 is: TGTTTTATCGGTCTGTATATCGAGGTTTATTTATTAATTTGAATAGATATTAAGTTTT ATTATATTTACACTTACATACTAATAATAAATTCAACAAACAATTTATTTATGTTTAT TTATTTATTAAAAAAAAACAAAAACTCAAAATTTCTTCTATAAAGTAACAAAACT (SEQ ID NO: 66).

In some embodiments, the PB transposon recognition site comprises a 3'-internal domain sequence. The 3'-internal domain sequence can be at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the 3'-internal domain sequence as shown in FIG. 5. The 3'-internal domain sequence in FIG. 5 is: TATCTATAACAAGAAAATATATATATAATAAGTTATCACGTAAGTAGAACATGAAAT AACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAATCATGC GTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGA CAAGCACGCCTCACGGGAGCTCCAAGCGGCGACTGAGATGTCCTAAATGCACAGCG ACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAATG (SEQ ID NO: 67).

In some embodiments, the PB transposon comprises a 5'-internal domain sequence that has at least or about 50, 100, 110, 120, 130, 140, 150, 160, 170 or 172 contiguous nucleotides of SEQ ID NO: 66. In some embodiments, the PB transposon comprises a 3'-internal domain sequence that has at least or about 50, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, or 272 contiguous nucleotides of SEQ ID NO: 67.

In some embodiments, the PB transposon does not have 5'-internal domain sequence. In some embodiments, the PB transposon does not have 3'-internal domain sequence.

As used herein, the term "PB transposase 5'-recognition site" refers one of the pair of the two sites that the PB transposase recognizes and interacts with, which is located on the 5'-direction of a nucleic acid (e.g., based on the 5' to 3' direction on the sense strand of the coding sequence or a reference sequence). The term "PB transposase 3'-recognition site" refers one of the pair of the two sites that the PB transposase recognizes and interacts with, which is located on the 3'- direction of a nucleic acid (e.g., based on the 5' to 3' direction on the sense strand of the coding sequence or a reference sequence). In some embodiments, the PB transposase 5'-recognition site sequence comprises or consists of TR_L. In some embodiments, the PB transposase 3'-recognition site sequence comprises or consists of TR_R. In some embodiments, the PB transposase 5'-recognition site comprises or consists of TR_L and IR_L, and optionally with a spacer and/or the internal domain sequence. In some embodiments, the PB transposase 3'-recognition site comprises or consists of TR_R and IR_R, and optionally with a spacer and/or the internal domain sequence.

The location for 5'-ITR and 3'-ITR can be exchanged on the transposon vector. Thus, in some embodiments, the PB transposase 3'-recognition site sequence comprises or consists of TR_L. In some embodiments, the PB transposase 5'-recognition site sequence comprises or consists of TR_R. In some embodiments, the PB transposase 3'-recognition site comprises or consists of TR_L and IR_L, and optionally with a spacer and/or the internal domain sequence. In some embodiments, the PB transposase 5'-recognition site comprises or consists of TR_R and IR_R, and optionally with a spacer and/or the internal domain sequence.

In some embodiments, the PB transposase 5'-recognition site comprises or consists of a sequence that is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 31. In some embodiments, the PB transposase 3'-recognition site comprises or consists of a sequence that is at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 32. In some embodiments, the PB transposase 5'-recognition site comprises or consists of a sequence that has at least or about 50, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, or 230 contiguous nucleotides or the entire sequence of SEQ ID NO: 31. In some embodiments, the PB transposase 3'-recognition site comprises or consists of a sequence that has at least or about 50, 100, 110, 120,

130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, or 310 contiguous nucleotides or the entire sequence of SEQ ID NO: 32.

In some embodiments, the length of the sequence between ITRs or PB transposase recognition sites is at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 kb. In some embodiments, the transposon comprises an insert of ranging between 1.5-3kb, 1.5-5kb, 1.5-10kb, 1.5-20kb, 1.5- 30kb, 1.5-50kb, 1.5-75kb, 2-5kb, 2- 10kb, 2-20kb, 2-30kb, 2-50kb, 2-75kb, 3-5kb, 3-10kb, 3- 20kb, 3-30kb, 3-50kb, 3-75kb, 5-10kb, 5-20kb, 5-30kb, 5-50kb, 5-75kb, 10-20kb, 10-30kb, 10- 50kb, or 10-75kb.

The present disclosure also provides a vector for making a transposon vector containing a gene of interest or a sequence of interest. In some embodiments, the vector comprises a 5'-ITR and a 3'-ITR as described herein, and a linker sequence between 5'-ITR and 3'-ITR. In some embodiments, the vector comprises a PB transposase 5'-recognition site and a PB transposase 3'-recognition site as described herein, and a linker sequence between the two recognition sites. In some embodiments, the linker sequence is a short length of DNA (e.g., less than 20, 30, 40, 50, 60, 70, 80, 90, 100 or 200 nucleotides) that contains numerous different endonuclease restrictions sites located in close proximity. The presence of the linker sequence is advantageous because it allows various exogenous sequences, such as expression cassettes, to be easily inserted and removed, thus simplifying the process of making a vector containing a particular targeted DNA fragment. As used herein, an expression cassette is a distinct component of a vector or a sequence having a gene to be expressed by a transfected cell and regulatory sequences for the gene. In some embodiments, the transposon vector can have various regulatory elements or genetic elements, such as promoters (e.g., inducible promoters), enhancers, or insulators, and the like.

When this transposon vector is introduced into a host cell, in the presence of transposase activity specific for the flanking inverted repeats, the targeted DNA sequence will be excised from the introduced vector and will be inserted into a location in the genome. Transposition of the targeted DNA is facilitated in the presence of transposase activity. The gene encoding the transposase can either be physically linked to the transposon vector, already present in the host cell's genome, or introduced into the cell as part of a separate vector. In some embodiments, inducible promoters can be used as a means of triggering the production or transposase activity.

Regulatory elements

The present disclosure also provides various regulatory elements. As used herein, the term "regulatory element" in the present disclosure refers to a sequence that is involved in the regulation of gene transcription and/or translation. In some embodiments, the regulatory element is a transcription regulatory element or a translation regulatory element. In some embodiments, the regulatory element can stabilize mRNA. In some embodiments, these regulatory elements are derived from CHO-K1 cells. In some embodiments, these regulatory elements can increase expression level of gene of interest.

The present disclosure demonstrates that the regulatory elements as listed in Table 1 can increase the expression level of the gene of interest. Without wishing to be bound by theory, it has been hypothesized that these regulatory elements can increase the transcription efficiency and stabilize the transcribed mRNA. In some embodiments, these regulatory elements can make the mRNA resistant to degradation.

SEQ ID NOs: 1-15 are the sequences for the regulatory elements. The reverse complementary sequences are provided in SEQ ID NOs: 16-30. In one aspect, the disclosure provides an isolated polynucleotide molecule comprising (i) a sequence of any of SEQ ID NOs: 1-15; (ii) a reverse complementary sequence of the sequence of any of SEQ ID NOs: 1-15; (iii) a reverse complementary sequence of a sequence capable of hybridizing with the sequence of (i) or (ii) under a high stringency hybridization condition or a very high stringency hybridization condition; and (iv) a sequence having at least or about 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to SEQ ID NOs: 1-30. As used herein, the term "reverse complementary sequence" in the present disclosure is a sequence which is opposite to the direction of the sequence of the original polynucleotide and is also complementary to the sequence of the original polynucleotide. Illustratively, if the original polynucleotide sequence is ACTGAAC, then the reverse complementary sequence thereof is GTTCAGT.

In one aspect, the disclosure also provides a sequence comprising a promoter and a gene of interest. The regulatory element sequence can be located at the 5' of the gene of interest (e.g., between the promoter and the gene of interest) or at the 3' of the gene of interest (e.g., between the gene of interest and a polyA signal sequence).

In some embodiments, the regulatory element has a sequence identity of at least or about 80%, 81 %, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,

96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) with the sequence of any of SEQ ID NOs: 1-30. In some embodiments, the sequence differs from a sequence selected from SEQ ID NOs: 1-30 by at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In some embodiments, the sequence differs from a sequence selected from SEQ ID NOs: 1-30 by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides.

The sequence can have a forward direction or a reverse direction. In some embodiments, the regulatory element sequence can increase the expression amount of a heterologous protein by about or at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% (e.g., as compared to a control sequence without the regulatory element sequence).

The regulatory element sequence can be located after the promotor (e.g., from 5' to 3' on the sense strand of the coding sequence) or after the polynucleotide encoding a polypeptide (e.g., from 5' to 3' on the sense strand of the coding sequence). In some embodiments, the regulatory element is located before the polynucleotide encoding a polypeptide (e.g., transcribed to a 5'-untranslated region (5'-UTR)). In some embodiments, there are at least or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides between the regulatory element sequence and the promoter or between the regulatory element sequence and the polynucleotide encoding a polypeptide. In some embodiments, there are no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, or l000 nucleotides between the regulatory element sequence and the promoter or between the regulatory element sequence and the polynucleotide encoding a polypeptide.

In some embodiments, the regulatory element is located after (e.g., immediately after) the polynucleotide encoding a polypeptide (e.g., transcribed to a 3'-UTR). In some embodiments, there are at least or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides between the regulatory element sequence and the end of the sequence encoding the polypeptide or between the regulatory element sequence and the polyA signal sequence. In some embodiments, there are no more than 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides between the regulatory element sequence and the end of the sequence encoding the polypeptide or between the regulatory element sequence and the polyA signal sequence.

The present disclosure also provides methods of screening regulatory element sequence. In some embodiments, RNA sequencing (RNA-seq) can be used to sequence and quantify

mRNAs in cells (e.g., CHO cells). Total RNA is extracted. In some embodiments, cDNA is generated from the extracted RNA. The amount of RNA is ranked by abundance. And the sequence at the untranslated regions of the top ranked RNA is selected. In some embodiments, experiments as described herein are performed to verify the effects of the regulatory elements on protein expression. In one aspect, mRNA of a desired host cell (e.g., CHO-K1 cell) at different stages are sequenced and quantified by a high throughput sequencing method (e.g., RNA-seq). In some embodiments, transient transfection as well as stable transfection can be performed. Total RNA can be extracted from an appropriate number of (e.g., at least or about 10, 20, or 30) samples after an appropriate time (e.g., on day 6, day 8, day 10, day 12, or day 14) by transient transfection, or from an appropriate number (e.g. at least or about 10, 20, or 30) of stable transfection samples (e.g., stable protein expressing cell lines) after an appropriate time (e.g., on day 6, day 8, day 10, day 12, or day 14) of a traditional fed-batch process (e.g., 14-day fed-batch process).

In some embodiments, cDNA can be generated accordingly (e.g., by reverse transcription) and used for sequencing (e.g., high throughput sequencing). With sequencing data and the relative reads number, mRNA can be extracted and ranked by average abundance across all samples. In some embodiments, regulatory element (RE) sequences can be extracted from the untranslated regions of top ranked mRNAs (e.g., at least or about the top 5, 10, 15, 20, 25, 30, 35, 40, or more).

In some embodiments, the regulatory element sequences as described herein can be incorporated into a fusion protein expression plasmid (e.g., immediately before, or immediately after the recombinant protein expressing gene). In some embodiments, a control sample that does not contain any regulatory element sequences can be used to determine the effects of the regulatory elements.

WXRE regulatory elements

A regulatory element can regulate gene transcription and/or translation. In some embodiments, the regulatory element is a transcription regulatory element, e.g., a WXRE regulatory element. As shown in FIG. 6, the WXREs in the present disclosure include transcription regulatory element A (SEQ ID NO: 35), transcription regulatory element B (SEQ ID NO: 36), transcription regulatory element C (SEQ ID NO: 37), transcription regulatory

element D (SEQ ID NO: 38), transcription regulatory element E (SEQ ID NO: 39), transcription regulatory element F (SEQ ID NO: 40), transcription regulatory element G (SEQ ID NO: 41), transcription regulatory element H (SEQ ID NO: 42), transcription regulatory element I (SEQ ID NO: 43), transcription regulatory element J (SEQ ID NO: 44), transcription regulatory element K (SEQ ID NO: 45), and transcription regulatory element L (SEQ ID NO: 46). Accordingly, as shown in FIG. 7, the reverse complementary sequence of the WXREs in the present disclosure include the reverse complementary sequence of transcription regulatory element A (SEQ ID NO: 47), the reverse complementary sequence of transcription regulatory element B (SEQ ID NO: 48), the reverse complementary sequence of transcription regulatory element C (SEQ ID NO: 49), the reverse complementary sequence of transcription regulatory element D (SEQ ID NO: 50), the reverse complementary sequence of transcription regulatory element E (SEQ ID NO: 51), the reverse complementary sequence of transcription regulatory element F (SEQ ID NO: 52), the reverse complementary sequence of transcription regulatory element G (SEQ ID NO: 53), the reverse complementary sequence of transcription regulatory element H (SEQ ID NO: 54), the reverse complementary sequence of transcription regulatory element I (SEQ ID NO: 55), the reverse complementary sequence of transcription regulatory element J (SEQ ID NO: 56), the reverse complementary sequence of transcription regulatory element K (SEQ ID NO: 57), and the reverse complementary sequence of transcription regulatory element L (SEQ ID NO: 58).

In some embodiments, the WXRE sequence has a sequence identity of at least or about 80%, 81 %, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) with the sequence of any of SEQ ID NOs: 35-58. In some embodiments, the WXRE sequence differs from a sequence selected from SEQ ID NOs: 35-58 by at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In some embodiments, the WXRE sequence differs from a sequence selected from SEQ ID NOs: 35-58 by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides.

The WXRE sequence can have a forward direction or a reverse direction. As used herein, a sequence of interest has a forward direction when the sense strand (from 5' to 3') has a sequence that is identical to the sequence of interest. A sequence of interest has a reverse direction when the sense strand has a sequence that is reverse complementary to the sequence of

interest. The sequences that are reverse complementary to SEQ ID NOs: 35-46 are set forth in SEQ ID NO: 47-58, respectively.

In some embodiments, the WXRE sequence can increase the expression amount of a heterologous protein by about or at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80% (e.g., as compared to a control sequence without the WXRE sequence).

In some embodiments, the WXRE sequence, the promotor, and the polynucleotide encoding a polypeptide are operably linked together. In some embodiments, the WXRE sequence, the promotor, the polynucleotide encoding a polypeptide, and one or more additional regulatory elements are operably linked together. In some embodiments, the additional regulatory elements are regulatory elements as described herein (e.g., SEQ ID NOs: 1-30). In some embodiments, an intron of EF-1α (e.g., the first intron of human EF-1α) can be used to increase the expression. The WXRE sequence, the promotor, and the polynucleotide encoding a polypeptide that are operably linked together can have various orders. For example, the WXRE sequence can be located before the promotor (e.g., from 5' to 3' on the sense strand of the coding sequence) or after the polynucleotide encoding a polypeptide (e.g., from 5' to 3' on the sense strand of the coding sequence). In some embodiments, there are at least or about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000 or 5000 nucleotides between the WXRE sequence and the promoter or between the WXRE sequence and the polynucleotide encoding a polypeptide. In some embodiments, there are no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 100, 200, 300, 400, 500, 600, 700, 800, 900, l000, 2000, 3000, 4000 or 5000 nucleotides between the WXRE sequence and the promoter or between the WXRE sequence and the polynucleotide encoding a polypeptide. In some embodiments, the one or more additional regulatory elements are located between the promoter and the sequence encoding a polypeptide.

In some embodiments, the use of the transcription regulatory element (WXRE) listed in the present disclosure can enable a heterologous protein to still maintain its biological activity while the expression level is greatly increased (e.g., by at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 folds).

In some embodiments, the transcription regulatory element (WXRE) listed in the present disclosure can be used together with other regulatory elements as a whole, and maintains its biological activity while enabling the expression level of the heterologous protein to increase greatly.

Thus, in one aspect, the disclosure is related to an isolated polynucleotide molecule comprising a nucleotide sequence selected from the group consisting of (i) to (iv): (i) a sequence of any of SEQ ID NOs: 35-46; (ii) a reverse complementary sequence of the sequence of any of SEQ ID NOs: 35-46; (iii) a reverse complementary sequence of a sequence capable of hybridizing with the sequence of (i) or (ii) under a high stringency hybridization condition or a very high stringency hybridization condition; and (iv) a sequence having at least 80% sequence identity, or at least 90% sequence identity, alternatively at least 95% sequence identity, preferably at least 97% sequence identity, more preferably at least 98% sequence identity, most preferably at least 99% sequence identity with the sequence of (i) or (ii).

Various WXRE sequences and the methods of use thereof are described e.g., in WO2020/034097A1 and WO2020/034986A1, which are incorporated herein by reference in the entirety.

In some embodiments of the present disclosure, the WXRE sequence can be inserted into two vectors. They can be the same or different, and may have forward direction or reverse direction. The exemplary combinations of WXREs in two vectors are listed in Table 2.

Table 2 Exemplary combinations of WXREs in two vectors

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
1	SEQ ID NO: 35	SEQ ID NO: 35	128	SEQ ID NO: 41	SEQ ID NO: 44
2	SEQ ID NO: 35	SEQ ID NO: 36	129	SEQ ID NO: 41	SEQ ID NO: 45
3	SEQ ID NO: 35	SEQ ID NO: 37	130	SEQ ID NO: 41	SEQ ID NO: 54
4	SEQ ID NO: 35	SEQ ID NO: 38	131	SEQ ID NO: 41	SEQ ID NO: 55
5	SEQ ID NO: 35	SEQ ID NO: 39	132	SEQ ID NO: 41	SEQ ID NO: 56
6	SEQ ID NO: 35	SEQ ID NO: 40	133	SEQ ID NO: 41	SEQ ID NO: 57
7	SEQ ID NO: 35	SEQ ID NO: 41	134	SEQ ID NO: 47	SEQ ID NO: 47
8	SEQ ID NO: 35	SEQ ID NO: 47	135	SEQ ID NO: 47	SEQ ID NO: 48
9	SEQ ID NO: 35	SEQ ID NO: 48	136	SEQ ID NO: 47	SEQ ID NO: 49
10	SEQ ID NO: 35	SEQ ID NO: 49	137	SEQ ID NO: 47	SEQ ID NO: 50
11	SEQ ID NO: 35	SEQ ID NO: 50	138	SEQ ID NO: 47	SEQ ID NO: 51
12	SEQ ID NO: 35	SEQ ID NO: 51	139	SEQ ID NO: 47	SEQ ID NO: 52
13	SEQ ID NO: 35	SEQ ID NO: 52	140	SEQ ID NO: 47	SEQ ID NO: 53
14	SEQ ID NO: 35	SEQ ID NO: 53	141	SEQ ID NO: 47	SEQ ID NO: 42
15	SEQ ID NO: 35	SEQ ID NO: 42	142	SEQ ID NO: 47	SEQ ID NO: 43
16	SEQ ID NO: 35	SEQ ID NO: 43	143	SEQ ID NO: 47	SEQ ID NO: 44

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
17	SEQ ID NO: 35	SEQ ID NO: 44	144	SEQ ID NO: 47	SEQ ID NO: 45
18	SEQ ID NO: 35	SEQ ID NO: 45	145	SEQ ID NO: 47	SEQ ID NO: 54
19	SEQ ID NO: 35	SEQ ID NO: 54	146	SEQ ID NO: 47	SEQ ID NO: 55
20	SEQ ID NO: 35	SEQ ID NO: 55	147	SEQ ID NO: 47	SEQ ID NO: 56
21	SEQ ID NO: 35	SEQ ID NO: 56	148	SEQ ID NO: 47	SEQ ID NO: 57
22	SEQ ID NO: 35	SEQ ID NO: 57	149	SEQ ID NO: 48	SEQ ID NO: 48
23	SEQ ID NO: 36	SEQ ID NO: 36	150	SEQ ID NO: 48	SEQ ID NO: 49
24	SEQ ID NO: 36	SEQ ID NO: 37	151	SEQ ID NO: 48	SEQ ID NO: 50
25	SEQ ID NO: 36	SEQ ID NO: 38	152	SEQ ID NO: 48	SEQ ID NO: 51
26	SEQ ID NO: 36	SEQ ID NO: 39	153	SEQ ID NO: 48	SEQ ID NO: 52
27	SEQ ID NO: 36	SEQ ID NO: 40	154	SEQ ID NO: 48	SEQ ID NO: 53
28	SEQ ID NO: 36	SEQ ID NO: 41	155	SEQ ID NO: 48	SEQ ID NO: 42
29	SEQ ID NO: 36	SEQ ID NO: 47	156	SEQ ID NO: 48	SEQ ID NO: 43
30	SEQ ID NO: 36	SEQ ID NO: 48	157	SEQ ID NO: 48	SEQ ID NO: 44
31	SEQ ID NO: 36	SEQ ID NO: 49	158	SEQ ID NO: 48	SEQ ID NO: 45
32	SEQ ID NO: 36	SEQ ID NO: 50	159	SEQ ID NO: 48	SEQ ID NO: 54
33	SEQ ID NO: 36	SEQ ID NO: 51	160	SEQ ID NO: 48	SEQ ID NO: 55
34	SEQ ID NO: 36	SEQ ID NO: 52	161	SEQ ID NO: 48	SEQ ID NO: 56
35	SEQ ID NO: 36	SEQ ID NO: 53	162	SEQ ID NO: 48	SEQ ID NO: 57
36	SEQ ID NO: 36	SEQ ID NO: 42	163	SEQ ID NO: 49	SEQ ID NO: 49
37	SEQ ID NO: 36	SEQ ID NO: 43	164	SEQ ID NO: 49	SEQ ID NO: 50
38	SEQ ID NO: 36	SEQ ID NO: 44	165	SEQ ID NO: 49	SEQ ID NO: 51
39	SEQ ID NO: 36	SEQ ID NO: 45	166	SEQ ID NO: 49	SEQ ID NO: 52
40	SEQ ID NO: 36	SEQ ID NO: 54	167	SEQ ID NO: 49	SEQ ID NO: 53
41	SEQ ID NO: 36	SEQ ID NO: 55	168	SEQ ID NO: 49	SEQ ID NO: 42
42	SEQ ID NO: 36	SEQ ID NO: 56	169	SEQ ID NO: 49	SEQ ID NO: 43
43	SEQ ID NO: 36	SEQ ID NO: 57	170	SEQ ID NO: 49	SEQ ID NO: 44
44	SEQ ID NO: 37	SEQ ID NO: 37	171	SEQ ID NO: 49	SEQ ID NO: 45
45	SEQ ID NO: 37	SEQ ID NO: 38	172	SEQ ID NO: 49	SEQ ID NO: 54
46	SEQ ID NO: 37	SEQ ID NO: 39	173	SEQ ID NO: 49	SEQ ID NO: 55
47	SEQ ID NO: 37	SEQ ID NO: 40	174	SEQ ID NO: 49	SEQ ID NO: 56
48	SEQ ID NO: 37	SEQ ID NO: 41	175	SEQ ID NO: 49	SEQ ID NO: 57
49	SEQ ID NO: 37	SEQ ID NO: 47	176	SEQ ID NO: 50	SEQ ID NO: 50
50	SEQ ID NO: 37	SEQ ID NO: 48	177	SEQ ID NO: 50	SEQ ID NO: 51
51	SEQ ID NO: 37	SEQ ID NO: 49	178	SEQ ID NO: 50	SEQ ID NO: 52
52	SEQ ID NO: 37	SEQ ID NO: 50	179	SEQ ID NO: 50	SEQ ID NO: 53

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
53	SEQ ID NO: 37	SEQ ID NO: 51	180	SEQ ID NO: 50	SEQ ID NO: 42
54	SEQ ID NO: 37	SEQ ID NO: 52	181	SEQ ID NO: 50	SEQ ID NO: 43
55	SEQ ID NO: 37	SEQ ID NO: 53	182	SEQ ID NO: 50	SEQ ID NO: 44
56	SEQ ID NO: 37	SEQ ID NO: 42	183	SEQ ID NO: 50	SEQ ID NO: 45
57	SEQ ID NO: 37	SEQ ID NO: 43	184	SEQ ID NO: 50	SEQ ID NO: 54
58	SEQ ID NO: 37	SEQ ID NO: 44	185	SEQ ID NO: 50	SEQ ID NO: 55
59	SEQ ID NO: 37	SEQ ID NO: 45	186	SEQ ID NO: 50	SEQ ID NO: 56
60	SEQ ID NO: 37	SEQ ID NO: 54	187	SEQ ID NO: 50	SEQ ID NO: 57
61	SEQ ID NO: 37	SEQ ID NO: 55	188	SEQ ID NO: 51	SEQ ID NO: 51
62	SEQ ID NO: 37	SEQ ID NO: 56	189	SEQ ID NO: 51	SEQ ID NO: 52
63	SEQ ID NO: 37	SEQ ID NO: 57	190	SEQ ID NO: 51	SEQ ID NO: 53
64	SEQ ID NO: 38	SEQ ID NO: 38	191	SEQ ID NO: 51	SEQ ID NO: 42
65	SEQ ID NO: 38	SEQ ID NO: 39	192	SEQ ID NO: 51	SEQ ID NO: 43
66	SEQ ID NO: 38	SEQ ID NO: 40	193	SEQ ID NO: 51	SEQ ID NO: 44
67	SEQ ID NO: 38	SEQ ID NO: 41	194	SEQ ID NO: 51	SEQ ID NO: 45
68	SEQ ID NO: 38	SEQ ID NO: 47	195	SEQ ID NO: 51	SEQ ID NO: 54
69	SEQ ID NO: 38	SEQ ID NO: 48	196	SEQ ID NO: 51	SEQ ID NO: 55
70	SEQ ID NO: 38	SEQ ID NO: 49	197	SEQ ID NO: 51	SEQ ID NO: 56
71	SEQ ID NO: 38	SEQ ID NO: 50	198	SEQ ID NO: 51	SEQ ID NO: 57
72	SEQ ID NO: 38	SEQ ID NO: 51	199	SEQ ID NO: 52	SEQ ID NO: 52
73	SEQ ID NO: 38	SEQ ID NO: 52	200	SEQ ID NO: 52	SEQ ID NO: 53
74	SEQ ID NO: 38	SEQ ID NO: 53	201	SEQ ID NO: 52	SEQ ID NO: 42
75	SEQ ID NO: 38	SEQ ID NO: 42	202	SEQ ID NO: 52	SEQ ID NO: 43
76	SEQ ID NO: 38	SEQ ID NO: 43	203	SEQ ID NO: 52	SEQ ID NO: 44
77	SEQ ID NO: 38	SEQ ID NO: 44	204	SEQ ID NO: 52	SEQ ID NO: 45
78	SEQ ID NO: 38	SEQ ID NO: 45	205	SEQ ID NO: 52	SEQ ID NO: 54
79	SEQ ID NO: 38	SEQ ID NO: 54	206	SEQ ID NO: 52	SEQ ID NO: 55
80	SEQ ID NO: 38	SEQ ID NO: 55	207	SEQ ID NO: 52	SEQ ID NO: 56
81	SEQ ID NO: 38	SEQ ID NO: 56	208	SEQ ID NO: 52	SEQ ID NO: 57
82	SEQ ID NO: 38	SEQ ID NO: 57	209	SEQ ID NO: 53	SEQ ID NO: 53
83	SEQ ID NO: 39	SEQ ID NO: 39	210	SEQ ID NO: 53	SEQ ID NO: 42
84	SEQ ID NO: 39	SEQ ID NO: 40	211	SEQ ID NO: 53	SEQ ID NO: 43
85	SEQ ID NO: 39	SEQ ID NO: 41	212	SEQ ID NO: 53	SEQ ID NO: 44
86	SEQ ID NO: 39	SEQ ID NO: 47	213	SEQ ID NO: 53	SEQ ID NO: 45
87	SEQ ID NO: 39	SEQ ID NO: 48	214	SEQ ID NO: 53	SEQ ID NO: 54
88	SEQ ID NO: 39	SEQ ID NO: 49	215	SEQ ID NO: 53	SEQ ID NO: 55

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
89	SEQ ID NO: 39	SEQ ID NO: 50	216	SEQ ID NO: 53	SEQ ID NO: 56
90	SEQ ID NO: 39	SEQ ID NO: 51	217	SEQ ID NO: 53	SEQ ID NO: 57
91	SEQ ID NO: 39	SEQ ID NO: 52	218	SEQ ID NO: 42	SEQ ID NO: 42
92	SEQ ID NO: 39	SEQ ID NO: 53	219	SEQ ID NO: 42	SEQ ID NO: 43
93	SEQ ID NO: 39	SEQ ID NO: 42	220	SEQ ID NO: 42	SEQ ID NO: 44
94	SEQ ID NO: 39	SEQ ID NO: 43	221	SEQ ID NO: 42	SEQ ID NO: 45
95	SEQ ID NO: 39	SEQ ID NO: 44	222	SEQ ID NO: 42	SEQ ID NO: 54
96	SEQ ID NO: 39	SEQ ID NO: 45	223	SEQ ID NO: 42	SEQ ID NO: 55
97	SEQ ID NO: 39	SEQ ID NO: 54	224	SEQ ID NO: 42	SEQ ID NO: 56
98	SEQ ID NO: 39	SEQ ID NO: 55	225	SEQ ID NO: 42	SEQ ID NO: 57
99	SEQ ID NO: 39	SEQ ID NO: 56	226	SEQ ID NO: 43	SEQ ID NO: 43
100	SEQ ID NO: 39	SEQ ID NO: 57	227	SEQ ID NO: 43	SEQ ID NO: 44
101	SEQ ID NO: 40	SEQ ID NO: 40	228	SEQ ID NO: 43	SEQ ID NO: 45
102	SEQ ID NO: 40	SEQ ID NO: 41	229	SEQ ID NO: 43	SEQ ID NO: 54
103	SEQ ID NO: 40	SEQ ID NO: 47	230	SEQ ID NO: 43	SEQ ID NO: 55
104	SEQ ID NO: 40	SEQ ID NO: 48	231	SEQ ID NO: 43	SEQ ID NO: 56
105	SEQ ID NO: 40	SEQ ID NO: 49	232	SEQ ID NO: 43	SEQ ID NO: 57
106	SEQ ID NO: 40	SEQ ID NO: 50	233	SEQ ID NO: 44	SEQ ID NO: 44
107	SEQ ID NO: 40	SEQ ID NO: 51	234	SEQ ID NO: 44	SEQ ID NO: 45
108	SEQ ID NO: 40	SEQ ID NO: 52	235	SEQ ID NO: 44	SEQ ID NO: 54
109	SEQ ID NO: 40	SEQ ID NO: 53	236	SEQ ID NO: 44	SEQ ID NO: 55
110	SEQ ID NO: 40	SEQ ID NO: 42	237	SEQ ID NO: 44	SEQ ID NO: 56
111	SEQ ID NO: 40	SEQ ID NO: 43	238	SEQ ID NO: 44	SEQ ID NO: 57
112	SEQ ID NO: 40	SEQ ID NO: 44	239	SEQ ID NO: 45	SEQ ID NO: 45
113	SEQ ID NO: 40	SEQ ID NO: 45	240	SEQ ID NO: 45	SEQ ID NO: 54
114	SEQ ID NO: 40	SEQ ID NO: 54	241	SEQ ID NO: 45	SEQ ID NO: 55
115	SEQ ID NO: 40	SEQ ID NO: 55	242	SEQ ID NO: 45	SEQ ID NO: 56
116	SEQ ID NO: 40	SEQ ID NO: 56	243	SEQ ID NO: 45	SEQ ID NO: 57
117	SEQ ID NO: 40	SEQ ID NO: 57	244	SEQ ID NO: 54	SEQ ID NO: 54
118	SEQ ID NO: 41	SEQ ID NO: 41	245	SEQ ID NO: 54	SEQ ID NO: 55
119	SEQ ID NO: 41	SEQ ID NO: 47	246	SEQ ID NO: 54	SEQ ID NO: 56
120	SEQ ID NO: 41	SEQ ID NO: 48	247	SEQ ID NO: 54	SEQ ID NO: 57
121	SEQ ID NO: 41	SEQ ID NO: 49	248	SEQ ID NO: 55	SEQ ID NO: 55
122	SEQ ID NO: 41	SEQ ID NO: 50	249	SEQ ID NO: 55	SEQ ID NO: 56
123	SEQ ID NO: 41	SEQ ID NO: 51	250	SEQ ID NO: 55	SEQ ID NO: 57
124	SEQ ID NO: 41	SEQ ID NO: 52	251	SEQ ID NO: 56	SEQ ID NO: 56

#	WXRE in a first vector	WXRE in a second vector	#	WXRE in a first vector	WXRE in a second vector
125	SEQ ID NO: 41	SEQ ID NO: 53	252	SEQ ID NO: 56	SEQ ID NO: 57
126	SEQ ID NO: 41	SEQ ID NO: 42	253	SEQ ID NO: 57	SEQ ID NO: 57
127	SEQ ID NO: 41	SEQ ID NO: 43			

Expression system

The present disclosure provides an efficient protein expression system that utilizes piggyBac transposons and/or one or more regulatory elements as described herein. As used herein, the term "protein expression system" or "expression system" refers to a system comprising a host and a vector containing a heterologous sequence (e.g., exogenous gene), and the expression of the heterologous sequence in the host can be achieved by this system. The protein expression system generally comprises the following parts: (1) a host, i.e., an organism expressing proteins, which can be selected from bacteria, yeast, plant cells, animal cells, and the like; (2) one or more vectors. The type of the vector matches with the host. According to the different hosts, the vectors can be prokaryotic (bacterial) expression vectors, yeast expression vectors, plant expression vectors, mammalian expression vectors, insect expression vectors, and the like. The vector contains a fragment of a heterologous gene. The heterologous gene can be expressed in the host via the mediation of the vector. In some embodiments, the expressed protein products are secreted. In some embodiments, the vectors are integrated into host cell DNA.

A vector can be a delivery vehicle for a polynucleotide. In some embodiments, the vector includes a polynucleotide sequence encoding a certain protein operatively inserted therein and allows the expression of this protein in a genetic engineering recombinant technique. The vector can be used to transform, transduce or transfect a host cell. The vector in the present disclosure can be any suitable vector, which includes chromosomal, non-chromosomal and synthetic nucleic acid vectors (including a group of suitable nucleic acid sequences which express the various elements). For example, a vector can be a recombinant plasmid vector, a recombinant eukaryotic viral vector, a recombinant phage vector, a recombinant yeast minichromosome vector, a recombinant bacterial artificial chromosome vector, or a recombinant yeast plasmid vector.

In some embodiments, the vector in the present disclosure can include the derivatives of SV40, bacterial plasmids, phage DNAs, baculovirus, yeast plasmid, vectors derived from a combination of a plasmid and a phase DNA, and vectors such as virus nuclear acids (RNA or DNA). In some embodiments, the vector is an adeno-associated virus (AAV) vector.

As used herein, the term "host cell" in the present disclosure refers to a cell that receives a heterologous polynucleotide and/or a vector introduced therein. The host cell can be a eukaryotic host cell or a prokaryotic host cell, wherein the eukaryotic host cell can be a mammalian host cell, an insect host cell, a plant host cell, a fungal host cell, a eukaryotic algae host cell, a nematode host cell, a protozoan host cell, and a fish host cell. Illustratively, the host cell in the present disclosure is a eukaryotic host cell, e.g., a mammalian host cell. In some embodiments, the mammalian host cell is a Chinese hamster ovary (CHO) cell, a COS cell, a Vero cell, a SP2/0 cell, a NS/O myeloma cell, an immature hamster kidney cell, a HeLa cell, a human B cell, a cv-1/EBNA cell, an L cell, a 3T3 cell, a HEPG2 cell, a PerC6 cell, a human embryonic kidney 293 (HEK 293) cell, or an MDCK cell. In some embodiments, the cell is a human embryonic retinal cell (PER.C6) with the transforming early region (E1) of adenovirus type 5 (ad5). CHO cells are routinely used for the production of biopharmaceutical proteins. In some embodiments, CHO cell is a CHO-K1 cell, a CHO-DG44 cell, or a CHO-S cell. In a preferred embodiment, the CHO cell is a CHO-K1 cell.

A key step in protein expression is the selection of recombinant host cells which have been successfully transfected with the vector comprising the heterologous gene coding the protein of interest. Most commonly a selection marker is included in the vector. The selection marker can be a gene or DNA sequence that allows separation of recombinant host cells containing the marker and those not containing it. The combination of a selection marker and a selection medium allows growth of recombinant host cells that have been transfected with the vector, while in some embodiments, inhibiting the growth of host cells that have not been successfully transfected.

Antibiotic resistance genes are the most commonly used markers for recombinant host cell selection. An antibiotic resistance gene as a selection marker, in combination with a selection medium containing the antibiotic, can be used in order to achieve selection. Exemplary antibiotic selection markers include but are not limited to ampicillin resistance gene, chloramphenicol resistance gene, kanamycin resistance gene, tetracycline resistance gene,

polymyxin B resistance gene, erythromycin resistance gene, carbenicillin resistance gene, streptomycin resistance gene, spectinomycin resistance gene, blasticidin resistance gene, neomycin resistance gene, puromycin resistance gene, zeocin resistance gene, and hygromycin B resistance gene. Accordingly, the selection antibiotics include but are not limited to ampicillin, chloramphenicol, kanamycin, tetracycline, polymyxin B, erythromycin, carbenicillin, streptomycin, spectinomycin, blasticidin, neomycin, puromycin, zeocin, and hygromycin B. In some embodiments, the selection marker is blasticidin resistance gene. In some embodiments, the selection marker is zeocin resistance gene.

In some embodiment, the selection medium can comprises one or more of the following ingredients: serum, polysaccharide (e.g. glucose, and/or dextrose) , sodium pyruvate, glutathione, ethanolamine, amino acid (e.g. glycine, alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, histidine, isoleucine, leucine, lysine, glutamine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and/or valine) or a salt thereof, vitamin (e.g. ascorbic acid phosphate, choline chloride, D-calcium pantothenate, folic acid, niacinamide, pyridoxine hydrochloride, riboflavin, thiamine hydrochloride, and/or i-inositol) , inorganic salt (e.g. calcium chloride, ferric nitrate, magnesium sulfate, potassium chloride, sodium bicarbonate, sodium chloride, and/or sodium phosphate dibasic) , protein (e.g. human transferrin and/or recombinant insulin) , and/or trace element (e.g. ammonium metavanadate, cupric sulfate, manganous chloride, and/or sodium selenite).

The expression system can be used to express various proteins or polypeptides, e.g., an antibody, a fusion protein, an enzyme, a soluble protein, a membrane protein, a structural protein, a ribosome protein, a zymogen, a cell surface receptor protein, a transcriptional regulatory protein, a translational regulatory protein, a chromatin protein, a hormone, a cell cycle regulatory protein, a G protein, a neuroactive peptide, an immunomodulatory protein, a blood component protein, an ion gate protein, a heat shock protein, dihydrofolate reductase, an antibiotic resistance protein, a functional fragment of any one of the proteins, an epitope fragment of any one of the proteins, and any combination thereof.

As used herein, the term “antibody” in the present disclosure refers to an immunoglobulin, a fragment thereof, or a derivative of them, and includes any polypeptide comprising an antigen-binding site, regardless of whether it is produced in vitro or in vivo. This term includes, but is not limited to, a polyclonal antibody, a monoclonal antibody, a monospecific antibody, a bispecific

antibody, a trispecific antibody, a multispecific antibody, a non-specific antibody, a humanized antibody, a fully human antibody, a chimeric antibody, a single-domain antibody, a single-stranded antibody, a synthetic antibody, a recombinant antibody, a heterozygous antibody, a mutated antibody, and a grafted antibody. The term "antibody" also includes antibody fragments such as Fab, Fab', $F(ab')_2$, Fv, scFv, Fd, dAb, and other antibody fragments that retain the antigen-binding function. Typically, such fragments will include an antigen-binding fragment.

As used herein, the term "fusion protein" in the present disclosure refers to a molecule comprising two or more proteins or the fragments thereof which are linked by the covalent bond via their respective main chains of the peptides, and more preferably, the fusion protein is generated by the genetic expression of the polynucleotide molecules encoding these proteins. In some embodiments, the fusion protein comprises an immunoglobulin domain. In some embodiments, the fusion protein is an Fc-fusion protein.

In some embodiments, the antibodies that can be used in connection with the expression system include, e.g., Adalimumab, Bezlotoxumab, Avelumab, Dupilumab, Durvalumab, Ocrelizumab, Brodalumab, Reslizumab, Olaratumab, Daratumumab, Elotuzumab, Necitumumab, Infliximab, Obiltoxaximab, Atezolizumab, Secukinumab, Mepolizumab, Nivolumab, Alirocumab, Evolocumab, Dinutuximab, Bevacizumab, Pembrolizumab, Ramucirumab, Vedolizumab, Siltuximab, Alemtuzumab, Trastuzumab, Pertuzumab, Obinutuzumab, Brentuximab, Raxibacumab, Belimumab, Ipilimumab, Denosumab, Ofatumumab, Besilesomab, Tocilizumab, Canakinumab, Golimumab, Ustekinumab, Certolizumab, Catumaxomab, Eculizumab, Ranibizumab, Panitumumab, Natalizumab, Omalizumab, Cetuximab, Efalizumab, Ibritumomab, Fanolesomab, Tositumomab, Gemtuzumab, Palivizumab, Necitumumab, Basiliximab, Rituximab, Capromab, Satumomab, and Muromonab.

In some embodiments, the fusion proteins that can be used in the present disclosure include, e.g., Etanercept, Alefacept, Abatacept, Rilonacept, Romiplostim, Belatacept, and Aflibercept.

In some embodiments, the expression system provides at least two transposon vectors. One transposon vector is designed to carry a sequence encoding a first polypeptide (e.g., an antibody heavy chain). The second transposon vector is designed to carry a sequence encoding a second polypeptide (e.g., an antibody light chain). In some embodiments, one singly transposon vector provides a sequence encoding two or more polypeptides. For example, the sequence

encoding the antibody heavy chain and the sequence encoding the antibody light chain can be on the same transposon vector. They can be located in the same expression cassette or in different expression cassettes within the transposon vector.

In some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: TTAA, 5'-ITR, optionally a 5'-internal domain, a promoter, a gene of interest, a selection marker, optionally a 3'-internal domain, 3'-ITR, and TTAA. In some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: PB transposase 5'-recognition site, a promoter, a gene of interest, a selection marker, and PB transposase 3'-recognition site. In some embodiments, the sequence can further include one or two regulatory elements. The one or two regulatory elements can be located between the promoter and the gene of interest and/or between the gene of interest and the polyA signal sequence. Thus, in some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: TTAA, 5'-ITR, optionally a 5'-internal domain, a promoter, a regulatory element, a gene of interest, a regulatory element, a polyA signal sequence, a promoter for a selection marker, the selection marker, a polyA signal sequence for the selection marker, optionally a 3'-internal domain, 3'-ITR, and TTAA. In some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: PB transposase 5'-recognition site, a promoter, a regulatory element, a gene of interest, a regulatory element, a polyA signal sequence, a promoter for a selection marker, the selection marker, a polyA signal sequence for the selection marker, and PB transposase 3'-recognition site.

The gene of interest can actually include a sequence encoding two or more polypeptides (e.g., an antibody heavy chain and an antibody light chain). These sequences can separated from one another by sequences encoding a self-cleavage peptide (e.g., P2A or T2A) or a protease recognition site (e.g., furin). The open reading frame (ORF) thus encodes a single polyprotein, which, either during or after translation, can be cleaved into the individual proteins. Similarly, the transposase vector can have a sequence comprising or consisting of a promoter, a piggyBac transpose coding sequence, a polyA signal sequence. In some embodiments, the transposase vector further comprises a selection marker.

The expression of gene of interest can be further enhanced by WXRE transcription regulatory elements. In some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: TTAA, 5'-ITR, optionally a 5'-internal domain, a WXRE transcription regulatory element, a promoter, optionally the first intron of human EF-1α, a regulatory element, a gene of interest, a regulatory element, a polyA signal sequence, a promoter for a selection marker, the selection marker, a polyA signal sequence for the selection marker, optionally a 3'-internal domain, 3'-ITR, and TTAA. In some embodiments, the transposon vector can comprise a sequence comprising or consisting of from 5' to 3' one or more of the following elements: PB transposase 5'-recognition site, a WXRE transcription regulatory element, a promoter, optionally the first intron of human EF-1α, a regulatory element, a gene of interest, a regulatory element, a polyA signal sequence, a promoter for a selection marker, the selection marker, a polyA signal sequence for the selection marker, and PB transposase 5'-recognition site.

In some embodiments, the promotor is a CMV promotor. In some embodiments, the CMV promoter has a sequence identity with the sequence as shown in SEQ ID: 59. In some embodiments, the sequence having sequence identity with the sequence as shown in SEQ ID: 59 has a sequence identity of at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% (including all the ranges and percentages between these values) with the sequence as shown in SEQ ID: 59.

In some embodiments, the promoter is an inducible promoter. Inducible promoters include any promoter capable of increasing the amount of gene product produced, by a given gene, in response to exposure to an inducer. Thus the use of this construct allows for control of the expression of the target functional gene or transposase introduced into the host cell. Inducible promoters are known to those familiar with the art and a variety exists that can be used to drive expression. Inducible systems include, for example, the heat shock promoter system, the metallothionein system, the glucocorticoid system, tissue specific promoters, etc. Promoters regulated by heat shock, such as the promoter normally associated with the gene encoding the 70-kDa heat shock protein, can increase expression several-fold after exposure to elevated temperatures. The glucocorticoid system also functions well in triggering the expression of genes. The system consists of a gene encoding glucocorticoid receptor protein (GR) which in the presence of a steroid hormone (i.e. glucocorticoid or one of its synthetic equivalents such as

dexamethasone) forms a complex with the hormone. This complex then binds to a short nucleotide sequence (26 bp) named the glucocorticoid response element (GRE), and this binding activates the expression of linked genes. Thus inducible promoters can be used as an environmentally inducible promoter for controlling the expression of the introduced gene.

In some embodiments, the inducible promoter is a T7 promoter. In some embodiments, the inducible promoter is PA1lacO1 promoter. In some embodiments, the inducible promoter is activated by an agent selected from a group that includes IPTG, sodium salicylate, octapine, nopaline, the quorum signal 3OC6HSL, aTc, cuminic acid, DAPG, and salicylic acid. In some embodiments, the inducible promoter has a terminator and the terminator is downstream from the inducible promoter. Examples of inducible promoters regulated by exogenously supplied promoters include the zinc-inducible sheep metallothionine (MT) promoter, the dexamethasone (Dex)-inducible mouse mammary tumor virus (MMTV) promoter, the T7 polymerase promoter system, the ecdysone insect promoter, the tetracycline-repressible system, the tetracycline-inducible system, and the rapamycin-inducible system.

The promoters can also be multicistronic (bicistronic or tricistronic). For example, in some embodiments, transcription units can be engineered as a bicistronic unit containing an IRES (internal ribosome entry site), which allows coexpression of gene products (e.g. encoding an antibody heavy chain and an antibody light chain) by a message from a single promoter. Alternatively, in some cases, a single promoter may direct expression of an RNA that contains, in a single open reading frame (ORF), two or three genes (e.g. encoding an alpha chain and/or beta chain of a TCR) separated from one another by sequences encoding a self-cleavage peptide (e.g., P2A or T2A) or a protease recognition site (e.g., furin). The ORF thus encodes a single polyprotein, which, either during (in the case of 2A e.g., T2A) or after translation, is cleaved into the individual proteins. In some cases, the peptide, such as T2A, can cause the ribosome to skip (ribosome skipping) synthesis of a peptide bond at the C-terminus of a 2A element, leading to separation between the end of the 2A sequence and the next peptide downstream.

The first intron of human EF-1α can be used to increase the expression level. The first intron of human EF-1α can have a sequence that has a sequence identity with the sequence as shown in SEQ ID: 34. In some embodiments, the sequence having sequence identity with the sequence as shown in SEQ ID: 34 has a sequence identity of at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%

(including all the ranges and percentages between these values) with the sequence as shown in SEQ ID: 34 (the first intron of human EF-1α).

In some embodiments, the expression system as described herein can increase the expression amount of a heterologous protein by about or at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, or 50 folds (e.g., as compared to a control expression system without the regulatory element sequences or the piggyBac transposon).

The disclosure also provides a nucleic acid sequence that is at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to any nucleotide sequence as described herein, and an amino acid sequence that is at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identical to any amino acid sequence as described herein. In some embodiments, the disclosure relates to nucleotide sequences encoding any peptides that are described herein, or any amino acid sequences that are encoded by any nucleotide sequences as described herein. In some embodiments, the nucleic acid sequence is less than 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 150, 200, 250, 300, 350, 400, 500, 600, 700, 800, 900, 1000, or 5000 nucleotides. In some embodiments, the amino acid sequence is less than 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 500, or 1000 amino acid residues.

As used herein, the term "sequence identity" and the "percent identity" in the present disclosure refer to the percentage of the same (i.e., identical) nucleotides or amino acids between two or more polynucleotides or polypeptides. The sequence identity between two or more polynucleotides or polypeptides can be determined by the following method. The nucleotide sequences or the amino acid sequences of the polynucleotides or polypeptides are aligned and the number of positions containing the same nucleotide or amino acid residue in the aligned polynucleotides or polypeptides is scored and compared with the number of positions containing different nucleotides or amino acid residues in the aligned polynucleotides or polypeptides. The polynucleotides may differ at one position, for example, by containing different nucleotides (i.e., substitutions or mutations) or by the deletion of nucleotide (s) (i.e., the insertion of nucleotide (s) or the deletion of nucleotide (s) in one or two polynucleotides). The polypeptides may differ at one position, for example, by containing different amino acids (i.e., substitutions or mutations) or

by the deletion of amino acid (s) (i.e., the insertion of amino acid (s) or the deletion of amino acid (s) in one or two polypeptides). The sequence identity can be calculated by dividing the number of positions containing the same nucleotide or amino acid residue by the total number of the amino acid residues in the polynucleotide or polypeptide. For example, the percent identity can be calculated by dividing the number of positions containing the same nucleotide or amino acid residue by the total number of the nucleotides or the amino acid residues in the polynucleotide or polypeptide. Thus, the percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. For purposes of the present disclosure, the comparison of sequences and determination of percent identity between two sequences can be accomplished e.g., using a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The present disclosure also provides a kit comprising the recombinant host cell as described herein, and/or a kit comprising the vectors or the expression system as described herein.

Methods of using expression system, transposon vectors, and transposase vectors

The present disclosure provides methods of using expression system, transposon vectors, and transposase vectors.

In one aspect, provided herein are methods of using the expression system to produce recombinant protein in host cells (e.g., Chinese hamster ovary (CHO) cells, CHO-K1 cells, or any cells that are commonly used for protein expression known in the art).

In some embodiments, an appropriate number of host cells (e.g., at least or about 1, 2, 3, 5, 6, 8, 9, 10, 20, 30, 40, 50, or 100 million) are transfected by an appropriate amount of transposon vectors (e.g., recombinant protein expressing plasmid) and the transposase vectors (e.g., transposase expressing plasmid). In some embodiments, the cell density is between 3.5 to 4.5 million cells/mL. In some embodiments, the host cells are transfected by a total of at least or about 1 μg of transposon vectors (e.g., at least or about 1 μg, 2 μg, 3 μg, 4 μg, 5 μg, 6 μg, 7 μg, 8 μg, 9 μg, 10 μg, 20 μg, 30 μg, 40 μg, 50 μg, or 100 μg).

In some embodiments, the transposon vectors (e.g., recombinant protein expressing plasmid) and the transposase vector (e.g., transposase expressing plasmids) used for transfection can have a mass ratio of about 1:1, about 2:1, about 3:1, about 4:1, about 5:1, about 6:1, about 7:1, about 8:1, about 9:1, about 10:1, about 11:1, about 12:1, about 13:1, about 14:1, about 15:1, about 16:1, about 17:1, about 18:1, about 19:1, or about 20:1. Then, the transfected cells are resuspended in an appropriate volume (e.g., at least or about 1 mL, 2 mL, 3 mL, 4 mL, 5 mL, 6 mL, 7 mL, 8 mL, 9 mL, 10 mL, 15 mL, 20 mL, 30 mL, 40 mL, or 50 mL) of the host cell culture. The resulting solution can be mixed and incubated in a shaking incubator (e.g., a Kuhner® shaking incubator).

Many transformation or transfection techniques are available to introduce vectors into a cell. Electroporation is also a commonly-used method for introducing DNA into a cell. In this technique, cells are subject to electrical impulses of high field strength which reversibly permeabilizes biomembranes, allowing the entry of exogenous DNA sequences. In some embodiments, fertilized eggs are microinjected with the vectors. In some embodiments, PEI (polyetherimide) can be added to transduce the plasmids as described herein to the host cells (e.g., CHO-K1 cells). In some embodiments, the vectors can be microinjected directly into cells though the use of micropipettes. Alternatively, high velocity ballistics can be used to propel small DNA associated particles into the cell. In some embodiments, the cell is permeabilized by the presence of polyethylene glycol, thus allowing DNA to enter the cell through diffusion. DNA can also be introduced into a cell by fusing protoplasts with other carriers which contain DNA. These carriers include minicells, cells, lysosomes or other fusible lipid-surfaced bodies. The resulting cell culture can be incubated in an appropriate container (e.g., a spin tube) in an incubator (e.g., a shaking incubator).

After an appropriate time (e.g., at least or about 2 hours, 4 hours, 6 hours, 12 hours, 18 hours, 24 hours, 30 hours, 36 hours or longer) of transfection, an appropriate volume (e.g., at least or about 1 mL, 2 mL, 5 mL, 10 mL, 15 mL, or 20 mL) of fresh medium containing antibiotics for selection (e.g., blasticidin, geneticin (G-418), hygromycin B, mycophenolic acid, puromycin, zeocin) can be added to the transfected cell culture. The antibiotic reagent corresponds to the antibiotic resistance gene as decided herein. The concentration of antibiotics can be determined (e.g., by the killing curve experiment) on the native host cell (e.g., cells without modifications). In addition, the concentration of the replenishing medium can be the

same or higher than the level of the normal antibiotic concentration for selection (e.g., at least or about 2 folds, 3 folds, 4 folds, 5 folds, or higher).

In some embodiments, cell passaging can be carried out (e.g., every day, every 2 days, every 3 days, every 4 days, every 5 days, or every 6 days) with the fresh medium containing the selected antibiotic reagent. The seeding density can be adjusted based on cell growth conditions (e.g., the viability rate, growth rate, and doubling time of the cells).

In some embodiments, after an appropriate time (approximately 1 week, approximately 2 weeks, approximately 3 weeks, approximately 4 weeks, approximately 5 weeks, or longer) of antibiotic selection, the cell culture can be used to inoculate the production basal medium at the about same seeding density as determined for each transfected cell culture, respectively. The production cell cultures can be incubated in a shaking incubator.

In some embodiments, the production process can be performed by fed-batch culturing. Appropriate type and amount of feeding medium can be supplemented to the culture accordingly.

In some aspects, the disclosure provides methods that are designed for quickly evaluating a heteromultimer (e.g., antibody) expression. For example, for efficient expression of antibodies, the antibody heavy chain and the antibody light chain needs to be expressed in roughly 1: 1 ratio. If the concentration for a selection antibiotic is too low, the amount of functional vectors in the cells can be too small. If the concentration for a selection antibiotic is too high, it may create a condition that is not favorable for culturing cells. In some embodiments, the expression system in the present methods involve a pair of two vectors, one carrying a heterologous gene encoding an antibody heavy chain and the other carrying a heterologous gene encoding an antibody light chain. The selection marker in the two vectors might be different. In one embodiment, the selection marker in the first vector is blasticidin while the selection marker in the second vector is zeocin. The concentration of blasticidin and zeocin can be any concentrations as described herein. In some embodiments, the methods can also involve one vector comprising a heterologous gene encoding an antibody heavy chain and a heterologous gene encoding an antibody light chain. The ratio of the two vectors needs to be properly adjusted. It has been determined, based on tests on many different conditions, the methods provided herein can express antibodies with a high efficiency, and can be used to reliably evaluate the heteromultimer expression in a reasonably short time. Furthermore, the methods provided herein can express antibodies with a high expression level. Thus, in some embodiments, the methods

involve transfecting the cell a pair of two transposon vectors, one carrying a heterologous gene encoding a first polypeptide and the other carrying a heterologous gene encoding a second polypeptide. Two selection markers are used. One selection marker is blasticidin resistance gene, and the other selection marker is zeocin resistance gene. In some embodiment, blasticidin is present in the selection medium in an amount of 1-15 μg/mL and zeocin is present in an amount of 50-1500 μg/mL. In some embodiments, after about 18～30 hours (e.g., about 24 hours) of transfection, the cells are cultured in an appropriate cell culture medium containing blasticidin (e.g., 9 μg/mL) and Zeocin (e.g., 400 μg/mL). The cells are then passaged to a new medium containing blasticidin and Zeocin every 2 to 4 days. When the cell viability is recovered to 90%or more, the expression level of the heteromultimer can be evaluated by fed-batch cultures. In some embodiments, the fed-batch cultures can be any medium as described herein. In some embodiments, the fed-batch cultures contain blasticidin and Zeocin.

The present disclosure also provides methods of making a transgenic animal. In one aspect, the methods involve introducing ex vivo into a non-human vertebrate embryo or fertilized oocyte a nucleic acid comprising a transposon vector as described herein, and, within the same or on a separate vector, a nucleotide sequence encoding a transposase as described herein. The resultant non-human vertebrate embryo or fertilized oocyte can be selected and then implanted into a foster mother of the same species under conditions favoring development of the embryo into a transgenic non-human vertebrate. The embryo can then develop into a transgenic non-human vertebrate, thereby generating a transgenic non-human vertebrate comprising an exogenous nucleotide sequence.

Many selection markers can be used, include, for example, genes that provide antibiotic, pesticide, insecticide, herbicide resistance; genes that modify the physiology of the host, such as for example eye color or green fluorescent protein, to produce an altered visible phenotype; etc. The inserted DNA is integrated in the genome and can be stably passed to the subsequent progenies. In some embodiments, cross-breeding is performed to generate a heterozygous or homozygous transgenic animal with the inserted sequence.

In some embodiments, the animal is a cow, cat, dog, horse, sheep, mouse, rat, guinea pig, hamster, mink, panda, or pig. In some embodiments, the cell is of mammalian origin, and can be obtained from various animals described herein.

In one aspect, the disclosure is related to a method of preparing a recombinant host cell which stably expresses a protein, comprising a step of inserting into a host cell the vector as described herein. In some embodiments, the host cell is a Chinese hamster ovary (CHO) cell. In some embodiments, provided herein is the method that comprises a step of culturing the recombinant host cell as described herein under conditions that allow production of the protein.

In one aspect, the disclosure also provides a method for identifying a polypeptide having a desired property (e.g., binding specificity or functionality). The method involve generating a diverse collection of polynucleotides, preferably plasmid vectors or double stranded DNA PCR amplicons, encoding polypeptides having different properties, wherein said polynucleotides comprise a sequence coding for a polypeptide disposed between ITR sequences that are recognized by and functional with a least one transposase enzyme. The diverse collection of polynucleotides are then introduced into host cells. At least one transposase enzyme functional with said inverted terminal repeat sequences is expressed in the host cells so that the diverse collection of polynucleotides are integrated into the host cell genome to provide a host cell population that expresses said diverse collection of polynucleotides encoding polypeptides having different properties. The host cells are then screened to identify a host cell expressing a polypeptide having a desired property (e.g., binding specificity or functionality). The inserted sequence is then determined from the host cell.

In one aspect, the disclosure is related to a kit comprising the recombinant host cell as described herein, a method of using the recombinant host cell as described herein in the preparation of a reagent or a kit for detecting a disease due to the abnormality of protein expression. In one aspect, the disclosure also provides a method of using the recombinant host cell as described herein in the preparation of a pharmaceutical composition for treating or preventing a disease. Also provided herein are pharmaceutical compositions that contain at least one (e.g., one, two, three, or four) of the proteins (e.g., antibodies or antigen-binding fragments) described herein. The pharmaceutical compositions may be formulated in any manner known in the art.

EXAMPLES

The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Example 1: Construction of a vector library and construction of a stable pool which expresses green fluorescent protein

1.1 Preparation of a vector library containing a genomic fragment of Chinese hamster ovary cells

1.1.1 1 μg of the GFP-expressing vector (i.e., the vector as shown in FIG. 9) was subjected to enzyme digestion with BamHI in the enzyme digestion kit (NEB) containing the restriction endonuclease BamHI so as to be linearized and stayed overnight at 37°C (the composition and the contents of the reagents in the enzyme digestion reaction were as shown in Table 3), wherein BamHI could be replaced with any other endonucleases corresponding to a specific restriction site which existed in the upstream of a promoter corresponding to GFP.

The schematic diagram of the GFP-expressing vector was as shown in FIG. 9.

Table 3 Composition and contents of the reagents in the enzyme digestion reaction

Reaction components	Volume
NEB CutSmart Buffer (Cat# B7204S)	5 μL
BamHI	5 μL
GFP-expressing vector	1 μg
Ultrapure water	up to a total volume of 50 μL

1.1.2 Approximate five million CHO host cells were harvested, a DNeasy Blood & Tissue Kit (QIAGEN) was used to extract the genomic DNA of the CHO host cells, and said genomic DNA was dissolved in 100 μL of the elution buffer in the above-mentioned kit.

1.1.3 5 μg of the genomic DNA was subjected to enzyme digestion with 100 units of restriction endonuclease BglII (NEB) or DpnII (NEB) (the composition and the contents of the reagents in the enzyme digestion reaction were as shown in Table 4). Other restriction endonucleases might also be used, as long as they matched with the cohesive end of the endonuclease of the linearized vector in step 1.1.1.

Table 4 Composition and contents of the reagents in the enzyme digestion reaction

Reaction components	Volume
NEB CutSmart Buffer (Cat# B7204S)	5 μL
BamHI	5 μL
CHO genomic DNA	1 μg
Ultrapure water	up to a total volume of 50 μL

1.1.4 The linearized vector in 1.1.1 was treated with 2 units of calf intestinal alkaline phosphatase (NEB) at 37°C for approximate 30 minutes. Other types of alkaline phosphatases could also be used.

1.1.5 The linearized GFP-expressing vector in 1.1.4 and the digested CHO genomic DNA in 1.1.3 were subjected to separation by agarose gel electrophoresis, respectively. The gel was cut to recover the fragments of the GFP-expressing vector and the 1-4kb fragments of the digested genome, DNA was extracted from the agarose gel after electrophoresis using a QIAquick Gel Extraction Kit (QIAGEN).

1.1.6 The fragments of the GFP-expressing vector and the fragments of the genome recovered in 1.1.5 were subjected to ligation using a DNA Ligation Kit (Takara, Cat# 6022) for 45 minutes at 16°C (the composition and the contents of the reagents in the ligation reaction were as shown in Table 5).

Table 5 Composition and contents of the reagents in the ligation reaction

Reaction components	Volume
the recovered CHO genomic DNA	4 μL
the recovered vector	6 μL
Solution I	20 μg
Ultrapure water	10 μL

1.1.7 10 μL of the ligation product obtained by 1.1.6 was taken, 100 μL of competent cells were added, placed in an ice bath for 30 minutes, thermally stimulated at 42°C for 1 minute, and then placed on ice for 1 minute. 500 μL of fresh LB medium free of antibiotic was added to each tube of cells, and the cells were subjected to a 45-minute-recovery at 37°C. The step of plating was skipped and 500 mL of medium containing 100 mg/L of Ampicillin was added directly. The vector extraction was performed using a Plasmid Maxi Kit (QIAGEN). The extracted DNA was used as the vector library.

1.1.8 The vector library obtained in 1.1.7 was linearized using the restriction endonucleases of which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight under the same reaction conditions as in 1.1.1 at 37°C. DNA was recovered by the phenol-chloroform method and used for transfection the next day.

1.2 Construction of the stable pool expressing green fluorescent protein

1.2.1 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in an Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 0.3 μg to 0.6 μg of the linearized vector library obtained by step 1.1.8 were mixed evenly, and the cells were resuspended with this mixed solution and transferred to an electroporation cuvette. The cells were subjected to transfection using a program corresponding to the respective host cells in a 4D-Nucleofector™ System electroporation instrument. The electroporated cells were resuspended with 5 mL of medium free of antibiotic and placed in a shaker at 37°C for cultivation.

1.2.2 24 hours after the transfection, equal volume of selective medium containing antibiotic corresponding to the resistance gene in the vector was added in the cell culture medium for screening (in this experiment, the antibiotic was Zeocin (800 μg/mL)).

1.2.3 The cells were counted every 2 to 4 days. Cell passage was performed according to the growth situation of the cells, and screening was performed by using the selective medium with antibiotic corresponding to the resistance gene in the vector (in this experiment, the antibiotic was Zeocin (400 μg/mL)). Monoclonal screening was prepared when the cell viability recovered to 90% or more.

Example 2: Screening of a clone highly expressing green fluorescent protein

2.1 Single-cell sorting and expansion

2.1.1 The cells in the recovered pool in step 1.2.3 of Example 1 with a higher GFP expression level (for example, the top 0.5% of the expression level) were sorted by a FACS AriaII flow cytometer into a 96-well plate for cultivation.

2.1.2 75% of the medium in the plate was changed every 2 to 4 days until the recovered cells were visible by naked eyes.

2.2 Screening for a clone highly expressing GFP

The cells recovered in 2.1.2 were successively transferred into a new 96-well plate respectively, altogether about 300 clones (all the cells in each well were derived from one cell and were referred to as a clone herein). The expression level of GFP was determined by a FACS AriaII flow cytometer, and the clones whose detected intensities were among the top 10% were transferred to a 24-well plate for expansion.

Example 3: Screening, identification and verification of the transcriptional regulatory element

3.1 Identification of a candidate sequence of the transcriptional regulatory element

3.1.1 When the cells that were expanded in the 24-well plate in 2.2 substantially covered the bottom of the plate, a DNeasy Blood & Tissue Kit (QIAGEN) was used to extract the genome of each clone.

3.1.2 A forward primer and a reverse primer were respectively designed in the vector (about 200bp away from the upstream and the downstream of the restriction site of BamHI, respectively), and the genomes extracted in 3.1.1 were successively subjected to PCR amplification, wherein the sequence of the forward primer of the PCR reaction was GCAAAAAAGGGAATAAGGGCGACACGG (SEQ ID NO: 69) and the sequence of the reverse primer of the PCR reaction was CATAGCCCATATATGGAGTTCCGCGTTA (SEQ ID NO: 70).

The reaction system of the above-mentioned PCR reaction was as shown in Table 6.

Table 6 Reaction system of the PCR reaction

Reaction components	Volume
5X Q5 Reaction Buffer	5 μL
10 mM dNTPs	0.5 μL
10 μM forward primer	1.25 μL
10 μM reverse primer	1.25 μL
genome	1 μL
Q5 DNA Polymerase (Cat# M0491S)	0.25 μl
Ultrapure water	15.75 μL

The reaction steps of the above-mentioned PCR reaction were as shown in Table 7.

Table 7 Reaction steps of the PCR reaction

Temperature	Time	Number of Cycles
98℃	1 min	1
98℃	30s	35
61℃	30s	
68℃	5 min	
68℃	10 min	1

3.1.3 PCR products were subjected to separation by agarose gel electrophoresis, the gel was cut to recover the specific band(s) of 1kb or more, and the QIAquick Gel Extraction Kit (QIAGEN) was used to extract DNA.

3.1.4 The recovered band(s) was sent for sequencing, and the sequences A~G of the candidate transcriptional regulatory elements were identified.

3.1.5 The sequences A~G of the transcriptional regulatory elements obtained by sequencing and identification were as follows, wherein

the sequence of the transcriptional regulatory element A was the sequence as shown in SEQ ID NO: 35 (the reverse sequence of the transcriptional regulatory element A was the sequence as shown in SEQ ID NO: 47);

the sequence of the transcriptional regulatory element B was the sequence as shown in SEQ ID NO: 36 (the reverse sequence of the transcriptional regulatory element B was the sequence as shown in SEQ ID NO: 48);

the sequence of the transcriptional regulatory element C was the sequence as shown in SEQ ID NO: 37 (the reverse sequence of the transcriptional regulatory element C was the sequence as shown in SEQ ID NO: 49);

the sequence of the transcriptional regulatory element D was the sequence as shown in SEQ ID NO: 38 (the reverse sequence of the transcriptional regulatory element D was the sequence as shown in SEQ ID NO: 50);

the sequence of the transcriptional regulatory element E was the sequence as shown in SEQ ID NO: 39 (the reverse sequence of the transcriptional regulatory element E was the sequence as shown in SEQ ID NO: 51);

the sequence of the transcriptional regulatory element F was the sequence as shown in SEQ ID NO: 40 (the reverse sequence of the transcriptional regulatory element F was the sequence as shown in SEQ ID NO：52);

the sequence of the transcriptional regulatory element G was the sequence as shown in SEQ ID NO: 41 (the reverse sequence of the transcriptional regulatory element G was the sequence as shown in SEQ ID NO: 53);

the sequence of the transcriptional regulatory element H was the sequence as shown in SEQ ID NO: 42 (the reverse sequence of the transcriptional regulatory element H was the sequence as shown in SEQ ID NO: 54);

the sequence of the transcriptional regulatory element I was the sequence as shown in SEQ ID NO: 43 (the reverse sequence of the transcriptional regulatory element I was the sequence as shown in SEQ ID NO: 55);

the sequence of the transcriptional regulatory element J was the sequence as shown in SEQ ID NO: 44 (the reverse sequence of the transcriptional regulatory element J was the sequence as shown in SEQ ID NO: 56); and

the sequence of the transcriptional regulatory element K was the sequence as shown in SEQ ID NO: 45 (the reverse sequence of the transcriptional regulatory element K was the sequence as shown in SEQ ID NO: 57).

3.2 Verification of the transcriptional regulatory element

3.2.1 The transcriptional regulatory elements A~K obtained by sequencing and identification in 3.1.5 were respectively inserted into a BamHI restriction site upstream of the corresponding promoter in a vector containing the GFP gene using an In-Fusion Cloning Kit (Takara). A vector with the transcriptional regulatory element inserted therein as shown in FIG. 10 (wherein WXRE showed one of the transcriptional regulatory elements A~K) was obtained. The above-mentioned vector was linearized using the restriction endonucleases of which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight at 37°C. DNA was recovered by phenol-chloroform and used for transfection the next day.

3.2.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of a linearized

vector containing the gene of the protein to be expressed (obtained by 3.2.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The electroporated cells were resuspended with 5 mL of medium free of antibiotic and placed in a shaker at 37°C for cultivation. Each sample contained one kind of transcriptional regulatory element or was a control without any transcriptional regulatory element.

3.2.3 24 hours after transfection, equal volume of selective medium containing antibiotic corresponding to the resistance gene in the vector was added in the cell culture medium for screening. The cells were passaged using a medium containing antibiotic(s) every 2 to 4 days.

3.2.4 After the cell viability recovered to 90% or more, the influence of the transcriptional regulatory elements A~K on the expression level of the proteins was evaluated by fed-batch culture.

Example 4: Influence of the transcriptional regulatory elements on the expression level of a protein expression system used to express a heterologous protein

4.1.1 The transcriptional regulatory elements A~K were respectively constructed into the upstream BamHI position of the promoter of a fusion protein (the above-mentioned fusion protein was the A chain of PD-L1, whose sequence was the sequence as shown in SEQ ID NO: 71) in both forward and reverse directions by using In-Fusion Cloning Kit (Takara). A vector with the transcriptional regulatory element inserted therein as shown in FIG. 10 (wherein WXRE was one of the transcriptional regulatory elements A~K) was obtained, wherein the tail number 1 of the transcriptional regulatory element indicated the forward direction and the tail number 2 of the transcriptional regulatory element indicated the reverse direction. For example, the transcriptional regulatory element A1 (as shown in SEQ ID NO: 35) indicated the forward direction ((i.e., from 5' to 3') of the sequence and was transcribed in the sense strand of the coding sequence. The transcriptional regulatory element A2 showed the reverse complementary sequence of the sequence and was as shown in SEQ ID NO: 47. That is, the transcription regulatory element A1 in the sense chain of the protein coding sequence in Example 4 was equivalent to the transcriptional regulatory element A in Example 3 of the present disclosure.

The above-mentioned vectors were linearized using the restriction endonucleases of which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)) and stayed overnight under a condition of 37°C. DNA was recovered by phenol-chloroform and used for transfection the next day.

4.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the linearized vector containing the gene of the fusion protein (obtained by 4.1.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The electroporated cells were resuspended with 5 mL of medium free of antibiotic and placed in a shaker at 37°C for cultivation. Samples in one group only contained one transcriptional regulatory element of a certain direction (i.e., the forward direction or the reverse direction) and a sample which did not contain any transcriptional regulatory element was taken as a control.

4.1.3 24 hours after transfection, equal volume of medium containing 800 μg/mL of Zeocin was added into the transfected cells.

4.1.4 The cells were passaged using a medium containing 400 μg/mL of Zeocin every 2 to 4 days.

4.1.5 When the cell viability recovered to 90% or more, the expression level of the fusion protein PD-L1 was subjected to evaluation by fed-batch culture.

4.1.6 Whether the sequence of the PD-L1 obtained by expression was identical to the sequence as shown in SEQ ID NO: 71 was verified.

4.2 Experimental results

As shown in FIG. 11, as compared with the control group which did not have the transcriptional regulatory element, inserting the transcriptional regulatory element in the upstream of the promoter of the fusion protein could increase the expression level of the target protein by about 10% to 25% (see A2, B1, B2, D2, E2, F2, G1, H1, I2, J1, and K2 in FIG. 11). The promoting effect of the above-mentioned sequence on protein expression in a certain direction was superior to that in the other direction, which might be related to the directionality of the promoter.

As shown in FIG. 12, corresponding to the expression level, the forward or reverse transcriptional regulatory elements can make specific productivity increase by about 10% could enable an increase about 10% in specific productivity (see A2, B1, B2, D2, E2, F2, G1, H1, I2, J1, and K2 in FIG. 12).

Meanwhile, it was confirmed by verification that the sequence of the PD-L1 obtained by expression was identical to the sequence as shown in SEQ ID NO: 71.

Example 5: Influence of the transcriptional regulatory elements on the expression level of a protein expression system used to express Adalimumab

5.1.1 The reverse sequence of the transcriptional regulatory element A (A2), the forward sequence of the transcriptional regulatory element B (B1) and the forward sequence of the transcriptional regulatory element G (G1) were respectively constructed into the upstream of the promoter of the gene that could express Adalimumab by using the In-Fusion Cloning kit of Takara (the specific conditions were as shown in Table 7). Vectors with transcriptional regulatory elements inserted therein as shown in FIG. 13 and FIG. 14 (wherein WXRE was one of the transcriptional regulatory elements A~G) were obtained respectively, wherein the "transcriptional regulatory element in the upstream of the heavy chain" was cloned into the vector as shown in FIG. 13 and the "transcriptional regulatory element in the upstream of the light chain" was cloned into the vector as shown in FIG. 14. Among them, the amino acid sequence of the heavy chain (HC) of Adalimumab in FIG. 13 was as shown in SEQ ID NO: 72 and the amino acid sequence of the light chain (LC) of Adalimumab in FIG. 14 was as shown in SEQ ID NO: 73.

The above-mentioned vector was linearized using the restriction endonucleases of which the restriction sites were merely located in the prokaryotic region of the backbone of the vector (for example, PvuI (NEB)), and stayed overnight at 37°C. The DNA was recovered by phenol-chloroform and used for transfection the next day.

Table 8 Corresponding transcriptional regulatory elements under different conditions

sample ID	transcriptional regulatory element in the upstream of the heavy chain	transcriptional regulatory element in the upstream of the light chain
1	B1	B1
2	B1	G1

3	G1	G1
4	B1	A2
5	G1	A2
6 (control for 1-5)	N/A	N/A
7	H1	H1
8	I2	I2
9	J1	J1
10	K1	K1
11	K2	K2
12 (control for 7-11)	N/A	N/A

5.1.2 Approximate five million CHO host cells were centrifuged, and the supernatant was discarded. At the same time, 90 μL of SF Cell Line Solution and 20 μL of Supplement I in the Amaxa SF Cell Line 4D-Nucleofector Kit L (Lonza, Cat# VCA-1005) and 30 μg of the linearized vector containing the sequence of Adalimumab (obtained by 5.1.1) were mixed evenly, and the cells were resuspended with this mixed solution and transferred to the electroporation cuvette. The cells were subjected to transfection using the program corresponding to the respective host cells in the 4D-Nucleofector™ System electroporation instrument. The electroporated cells were resuspended with 5 mL of a medium free of antibiotic and placed in a shaker at 37°C for cultivation. Samples in one group only contained one transcriptional regulatory element of a certain direction (i.e., the forward direction or the reverse direction), and a sample which did not contain any transcriptional regulatory element was taken as a control.

5.1.3 A method that was designed for quickly assessing antibody expression was used. This method could ensure that the antibody heavy chain and light chain were roughly expressed in a ratio of 1:1, and could reliably assess the antibody expression within a considerably short period of time. Twenty-four hours after transfection, equal volume of medium containing 18 μg/mL of blasticidin and 800 μg/mL of Zeocin was added into the transfected cells.

5.1.4 The cells were passaged using a medium containing 9 μg/mL of blasticidin and 400 μg/mL of Zeocin every 2 to 4 days.

5.1.5 When the cell viability recovered to 90% or more, the expression level of Adalimumab was subjected to evaluation by fed-batch culture. Since both the heavy chain expression vector and the light chain expression vector of Adalimumab could be transfected into the same host cell, the heavy chain and the light chain of Adalimumab were capable of being

expressed simultaneously. Since the heavy chain and the light chain mentioned above were capable of self-assembly in the host cells, a complete Adalimumab was obtained.

5.1.6 The biological activity of the obtained Adalimumab was determined.

5.2 Experimental results

As compared with the control group, in some of the forward sequences containing the transcriptional regulatory element B (see sample 1, 2 and 4), the expression level of Adalimumab had an increase of 10% to 20% (as shown in FIG. 15).

By determining the biological activity of Adalimumab expressed by the present heterologous protein expression vector, it was found that its biological activity was identical to the biological activity of the known commercial Adalimumab.

EXAMPLE 6: Transposon-based Stable Pool Development

A recombinant protein expressing gene and an antibiotic resistance gene were inserted between a pair of terminal repeat sequences recognizable by the piggyBac transposase. For example, if the recombinant proteins (e.g., antibodies, or Fc fusion proteins) contain multiple expressing units, these different expressing units can be cloned into vectors containing different antibiotic resistance genes.

The plasmid containing piggyBac transposase expressing gene and the plasmid containing the recombinant protein expressing gene were separately prepared by endotoxin-free plasmid kits. Neither of the plasmids were linearized by digestion.

The host cell used for the transfection was suspension-adapted CHO-K1 cell. The expected viable cell density of the host cell were between 1.0 to 3.0 million cells/mL.

To generate the cell pool expressing different human IgG1 antibodies (A, B, and C), 10 million host cells were transfected by a total of 20 μg of the transposon vectors for expressing an antibody and the transposase vector at a mass ratio of 10:1. The transposon vectors included transposon vectors for the antibody heavy chain and transposon vectors for the antibody light chain at a roughly 1:1 ratio.

Then, the transfected cells were resuspended in 10 mL of the cell culture. For controls, the same amount of the host cells were transfected by the same amount of linearized antibody expressing plasmid. These linearized antibody expressing nucleic acid were not inserted into the

transposase expressing plasmids. The cells were mixed the plasmids and incubated in a Kuhner® shaking incubator.

After 24 hours of transfection, 10 mL of fresh medium containing antibiotics for selection was added to the cell culture. The antibiotic reagent corresponded to the antibiotic resistance gene. The concentration of antibiotics was determined by the killing curve experiment on the host cells. In addition, the concentration of the replenishing medium was twice the level of the antibiotic concentration for selection. This will allow the actual antibiotic concentration of the resulting cell culture to be at the selected antibiotic concentration level.

Cell passaging was carried out every 2 to 4 days with the fresh medium containing the selected antibiotic reagent. The seeding density was adjusted based on the viability, growth rate and doubling time of the cells.

After approximately 2 weeks of culture, the cell culture was used to inoculate the production basal medium at the same seeding density as determined for each transfected cell culture. The production cell cultures were incubated in a shaking incubator.

The production process was performed by fed-batch culturing. Appropriate type and amount of feeding medium was supplemented to the culture accordingly.

The results are shown in **FIG. 2**. The fed-batch culturing of the host cells applied with the piggyBac transposon system yielded 2.4 g/L antibody A, 4.1 g/L antibody B, and 4.6 g/L antibody C after 14 days of culture (PB D14 Titer). As shown in the figure, different sequences of the variable regions may lead to different expression levels. But for each antibody, the piggyBac transposon expression system significantly increased the expression level.

In contrast, the traditional stable pool controls yielded less than 0.2 g/L of antibodies (Traditional Pool D14 Titer). Therefore, the piggyBac transposon system increased protein expression by more than 10 folds under the same conditions.

EXAMPLE 7: Regulatory Element Screening

RNA-seq was performed on mRNAs of CHO-K1 cell at different stages as described below. Transient transfection as well as stable transfection were performed and total RNA was extracted from: 1) 10 samples on day 6 after transient transfection; 2) 10 samples on day 8 after transient transfection; and 3) 10 stable transfection samples on day 10 of a traditional 14 days fed-batch process.

cDNA was generated and was sequenced. Based on the relative reads number, mRNA was extracted and ranked by the average abundance across all 30 samples. Regulatory element (RE) sequences were extracted from top ranked ones and were listed in the following table.

Table 1.

RE ID	Regulatory Element Sequences	Reverse Complement Regulatory Element Sequences
01	GCCTCTTTCTTGTTAACATGTCCAATAAAAAGAAACTTTAGTTGTACTAGT (SEQ ID NO: 1)	ACTAGTACAACTAAAGTTTCTTTTTATTGGACATGTTAACAAGAAAGAGGC (SEQ ID NO: 16)
02	GAGGACTCTAGCTAACTCCCTGGAACAAATAAAGTTATTTTCCAGCTTAA (SEQ ID NO: 2)	TTAAGCTGGAAAATAACTTTATTTGTTCCAGGGAGTTAGCTAGAGTCCTC (SEQ ID NO: 17)
03	GCCTGATCCCTGGCATTTCAGGCAGCTCTGAACCGTGCTGTGTGTGCTCTGGAACCTCCTTCTCTGCTCTCAGGTTCCCCAGCTCCCATCTTGGATCCAGTGGAGAGGGTTTGCTTCTGCCACCAACAGCTCCCTTTGGTACATGCTCAGCATTCAGGAGTCTTTAAGGCAATACCATCAGAGAGCAAATAAATAAACGCGTTTATGTCTCTAAGCACA (SEQ ID NO: 3)	TGTGCTTAGAGACATAAACGCGTTTATTTATTTGCTCTCTGATGGTATTGCCTTAAAGACTCCTGAATGCTGAGCATGTACCAAAGGGAGCTGTTGGTGGCAGAAGCAAACCCTCTCCACTGGATCCAAGATGGGAGCTGGGGAACCTGAGAGCAGAGAAGGAGGTTCCAGAGCACACACAGCACGGTTCAGAGCTGCCTGAAATGCCAGGGATCAGGC (SEQ ID NO: 18)
04	ACAGGTTCAATCAGCTGTGCATTTGGAAAAATAAAACTTTATTAAATCAGA (SEQ ID NO: 4)	TCTGATTTAATAAAGTTTTATTTTTCCAAATGCACAGCTGATTGAACCTGT (SEQ ID NO: 19)
05	AGTCAACAAGCCCCTAGGCCTCAATAAAGGCAGCTGCCTCTGTTCCCCACAGCCTAAACCCTCA (SEQ ID NO: 5)	TGAGGGTTTAGGCTGTGGGGAACAGAGGCAGCTGCCTTTATTGAGGCCTAGGGGCTTGTTGACT (SEQ ID NO: 20)
06	GCCCAATAAAGACTGTTTGTGCTAA (SEQ ID NO: 6)	TTAGCACAAACAGTCTTTATTGGGC (SEQ ID NO: 21)
07	GGGCCCCTCATACACTGCTTCCATTAAAGACTGTTTAAGTAGT (SEQ ID NO: 7)	ACTACTTAAACAGTCTTTAATGGAAGCAGTGTATGAGGGGCCC (SEQ ID NO: 22)
08	GGATTCATACAATCAATGGCAGGACTTGAGAGTTTGTACTGAATCATGATCAATACCATGTATGCTGCCAGATGGAGTTCAACATTGTTAATCGGGAGACTTGTTCATGCTTAAGCTGGGAATGGTTTTGTCCTGTAATAAAAATATAGAGCCTTTCAAA (SEQ ID NO: 8)	TTTGAAAGGCTCTATATTTTTATTACAGGACAAAACCATTCCCAGCTTAAGCATGAACAAGTCTCCCGATTAACAATGTTGAACTCCATCTGGCAGCATACATGGTATTGATCATGATTCAGTACAAACTCTCAAGTCCTGCCATTGATTGTATGAATCC (SEQ ID NO: 23)
09	GACCTAAGTTAACCAGTTCCAGAAACAAGATCCTGAATTAAGTACGATTTGGTGTGTCTTTTGGGACAATAAAGACTTGTATTGAT (SEQ ID NO: 9)	ATCAATACAAGTCTTTATTGTCCCAAAAGACACACCAAATCGTACTTAATTCAGGATCTTGTTTCTGGAACTGGTTAACTTAGGTC (SEQ ID NO: 24)
10	AGATGTAAAACGTAAATAAAAAGCCTCCATAGACTGTT (SEQ ID NO: 10)	AACAGTCTATGGAGGCTTTTTATTTACGTTTTACATCT (SEQ ID NO: 25)
11	GCCCATCTCAAGGATCAGGGTTACCTTTGTAATAAACATCCCAGAGCTTTAGTG (SEQ ID NO: 11)	CACTAAAGCTCTGGGATGTTTATTACAAAGGTAACCCTGATCCTTGAGATGGGC (SEQ ID NO: 26)
12	ATCTGTTCTGTCAGATTTTCAATAAACCTG (SEQ ID NO: 12)	CAGGTTTATTGAAAATCTGACAGAACAGAT (SEQ ID NO: 27)
13	TTGTGTATGAATAAATAAAAAGACAGGAACTGA (SEQ ID NO: 13)	TCAGTTCCTGTCTTTTTATTTATTCATACACAA (SEQ ID NO: 28)
14	AATGGTCTCTAGGAGACATGCTGGAGAAATGTCTGTACTCTTGCCTTTTTAGGCAACTGTGCTCAATTAAACAGCATGATAAAATT (SEQ ID NO: 14)	AATTTTATCATGCTGTTTAATTGAGCACAGTTGCCTAAAAAGGCAAGAGTACAGACATTTCTCCAGCATGTCTCCTAGAGACCATT (SEQ ID NO: 29)

15	CAAATTGGATCTGTCACCTGTCACCATAGCTGACTG CTGCTTGCCATCCATACAACACCAGGGCTTAGGACA AATGGGACTGATGTCATCTTGAGCTTTTATTTTGACC ATGATTTATTTGGAGTGGAGACATTGTTTTTTTTCTT TTCTTTTTTTTAAAAAGAAAGAACATGTCGTGTAGGT TGTCTGAAAATAAAGTGCATTTAAATTCACTTA (SEQ ID NO: 15)	TAAGTGAATTTAAATGCACTTTATTTTCAGACAACCTACA CGACATGTTCTTTCTTTTTAAAAAAAAGAAAAGAAAAAAA ACAATGTCTCCACTCCAAATAAATCATGGTCAAAATAAAA GCTCAAGATGACATCAGTCCCATTTGTCCTAAGCCCTGGT GTTGTATGGATGGCAAGCAGCAGTCAGCTATGGTGACAG GTGACAGATCCAATTTG (SEQ ID NO: 30)

Experiments were performed to evaluate the effects of these RE sequences on protein expression. These RE sequences were incorporated into a fusion protein expression plasmid immediately after the recombinant protein expressing gene. The control sample did not contain any regulatory element sequences.

A total of 10 million host cells were transfected by 20 μg of the fusion protein expressing plasmids with different RE sequences. The transfected cells were resuspended in 10 mL of the host cell culture. The resulting solution was mixed and incubated in a Kuhner® shaking incubator.

After 24 hours of transfection, 10 mL of fresh medium containing antibiotics for selection was added to the transfected cell culture. Cell passaging was carried out every 2 to 4 days with the fresh medium containing selected antibiotic reagent. The seeding density was adjusted based on the viability, growth rate and doubling time of the cells.

After approximately 2 weeks of antibiotic selection, the cell culture was used to inoculate the production basal medium at the same seeding density as determined for each transfected cell culture, respectively. The production cell cultures were incubated in a shaking incubator.

The production process was performed by fed-batch culturing. Appropriate type and amount of feeding medium was supplemented to the culture accordingly.

As illustrated in **FIG. 3**, most of the regulatory element increased the protein expression by at least 10% productivity as compared to the control.

EXAMPLE 8: Combination of piggyBac and Regulatory Element

A recombinant protein expressing gene, one representative regulatory element sequence and an antibiotic resistance gene were cloned between a pair of terminal repeat sequences recognizable by piggyBac recombinase. For example, if the recombinant proteins (e.g., antibodies, or Fc fusion proteins) contain multiple expressing units, these different expressing units can be cloned into vectors containing different antibiotic resistance genes.

The plasmid containing piggyBac transposase expressing gene and the plasmid containing the recombinant protein expressing gene were separately prepared by endotoxin-free plasmid kits. None of the plasmids were linearized by digestion.

The host cell used for the transfection was suspension-adapted CHO-K1 cell. The expected viable cell density of the host cell were between 1.0 to 3.0 million cells/mL.

To generate the cell pool expressing different human IgG1 antibodies (D, E, and F), 10 million host cells were transfected by 20 μg of the piggyBac transposon vectors (with or without WXRE ID: B (SEQ ID NO: 36)) and the transposase vectors at a mass ratio of 10:1. Then, the transfected cells were resuspended in 10 mL of the host cell culture. As traditional stable pool controls, the same amount of the host cells were transfected by the same amount of linearized antibody expressing plasmids. The resulting solution was mixed and incubated in a Kuhner® shaking incubator.

After 24 hours of transfection, 10 mL of fresh medium containing antibiotics for selection was added to the transfected cell culture. The antibiotic reagent corresponded to the antibiotic resistance gene as decided herein. The concentration of antibiotics was determined by the killing curve experiment on the host cell. In addition, the concentration of the replenishing medium was twice the level of the selected antibiotic concentration for selection, so that the actual antibiotic concentration of the resulting cell culture is the same as the selected antibiotic concentration.

Cell passaging was carried out every 2 to 4 days with the fresh medium containing the selected antibiotic reagent. The seeding density was adjusted based on the viability rate and doubling time of the cells.

After 2 weeks of antibiotic selection, the cell culture was used to inoculate the production basal medium at the same seeding density as determined for each transfected cell culture, respectively. The production cell cultures were incubated in a shaking incubator.

The production process was performed by fed-batch culturing. Appropriate type and amount of feeding medium was supplemented to the culture accordingly.

As illustrated in **FIG. 4**, after 14 days, the protein expression (titer) of the three groups of production cell cultures using the piggyBac transposon system alone (without the RE sequence) were 2.1 g/L, 3.9 g/L and 5.5 g/L, respectively. However, the expression increased to 2.8 g/L, 5.2 g/L and 6.5 g/L, respectively, when the RE sequence was present in the antibody expressing

plasmids. The results indicated that combination of piggyBac transposon system and regulatory element sequences (e.g., WXRE) can further increase protein production.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A nucleic acid comprising
   a 5'- ITR (inverted terminal repeat) sequence;
   a 3'-ITR sequence; and
   a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-15 and SEQ ID NOs: 35-58.

2. The nucleic acid of claim 1, wherein the 5' ITR sequence comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 68, and the 3' ITR sequence comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 60.

3. The nucleic acid of claim 1, wherein the 5' ITR sequence comprises SEQ ID NO: 68 and the 3' ITR sequence comprises SEQ ID NO: 60.

4. The nucleic acid of claims 1, further comprising a 5'-internal domain and a 3'-internal domain, wherein the 5'-internal domain comprise a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66, wherein the 3'-internal domain comprise a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67, wherein the 5'-internal domain is immediately adjacent to the 5'-ITR, and the 3'-internal domain is immediately adjacent to the 3'-ITR.

5. The nucleic acid of claim 1, wherein the nucleic acid comprises one or more regulatory element sequences selected from the group consisting of SEQ ID NOs: 1-15 (e.g., SEQ ID NO: 2).

6. The nucleic acid of claim 1, wherein the nucleic acid comprises one or more regulatory element sequences selected from the group consisting of SEQ ID NOs: 35-46 (e.g., SEQ ID NO: 36).

7. The nucleic acid of any one of claims 1-6, further comprising a promoter and a sequence encoding a polypeptide, wherein the sequence encoding the polypeptide is operably linked to the promoter.

8. The nucleic acid of claim 7, wherein the sequence encoding a polypeptide is located between two regulatory element sequences.

9. The nucleic acid of any one of claims 1-6, further comprising a promoter and a sequence encoding two or more polypeptides, wherein the sequence encoding the two or more polypeptides is operably linked to the promoter.

10. The nucleic acid of claim 9, wherein the sequence encodes an antibody heavy chain and an antibody light chain.

11. The nucleic acid of any one of claims 1-10, wherein the nucleic acid further comprises a WXRE sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

12. The nucleic acid of any one of claims 1-10, wherein the nucleic acid comprises two or more expression cassettes.

13. The nucleic acid of any one of claims 1-12, wherein the nucleic acid comprises a selection marker.

14. The nucleic acid of claim 13, wherein the selection marker is an antibiotic resistance gene, a sequence encoding a fluorescent protein, or lacZ.

15. A vector comprising the nucleic acid of any one of claims 1-14.

16. A transposon vector comprising from 5' to 3':

a 5' ITR sequence consisting of a sequence that is at least 95% identical to SEQ ID NO: 68;

a non-transposon heterologous DNA sequence; and

a 3'-ITR sequence consisting of a sequence that is at least 95% identical to SEQ ID NO: 60.

17. The transposon vector of claim 16, wherein the 5' ITR sequence consists of SEQ ID NO: 68 and the 3' ITR sequence consists of SEQ ID NO: 60.

18. The transposon vector of claims 16, further comprising a 5'-internal domain and a 3'-internal domain, wherein the 5'-internal domain comprises a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66, wherein the 3'-internal domain comprises a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67, wherein the 5'-internal domain is immediately adjacent to the 5'-ITR, and the 3'-internal domain is immediately adjacent to the 3'-ITR.

19. The transposon vector of any one of claims 16-18, wherein the non-transposon heterologous DNA sequence comprises a promoter and a sequence encoding one or more polypeptides, wherein the sequence encoding one or more polypeptides is operably linked to the promoter.

20. The transposon vector of claim 19, wherein the promoter is a cytomegalovirus (CMV) promoter.

21. The transposon vector of any one of claims 16-20, wherein the non-transposon heterologous DNA sequence further comprises a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30.

22. The transposon vector of any one of claims 16-21, wherein the non-transposon heterologous DNA sequence further comprises a WXRE sequence that is at least 80%,

85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

23. The transposon vector of any one of claims 16-22, wherein the non-transposon heterologous DNA sequence comprises a multiple cloning site.

24. An expression system comprising:
(a) a first nucleic acid comprising
a 5'- ITR sequence,
a non-transposon heterologous DNA sequence comprising a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30,
a 3'-ITR sequence; and
(b) a second nucleic acid encoding a piggyBac transposase.

25. The expression system of claim 24, wherein the second nucleic acid encoding a piggyBac transposase having an amino acid sequence that is at least 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 33.

26. The expression system of claim 24, wherein the non-transposon heterologous DNA sequence comprises a promoter and a sequence encoding one or more polypeptides, wherein the sequence encoding one or more polypeptides is operably linked to the promoter.

27. The expression system of claim 26, wherein the sequence encodes an antibody heavy chain and/or an antibody light chain.

28. The expression system of claim 26, wherein the sequence encodes a monoclonal antibody, a bispecific antibody, a recombinant protein, or a fusion protein.

29. The expression system of claim 26, wherein the promoter is a CMV promoter.

30. The expression system of claim 26, wherein the promoter is an inducible promoter (e.g., a heat shock promoter, a metallothionein promoter, or a glucocorticoid response element).

31. The expression system of any one of claims 24-30, wherein the non-transposon heterologous DNA sequence further comprises a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30.

32. The expression system of any one of claims 24-31, wherein the non-transposon heterologous DNA sequence further comprises a WXRE regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

33. An isolated nucleic acid comprising a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30.

34. The isolated nucleic acid of claim 33, wherein the nucleic acid further comprises a promoter and a protein-coding sequence.

35. The isolated nucleic acid of claim 34, wherein the regulatory element sequence is located between the promoter and the protein-coding sequence.

36. The isolated nucleic acid of claim 34, wherein the regulatory element sequence is located at the 3' of the protein-coding sequence.

37. The isolated nucleic acid of any one of claims 33-36, wherein the regulatory element sequence can be transcribed to a 5'-UTR or a 3'-UTR.

38. The isolated nucleic acid of any one of claims 33-37, wherein the isolated nucleic acid further comprises a 5'-ITR and a 3'-ITR, and the regulatory element sequence is located between the 5'-ITR and the 3'-ITR.

39. A vector comprising the nucleic acid of any one of claims 33-38.

40. A vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction:
a 5'-ITR comprising a TR_L, a 5'-ITR spacer, a IR_L;
a promoter;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30;
a protein-coding sequence; and
a 3'-ITR comprising a IR_R, a 3'-ITR spacer, a TR_R.

41. A vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction:
a 5'-ITR comprising a TR_L, a 5'-ITR spacer, a IR_L;
a promoter;
a protein-coding sequence;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30; and
a 3'-ITR comprising a IR_R, a 3'-ITR spacer, a TR_R.

42. A vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction:
a 3'-ITR comprising a TR_R, a 3'-ITR spacer, a IR_R;
a promoter;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30;
a protein-coding sequence; and
a 5'-ITR comprising a IR_L, a 5'-ITR spacer, a TR_L.

43. A vector comprising a piggyBac transposon, the piggyBac transposon comprising the following genetic elements in a 5' to 3' direction:
a 3'-ITR comprising a TR_R, a 3'-ITR spacer, a IR_R;
a promoter;
a protein-coding sequence;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30; and
a 5'-ITR comprising a IR_L, a 5'-ITR spacer, a TR_L.

44. The vector of claim 42 or 43, wherein the TR_L comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 61; wherein the 5'-ITR spacer comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 62; wherein the IR_L comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 63; wherein the TR_R comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 64; and wherein the IR_R comprises or consists of a sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 65.

45. The vector of any one of claims 40-44, wherein the vector comprises two or more regulatory element sequences, each of which is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30.

46. The vector of any one of claims 40-45, wherein the vector further comprises
a 5'-internal domain sequence and a 3'-internal domain sequence.

47. The vector of claim 46, wherein the 5'-internal domain sequence comprises or consists of a sequence that is at least or about 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 66 and
the 3'-internal domain sequence comprises or consists of a sequence that is at least or about 80%, 85%, 90%, 95% or 100% identical to SEQ ID NO: 67.

48. The vector of any one of claims 40-47, wherein the regulatory element sequence can be transcribed to a 5'-UTR or a 3'-UTR.

49. The vector of any one of claims 40-48, wherein the vector further comprises a sequence encoding a piggyBac transposase, wherein the sequence encoding the piggyBac transposase is outside a region between the 5'-ITR and the 3'-ITR.

50. A method of generating a cell in culture for expressing a polypeptide of interest, comprising:
(a) introducing into a cell:
a transposon vector comprising:
a 5'- ITR (inverted terminal repeat) sequence;
a 3'-ITR sequence;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and SEQ ID NOs: 35-58; and
a sequence encoding the polypeptide of interest; and
(b) culturing the cell under an appropriate condition, wherein the piggyBac transposon is integrated into the genome of the cell,
thereby generating a cell for expressing the polypeptide of interest.

51. The method of claim 50, wherein the method further comprises introducing a vector comprising a sequence encoding a piggyBac transposase to the cell.

52. The method of claim 50, wherein the transposon vector comprises a sequence encoding a piggyBac transposase.

53. The method of any one of claims 50-52, wherein the transposon vector is introduced into the cell by microinjection, high velocity propulsion, permeabilization, fusion, or electroporation.

54. The method of any one of claims 50-53, wherein the cell is a Chinese hamster ovary (CHO) cell.

55. The method of any one of claims 50-53, wherein the cell is a mammalian cell or an insect cell.

56. A cell comprising the nucleic acid of any one of claims 1-14 and 33-38, the vector of any one of claims 15-23, 39-49, or the expression system of any one of claims 24-32.

57. The cell of claim 56, wherein the cell is a Chinese hamster ovary (CHO) cell.

58. A method of expressing a protein, comprising:
culturing the cell of claim 56 or 57 under conditions that allow the cell to express the protein; and
collecting and purifying the protein.

59. A protein expressed by the cell of claim 56 or 57, or produced by the method of claim 58.

60. A pharmaceutical composition comprising the protein of claim 59 and a pharmaceutically acceptable carrier.

61. A nucleic acid comprising
a 5'- ITR sequence;
a 3'-ITR sequence; and

one or more regulatory element sequences derived from CHO.

62. An expression system comprising:
(a) a first nucleic acid comprising a piggyBac transposon comprising the following genetic elements in a 5' to 3' direction:
a first TTAA sequence;
a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L;
a 5'-internal domain (ID);
a sequence of interest;
a 3'-ID;
a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR_R; and
a second TTAA sequence; and
(b) a second nucleic acid encoding a piggyBac transposase.

63. A method of generating a cell for expressing a polypeptide of interest, comprising:
(a) introducing into a cell:
a transposon vector comprising the following genetic elements in a 5' to 3' direction:
a first TTAA sequence;
a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L;
a 5'-internal domain (ID);
a sequence of interest;
a 3'-ID;
a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR_R; and
a second TTAA sequence; and
(b) culturing the cell under an appropriate condition.

64. A cell line whose genome is stably integrated with a piggyBac transposon comprising the following genetic elements:
a 5'- ITR sequence;
a regulatory element sequence that is at least 80%, 85%, 90%, 95% or 100% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-30 and 35-58; and

a 3'-ITR sequence.

65. A cell line whose genome is stably integrated with a piggyBac transposon comprising the following genetic elements:

a 5'-ITR comprising a TR_L, a 5'-ITR spacer and an IR_L;

a 5'-internal domain (ID);

a sequence of interest;

a 3'-ID;

a 3'-ITR comprising an IR_R, a 3'-ITR spacer and a TR_R.