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 example, 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 example, 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 example, 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 one aspect, the present disclosure refers to a nucleic acid comprising a 5'- ITR (inverted terminal repeat) sequence; a 3'-ITR sequence; and a regulatory element sequence that is at least 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 1-15 and SEQ ID NOs: 35-58.

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

In some examples, the nucleic acid as described herein, further comprising a 5'-internal domain and a 3'-internal domain. In some examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, the nucleic acid as described herein, further comprises a promoter and a sequence encoding a polypeptide. In some examples, the sequence encoding the polypeptide is operably linked to the promoter. In some examples, the sequence encoding a polypeptide is located between two regulatory element sequences.

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

In some examples, 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 examples, the nucleic acid comprises two or more expression cassettes. In some examples, the nucleic acid comprises a selection marker. In some examples, the selection marker is an antibiotic resistance gene, a sequence encoding a fluorescent protein, or lacZ.

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

In another aspect, the present disclosure refers to a vector comprising the nucleic acid disclosed herein.

In one example, 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 examples, the 5' ITR sequence consists of SEQ ID NO: 68 and the 3' ITR sequence consists of SEQ ID NO: 60.

In some examples, the transposon vector as described herein further comprises a 5'-internal domain and a 3'-internal domain. In some examples, 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 examples, 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 examples, 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 examples, the non-transposon heterologous DNA sequence comprises a promoter and a sequence encoding one or more polypeptides. In some examples, the sequence encoding one or more polypeptides is operably linked to the promoter. In some examples, the promoter is a cytomegalovirus (CMV) promoter.

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

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

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

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

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

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

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

In some examples, the regulatory element sequence can be transcribed to a 5'-UTR or a 3'-UTR. In some examples, 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 example, provided herein is a vector comprising the nucleic acid as described herein.

In one example, 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 example, 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 example, 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 example, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, 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 examples, the vector further comprises a 5'-internal domain sequence and a 3'-internal domain sequence.

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

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

In one example, 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 examples, the piggyBac transposon is integrated into the genome of the cell, thereby generating a cell for expressing the polypeptide of interest.

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

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

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

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

In one example, 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 yet another aspect, the present disclosure refers to a cell comprising the nucleic acid disclosed herein.

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

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

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

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

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

In some examples, 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 example, 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 disclosure belongs. Methods and materials are described herein for use in the present disclosure; 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 disclosure 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.

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 90% 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 of a sequence that is at least 90% identical to SEQ ID NO: 68, and the 3' ITR sequence comprises a sequence that is at least 90% 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 claim 1, further comprising a 5'-internal domain and a 3'-internal domain, wherein the 5'-internal domain comprise a sequence that is at least 90% identical to SEQ ID NO: 66, wherein the 3'-internal domain comprise a sequence that is at least 90% 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.

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.

7. The nucleic acid of claim 1, 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 claim 1, 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 claim 1, wherein the nucleic acid further comprises a WXRE sequence that is at least 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 35-46.

12. The nucleic acid of claim 1, wherein the nucleic acid comprises two or more expression cassettes.

13. The nucleic acid of claim 1, 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 claim 1.

16. The nucleic acid of claim 1, wherein the 5'-ITR comprises a TR_L, a 5'-ITR spacer and an IR_L, and the 3'-ITR comprises an IR_R, a 3'-ITR spacer and a TR_R.

17. The nucleic acid of claim 16, wherein the TR_L comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 61; wherein the 5'-ITR spacer comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 62; wherein the IR_L

comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 63; wherein the TR_R comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 64; and wherein the IR_R comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 65.

18. The vector of claim 15, comprising the following genetic elements in a 5' to 3' direction:
the 5'-ITR comprising a TR_L, a 5'-ITR spacer, a IR_L;
a promoter;
the regulatory element sequence;
a protein-coding sequence; and
the 3'-ITR comprising a IR_R, a 3'-ITR spacer, a TR_R.

19. The vector of claim 18, wherein the TR_L comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 61; wherein the 5'-ITR spacer comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 62; wherein the IR_L comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 63; wherein the TR_R comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 64; and wherein the IR_R comprises a sequence that is at least 90% identical to a reverse sequence or a reverse complementary sequence of SEQ ID NO: 65.

20. A cell comprising the nucleic acid of claim 1.