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 represented by SEQ ID NO: 1.

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).

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 represented by SEQ ID NO: 1.

2. The nucleic acid of claim 1, wherein the 5' ITR sequence comprises 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 comprises a sequence that is at least 90% identical to SEQ ID NO: 66, wherein the 3'-internal domain comprises 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, further comprising a promoter and a sequence encoding a polypeptide, wherein the sequence encoding the polypeptide is operably linked to the promoter.

6. The nucleic acid of claim 5, wherein the sequence encoding a polypeptide is located between two regulatory element sequences.

7. 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.

8. The nucleic acid of claim 7, wherein the sequence encodes an antibody heavy chain and an antibody light chain.

9. The nucleic acid of claim 1, wherein the nucleic acid further comprises a WXRE sequence that is selected from the group consisting of SEQ ID NOs: 35-46.

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

11. The nucleic acid of claim 1, wherein the nucleic acid comprises a selection marker.

12. The nucleic acid of claim 11, wherein the selection marker is an antibiotic resistance gene, a sequence encoding a fluorescent protein, or lacZ.

13. A vector comprising the nucleic acid of claim 1.

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