TITLE
Transposable elements: a jump toward the future of expression vectors

AUTHORS
Antonio Palazzo[1], René Massimiliano Marsano[2]

1 Laboratory of Translational Nanotechnology, "Istituto Tumori Giovanni Paolo II" I.R.C.C.S. Viale Orazio Flacco 65, 70125 Bari, Italy
2 Dipartimento di Biologia, Università degli Studi di Bari "Aldo Moro". Via Orabona 4, 70125 Bari (Italy)

Corresponding Author: René Massimiliano Marsano
email: renemassimiliano.marsano@uniba.it
Address: Dipartimento di Biologia, Università degli Studi di Bari "Aldo Moro". Via Orabona 4, 70125 Bari (Italy) tel.+390805442241

ABSTRACT
Expression vectors have undoubtedly become cross-disciplinary tools that find many applications in Life Sciences and have a constantly growing market. Expression vectors are artificial nucleic acid molecules with a modular structure that allows the transcription of DNA sequences of interest in either cellular or cell-free environments. The *cis*-regulatory sequences (CRSs) that control the transcription in expression vectors are usually sourced either from viruses or from characterized genes. However, the recent progress made on transposable elements (TEs) offers attractive alternatives that could enable significant advancement in the design of expression vectors.
TEs are mobile genetic entities and constitutive components of both eukaryotic and prokaryotic genomes. TE-related CRSs allow the regulated transcription of transposition-related genes. However, some of them display striking characteristics, providing the opportunity to reconsider TEs as lead actors in the design of expression vectors.
Here, we provide a synthesis of the transcriptional control elements commonly found in expression vectors together with an extensive discussion of their advantages and limitations. We also highlight the latest findings that could make the implementation of TE-derived sequences in expression vectors feasible, with a possible improvement of the existing vectors. By introducing a new concept of TEs as a source of regulatory sequences, we aim to trigger a profitable discussion of the potential advantages and benefits of developing a new generation of expression vectors based on the use of TE-derived control sequences.

Keywords
Promoter; enhancer; insulator; silencer; expression vectors; transposable elements; species-specificity; patenting; evolution

List of abbreviations
EV: expression vector; TSS: Transcription Start Site; TE: Transposable Element; CRS: Cis-Regulatory Sequence; Chip-seq: Chromatin immunoprecipitation-sequencing; NRE elements: Negative Regulatory Elements; TF: Transcription Factor; IRES: Internal Ribosome Entry Site; CMV: Cytomegalovirus; UTR: Untranslated Region: LINE: Long Interspersed Nuclear Element; SINE: Short Interspersed Nuclear Element.

Word Count: 10152

INTRODUCTION

Biotechnology is undoubtedly one of the most rapidly evolving fields of science, which is mainly driven by the high demand in a large number of scientific and economic sectors.

The practices of expressing recombinant proteins in the preferred cellular system and transgenic organisms have become increasingly important in many fields of Life Sciences, especially in the biomedical field, with the purpose of try explaining many relevant biological issues, related or not to the human health, and as a primary method for the production of recombinant proteins in the Bio-Pharma industry. Moreover, the protein synthesis industry has a great impact on human lives because it enters in the production processes of everyday items, such as clothes, foods, beverages, and personal care products[1]. According to the most important market research and consulting companies, the demand for expression vectors (EVs) that efficiently grant the transcription of the transgene is constantly growing (appraisal by https://www.marketreportsworld.com) suggesting that the recombinant protein expression systems market is growing up as an important economic sector.

In the biomedical field, the demand for novel expression host platforms increases with the number of genes that can be targeted by drugs produced by the Pharma industry. In the past years, tens of reviews have covered the topic of the expression systems, looking at both the host and the vector sides with great detail, which collectively gathered thousands of citations. Yet, in the field of recombinant protein expression, progress is continuously being made. In this review, we comment on the possible ways to further improve an apparently perfect system based on decades of studies on the regulation of prokaryotic and eukaryotic gene expression, through the application and integration of information coming from studies concerning transposable elements (TEs).

Expression systems: an overview of vectors and hosts

Expression systems enable the expression of gene products with the aid of a recombinant expression vector. A great variety of expression systems are currently available to the research community, which makes it difficult to uniquely classify them into few categories. A schematic overview is given in Table 1, in which the expression systems are catalogued using different criteria.

TABLE 1 Schematic overview of expression systems.

By vector type	By promoter type	By presence of additional control sequences	By host type	By End User	By Application
Plasmidic/ non-integrating	Constitutive Weak Strong Type II* Type III*	Enhancer	**Bacterial cells:** **E.coli (G-)** **Bacillus spp (G+)** **L.lactis (G+)** **Pseudomonas spp** **Corynebacterium spp (G+)**	Pharmaceutical And Biotechnology Companies	Laboratory Use
Viral/trans poson-based	inducible	Insulator	**Animal cells**: insect Mammal	CROs and CMOs **	Production of Peptide and Protein Pharmaceuti cals
Minicircle-based		Repression/Activat ion	Plant cells	Academic Research Institute	Industrial production of enzymes
		Integration	**Fungi**: *S. cerevisiae* *K. lactis* *P. pastoris*		Transgenic Plant and Animals
			Transgenic-in vivo		Gene Therapy
			Cell-free		

Table legend. G+: Gram positive bacteria. G- Gram negative bacteria. (*) Type II and Type III promoters are responsive to RNA polymerase II and III respectively. Type I promoters are not used in the design of EVs. (**) CROs (Contract Research Organizations) and CMOs (Contract Manufacturing Organizations) are two key types of companies who support the pharmaceutical, biotechnology, and medical devices industries in the form of outsourced services.

EVs are DNA molecules with either linear or circular topology that need to be either introduced into a cellular environment or to be put in contact with a cell-free (typically a cell extract) to express the desired gene product. As a general rule, EVs are designed as small as possible in size in order to reduce the metabolic load of the cells into which they are introduced and, in therapeutic applications, to reduce the harmful effect of non-essential sequences. In addition to the usual features of standard cloning vectors - i.e. a replication origin, a selectable marker, and a multiple cloning site - EVs contain sequences that allow the transcriptional control and the early steps of the translation of the cloned sequence, called *cis*-regulatory sequences (CRSs) after their in *cis* mode of action. Furthermore, EVs contain sequences that make the transcript susceptible to post-transcriptional modifications (such as polyadenylation) and to translation, protein localization and post-translational modifications.
CRSs are usually characterized when studying genetic loci of interest and next they might be used to design EVs. The representative CRSs involved in gene control are depicted in Figure 1 and will be described and discussed, in the context of the EVs, in the next paragraphs.
The choice of the host for expression is a critical step that affects not only the amount of expressed gene product but also the easiness and the quality of the purification, and the biological activity of the purified product.
E. coli is undoubtedly the workhorse of protein expression and represents the easiest, quickest, and cheapest host for protein expression. The rapid growth rate, the low culturing costs, and the availability of commercial and non-commercial EVs along with many strains optimized for *ad hoc* applications make *E.coli* the host of choice in many instances. However, *E.coli* suffers limitations when expressing eukaryotic proteins that require proper folding or post-translational modifications to gain their biological activity. Other limitations come from differences in the codon usage between eukaryotes and prokaryotes, and protein toxicity in the host.
Mammalian cells are frequently used as hosts for the expression of foreign genes especially in recombinant drug production, because of the requirement of specific post-translational modifications that enable the biological functions of the ectopically expressed protein. Other cellular platforms, both eukaryotic and prokaryotic, can be routinely used in place of the mammalian cells.
Alternative cellular systems, which include various types of bacteria, yeast, insect, and plant cells, become the preferred choice when a high yield of gene product is required and when there is no need for extremely specific post-translational modifications[2].
Finally, transgenic organisms, in which the transgene is integrated into the host chromosomes and expressed following the desired pattern, are currently used to produce recombinant proteins, While this field has been extremely boosted by the recent development of new genome editing technologies, it requires highly sophisticated levels of technologies and highly regulated protocols and manufacturing especially when recombinant proteins are intended for therapeutic purposes[3].

Applications and production of EVs.
In principle, EVs find applications whenever a recombinant gene product is needed. Based on Table 1, EVs find at least four types of applications. The introduction of an EV into our favorite host is the most common method to produce and isolate recombinant proteins or to discover the subcellular localization of the protein of interest, or to localize the gene product into specific subcellular compartments. In many cases, gene expression is performed in the laboratory to purify the gene product

that researchers will use for subsequent applications (quantification, sequencing, physical and chemical characterization, and other). The use of EVs in Pharmaceutical applications aims to produce gene products (more frequently protein products) that will be used in the preparations of drugs in which the active compound is not the EV itself. A similar use of EVs is made in Industrial applications that have the purpose of producing enzymes (mainly proteases) that are added to the final product (e.g. foods and detergents)[4,5]. It is worth noting that this specific field of Biotechnology mostly explores new approaches for the production of proteins of interest, since custom protein synthesis is 50+ billion US dollars industry[6].

The production of genetically modified organisms (GMOs) or "biotechnologically modified organism" as redefined by Oliver[7] makes extensive use of EVs. GMOs, either plants, animals, or microorganisms are produced with different purposes, but always an integrating expression cassette is needed for the stable expression of foreign genes. Again, GMOs production is performed either in research laboratories, with the purpose of modeling biological phenomena into the favorite organism, or at an industrial scale, to produce foods[8] and crops[9]. Genetic modification of germ cells, zygotes, embryos (animal GMOs), and somatic (plant GMOs) cells is at the basis of GMOs production. Gene therapy is a pharmacological application in which the EV is the therapeutic agent. It is also a particular application of transgenesis since it does not apply to germ cells. Indeed, gene therapy concerns somatic cells that should replenish the weak tissue(s) with the right or the right amount of gene product that is lacking upon the onset of a genetic-based disease.

Different quality grade levels should be considered and achieved when producing vectors[10], depending on their downstream use.

The use of molecular vectors in therapeutic and pharmaceutical applications requires the permission of the local regulatory agencies. These rules are not equal in different countries, but they are quite similar, harmonized and oriented to protect individuals from any potential risk associated with their use.

The quality of vector preparations ranges from Research grade to Good Manufacturing Practice (GMP) grade[10]. Research grade or Research grade animal-free vector preparations contain mixed topological forms of plasmid DNA, whereas ccc-Grade contains only covalently closed circular supercoiled DNA.

HQ Grade plasmids are compliant to the production of GMP grade of active pharmaceutical ingredients like recombinant viruses, proteins, antibodies, and RNAs. GMP Grade vectors are needed every time they are used as active pharmaceutical ingredients. These are highly purified vectors produced in accordance with the EM(E)A guidelines CHMP/BWP/2458/03 and CPMP/BWP/3088/99. Methods of GMP grade plasmids production, from small[11] to large scale[12], are available.

Control of gene expression and the CRSs in EVs design

The typical structure of prokaryotic and eukaryotic transcriptional units with the respective CRSs are depicted in Figure 1A-B. Currently available EVs are the result of many years of experimental studies on gene functioning and integrate various types of CRSs. The organization of CRSs in a generic EV is depicted in Figure 1C. These modules can be hardly found all in the same vector, since each EV has its own configuration, depending on the purpose for which it was created.

The promoter is undoubtedly the most important component in transcriptional control, which deserved many of studies (over 258000 items in PubMed – last accessed June 2020). The promoter and the transcriptional terminator sequences enable an efficient initiation and completion of the transcription, respectively. The

promoter ensures that the transcription is initiated in a reiterated way and with a certain frequency at a precise position within the transcriptional unit, the transcriptional start site (TSS). This feature is used to classify promoters as weak (low transcription efficiency), strong (high transcription efficiency) or inducible (activated by organic or inorganic molecules). The promoter performances rely on the interaction with the RNA polymerase and their associated transcription factors. While bacteria possess a single RNA polymerase enzyme[13], eukaryotes have multiple RNA polymerases, which account for the transcription of specific subsets of genes[14]. The specificity of prokaryotic RNA polymerase is determined by the interaction with multiple forms of the σ subunit, which, in turn, enables the transcription of specific operons. Similarly, the eukaryotic RNA polymerases gain specificity through the association with a plethora of transcription factors that determine the ability to recognizing and binding to specific promoter elements[15].
The choice of the promoter in EVs designing is almost based on its strength and its performance in the preferred host, depending on the application. When aiming at high yield of the gene product, a proper promoter should be strong, with the caveat that any unfavorable impact, like toxicity or cell growth impairment should be avoided. The characterization of many promoters in model organisms such as *S. cerevisiae*[16], *E. coli*[17], and in mammalian cells[18] has increased the availability of promoters with a wide range of potency. The promoter choice also depends on the gene product type. mRNAs should be expressed from RNA polymerase II responsive promoters, whereas small RNAs should be expressed from RNA polymerase III responsive promoters. Inducible and repressible promoters offer the option to control transcriptional initiation. Natural or designed stimulus-responding TFs allow either to turn the expression from off-to-on state (inducible promoters) or vice-versa (repressible promoters). These TFs can respond either to physical stimuli (light or temperature) or to ligands[19,20]. An alternative induction strategy relies on the use of binary systems, such as the GAL4/UAS from *S. cerevisiae*[21] and the Q system of *N. crassa*[22]. They consist of a promoter and the specific TF(s) it binds, which can be introduced in the expression system either together or independently and subsequently combined by genetic crossing. These systems are quite more complicated than simple expression systems in cultured cells and are mainly used to drive gene expression in a highly tuned way in transgenic organisms[23].

On the other hand, the transcriptional terminators ensure an efficient transcript 3'-end formation that affects the stability of the RNA molecules and influences the translation efficiency. Two main termination mechanisms are known in prokaryotes whereas the three eukaryotic RNA polymerases adopt different termination mechanisms[24]. Transcriptional terminators are therefore essential component of EVs.

Additional CRSs regulate the expression of eukaryotic genes. Prokaryotes have indeed a smaller repertoire and a simpler organization of their CRSs compared to their eukaryotic counterparts (Figure 1 A). Additional features can indeed be located in a typical eukaryotic transcriptional unit to determine a precise gene expression pattern (both developmental- and tissue-specific), and to establish the transcription and translation efficiencies. Examples are enhancers, insulators, splice sites, and translational recoding sequences.

Enhancers are distal regulatory sequences, located outside of the proximal promoter regions, which increase the expression of genes through their interaction with transcription factors. Enhancers can be found up to 1 Mbp distant from the genes they control[25]. The DNA folding can place two regulative sequences (enhancers and promoters) in close proximity to each other, thus explaining the enhancers' action *in*

trans[26]. Enhancers also activate transcription irrespective of their position (upstream, downstream, and within the transcribed sequence) and orientation (sense or antisense) relative to the TSS. Although their best mode of action is *in cis*, enhancers can also act *in trans*, a phenomenon known as transvection[27,28]. Transvection allows activation of a promoter from the enhancer located on the homologous chromosome, allowing the expression of the gene in trans-heterozygous null mutant alleles[28].
Usually, the enhancers found in EVs are of viral origin, such as that of the herpes virus, Rous sarcoma virus, hepatitis B virus, SV40 virus and the enhancer of the human CMV (hCMV). The latter has the highest activity in several mammalian cell types[29], whereas its simian counterpart (sCMV) is used to enhance gene expression in amphibians[30] and fishes[31].

Silencers are orientation and position-independent elements that have a down-regulation function in gene expression[32]. The silencer function is distinct from that of the NRE elements (Negative Regulatory Elements), which are position-dependent and act in a specific context through a passive repression mechanism[33]. While NREs recruit TFs (repressors), which negatively affect the formation of functional Pol II transcription complexes[34], silencers establish an inheritable repressive chromatin state. Although a variety of silencers have been characterized so far, they constitute an overall under-studied class of CRSs (2458 PubMed items using "Silencer" as search keyword – last accessed June 2020). Furthermore, while silencers have not found application in the design of EVs, NREs have had more success, and in this sense, the eukaryotic NREs function like the prokaryotic repression systems at the operator sites. A widely used repression systems, is based on the tetracycline operator (tetO) which has been modified and improved for its function in eukaryotic cells[35].

Insulators are important sequences that organize chromosomes into distinct domains in which gene expression occurs in a coherent way and into which genes acquire specific spatio-temporal expression patterns. Insulators prevent enhancer to be hijacked out of the chromosomal domain of competence, an effect that could cause severe phenotypes[36,37]. The effect of insulators on gene expression is exerted through their enhancer-blocking activity that disrupts the enhancer-promoter interaction, mediated by the CTCF transcriptional repressor, a highly conserved and ubiquitously expressed protein in eukaryotes[38]. Insulators are often used in the design of expression cassettes that are meant to integrate into the genome to prevent the transgene silencing from the genomic position effect[39].

Translational recoding signals (stop-codon readthrough, ribosome frameshifting, and translational bypassing) can be found both in eukaryotic and prokaryotic genes and allow the ribosome to bypass stop codons, to change the reading frame on the RNA, and to synthesize one protein from two discontinuous reading frames respectively[40].

Some EVs offer the possibility to express two or more sequences in equimolar amounts, through the use of internal ribosome entry sites (IRES), an organization recalls the bacterial operon. Many of these sequences have viral origin, like the IRES of the encephalomyocarditis virus (ECMV) that is widely used in EVs.

Species-specificity of CRSs limits the use of EVs.
How can we improve the currently available EVs? The features of the currently available EVs alongside with the features of a hypothetical ideal EV are reported in Table 2.

Table 2. Desirable improvements to the currently used EVs

	CURRENT VECTORS	IDEAL VECTORS	BENEFITS
Host specificity	High	Null or very low	Reduced time and costs of experimental setup
Modulation/induction of transcription	Depends on the vector	Integrated feature	Optimization of experimental conditions
Leakiness	Often present	Absent	Straightforwardness of expression results
Episomal or chromosomal-integrated	Separate features	Integrated features	Switch between stable or transient expression
RNA polymerase dependency	RNApol specific	RNApol independent	Transcription optimal for all transcript types (mRNA, siRNA…)
Insulation of expression cassette	Optional	Integrated feature	Protection from position effect

As described above, all the features found in EVs are taken from standardized and well-characterized *naïve* expression systems in a given species, which make a given expression system useful for testing in that species only, or in a very narrow range of organisms. Such stereotyped workflow has the disadvantage that each time researchers have to change their cellular platform (e.g. from bacteria to eukaryotic cells) they also need to change the EV.
The promoter species-specificity is recognized as an issue in heterologous gene expression since the late '70. Helinski reported that the "[…] molecular barriers at the transcription and translation levels will greatly reduce or prevent the expression of cloned eukaryotic genes in a prokaryotic (bacterial) cell or cloned prokaryotic genes in a eukaryotic cell as plant cells. These barriers [….] are most likely responsible for the very limited success in the expression of cloned eukaryotic genes as intact and functional proteins within the bacterial cell"[41].
The transcriptional bottleneck is largely due to the mode of evolution of promoters, their associated TFs, and the RNA polymerases. Although promoters are heterogeneous in sequence, even in the same species[42], sequence divergence of the promoter could easily disrupt the ability to start the transcription[43]. Additional factor affecting the promoter function are the DNA secondary structure[44] and the epigenetic environment that wraps the locus[18]. All these features are the result of the coevolution between promoters and other coding and non-coding sequences in the respective genomes, which limits their use in evolutionarily distant types of cells.
It follows that the search of unconstrained promoters to be used in EVs design is a hard task.
In this context, one of the most promising promoters characterized so far is the CaMV promoter. CaMV is a virus of the plant family Caulimoviridae with a tropism limited to the Brassicaceae and Solanaceae. The 35S CaMV promoter is a strong constitutive plant promoter[45] also able to drive transgene transcription in a variety plants cells outside the host range of the virus[46], and in completely unrelated cells such as bacteria, in fungi, and in several higher vertebrates cells[47]. Unfortunately the 35S CaMV promoter represents an exception to the rule, and the molecular reasons of this peculiar behavior are still unreported. Contrarily, many, if not all, the promoters that are currently used in EVs have a transcriptional activity in a very limited host range, a limitation, which does not depend on the strength of the promoter[48].
Finally, it is worth noting that other CRSs also have the same limitation described for the promoters. Enhancers seem to have a stronger species preference. Few reports suggest that an enhancer from a given species could preserve its function in a different one, and, to our knowledge, no reports at all describing such a wide-species range of action of TE-linked enhancer. Early studies on the SV40 enhancer suggested that enhancers have strict cell and host specificity[49], and more recently this was confirmed in studies involving flies and nematodes[50]. To the best of our knowledge, the only cross-species enhancer reactivity reported so far includes the *T. castaneum nub* wing enhancers, which cross-react *in vivo* in *D. melanogaster*[51], and several monocot plant enhancers, which cross-react in *N. benthamiana in vivo*[52].
It can be concluded that new CRSs should be searched to expand the use of the current EVs.

Identifying novel CRSs: where and how
The simplest solution for the identification of new CRSs with a wider spectrum of activity is to search and characterize more regulatory sequences. However, identifying novel promoters useful for creating new EVs can be a challenging task. Indeed, while

all promoters have similar function (i.e. they all drive the transcription), there is a clear lack of sequence similarity when different promoters of the same species are compared. This lack of similarity is exacerbated when promoters that drive the expression of orthologous genes in distant species are compared. For this reason, the identification of new regulatory sequences using similarity-based approaches could not be such an effective strategy in finding new promoters, as well as other CRSs.
The available methods to identify new regulatory sequences are based on computational and experimental approaches. The computational strategies are based on the identification of small DNA modules that characterize the individual functional sequences, and the association of a prediction score suggests the reliability of the predictions[53].
However, the most effective method to test CRSs is to assay them functionally. Sequences can be either randomly assayed from a sheared genome, resulting in high throughput screenings or systematically picked from a subset of selected annotated sequences[54]. Massive identification of enhancers and promoters can be also performed using more sophisticated methods and technologies such as Chip-Chip, and Chip-seq assays and on chromatin accessibility either to DNase[55] or to transposase (ATAC-seq)[56] that are also based on the structural and the epigenetic status of these sequences *in vivo*. Functional assays of promoters make often use of promoter-less reporter gene cassettes that express an assayable protein product, thus providing an indirect measurement of the promoter strength in the experimental cellular system used.

Identification of novel CRSs: insights from transposon biology

A special subset of genomic sequences in which new types of CRSs can be searched is the mobilome, represented by the entire set of mobile sequences in a genome. Transposable Elements (TEs) are important contributors to genome complexity [57], evolutionary variation[58] and environmental adaptation[59].
After their discovery in prokaryotes[60,61] and in eukaryotes[62], TEs were long considered as useless sequences. Lately, TEs were re-discovered as powerful and essential evolutionary drivers in all living organisms[63]. Indeed, TEs have been mainly studied for their power to jump from a DNA molecule to another, and successfully implemented in insertional mutagenesis protocols[64]. This hallmark soon made TEs the top candidates for special applications such as gene therapy[65,66], also considering the non-viral origin of these sequences that render TEs safe tools with respect to many therapeutic applications[67].
TEs move in the host genome using either the replicative or “copy & paste”, or the conservative or “cut & paste” mechanisms. Prokaryotic transposons are more heterogeneous, and form three groups on the basis of their structural organization[68], (Figure 2A). Two main Classes of eukaryotic TEs are defined based on their structural and biological characteristics[69] (Figure 2B).
Class I TEs or retrotransposons, replicate via an RNA intermediate, that is converted to cDNA and integrated in a new genomic site by the Reverse Transcriptase, RNAseH and Integrase enzymatic activities (Class I TEs or retrotransposons). Class II, or DNA transposons, TEs move directly by excision and integration of the element in a new genomic locus, mediated by the transposase, with some exceptions[70]. It is worth mentioning that no equivalent of RNA-based TEs have been found in prokaryotes[71].
Usually, both autonomous (i.e. transposition competent) and non-autonomous elements coexist within the same genome. In some instances, non-autonomous TEs can be over-represented[72,73]since they can be trans-mobilized by hijacking the transposition machinery of active elements[74].

TEs carry sequences whose primary function is to regulate the transcription of transposition-related genes. However, there is growing evidence that TEs can be adopted as CRSs, donating enhancers, promoters, and insulators to the host genome[75] and contributing to rewire existing gene regulatory networks[76]. Model and non-model organisms can provide many examples of CRSs borrowed from TEs to both simple and complex genetic loci that account for a variety of adaptive and evolutionary events[77,78]. As described below, a number of TE-linked *cis*-acting sequences have been characterized so far, that could be relevant to developing a new generation of EVs.

Relevant CRSs within TEs

In all the above-described Classes of TEs, both eukaryotic and prokaryotic, the most interesting portions where searching CRSs are the terminal segments, and in particular their 5' region. Indeed, the LTR and the 5'UTR of LTR-retrotransposons, the 5'UTR of LINEs, and the 5' region of TIRs transposons contain acknowledged *cis*-functions. While the main function of the 5' terminal portion of TEs is related to the recognition of the integration machinery, it also contains all the CRSs primarily needed to express the self-encoded transposition-related enzymes and regulators. Surprisingly, TEs contain the most diverse CRSs, for which a description will be provided in the next paragraphs.

Promoters are naturally found in TEs to transcribe genes encoding the transposition machinery, comprising multiple enzymatic activities of Class I TEs, or the single transposase gene of Class II TEs, with few exceptions. In non-autonomous elements the promoter does not significantly impact transposition but it could be used as an evolutionary substrate to shape or modify the transcription pattern of nearby genes. Also in bacteria, the promoter of IS elements can alter the expression of adjacent genes. The IS promoter is either entirely enclosed within the 5' of the element or it can be created upon insertion as a chimeric flanking/transposon sequence, and its action can be oriented either towards the right or the left end of the element[79]. Several efficient promoters have been identified and characterized so far in many types of TEs, especially in model organisms [77,78].

Increasing evidence suggests that endogenous genomic enhancers are the result of TE exaptation[75,80], indicating a natural inclination of these elements to evolve as enhancers. Indeed, endogenous enhancers have been described within elements belonging to both Classes of TEs[77,78] and belonging to the most diverse organisms. Model organisms frequently provide examples of how strong TE-derived enhancers are, and to what extent they can modify the expression of endogenous genes[81]. Recent results from epigenomic and transcriptomic analyses have suggested a role for TEs in establishing tissue-specific gene expression in mouse embryonic and trophoblast stem cells[82].

The presence of silencers in TEs has been disclosed firstly in the 5'UTR of the Drosophila *gypsy* retrotransposon that can switch its function from an insulator to a silencer[83] in a host mutant genetic background for the *modMdg4* gene[84]. Thereafter, a sequence acting as silencer has been identified in the *mariner* element *Mos1*, and it was found to be conserved in other *mariner* elements identified in other species[85]. Interestingly, a transcriptional repression action has been associated with young full-length *L1* elements located within introns of active genes. These intronic L1 copies are the preferential targets for H3K9me3-deposition[86] by the Human Silencing Hub Complex[87]. This kind of transcriptional repression could be regarded as the result of L1-mediated silencer as suggested[78].

TE-linked insulators have been identified in a variety of human TEs, such as the human MIRs[88], the murine SINEB2 elements[89], and several Drosophila LTR-retrotransposons[77]. Additionally, functional and evolutionary analyses have revealed that repeated bursts of retrotransposition have contributed to expand the number of CTCF binding sites in six mammals[90].

Retrotransposons and endogenous retroviruses make large use of translational recoding, such as programmed frameshifting and IRES, to express their transposition-related genes. The presence of programmed frameshifting signals between the Gag-Pol genes is a common strategy adopted to express from the same promoter the Pol product, which is on a different frame with respect to the Gag gene. The stop codon read-through is an alternative strategy to express Pol protein when the Gag-Pol genes are on the same frame and are separated by a stop codon.

The programmed frameshift has also been demonstrated to be an efficient strategy to express the transposase of many prokaryotic IS elements that exhibit two consecutive overlapping ORFs placed on different reading frames [91].

Special features of TE-related CRSs enable the improvement of EVs

The specific research in the field of TE biology has disclosed unprecedented features of their CRSs that allow their functioning in multiple hosts, even if evolutionarily far from the species of origin. This feature, although very rare and uncommon, concerns promoters, insulators, and silencers.

The ability of a promoter to drive a reporter gene expression in evolutionarily distant hosts has been formerly reported for the CaMV 5S gene promoter (see above paragraphs). This peculiarity has been identified for the first time in TEs when the promoter of the Drosophila *Bari* transposons was characterized. The *Bari1* transposon from *D. melanogaster*[92] and *Bari3* transposon from *D. mojavensis*[93,94] are members of the *IS630/Tc1/mariner* superfamily, which are among the most widespread DNA transposon among extant organisms[95,96]. Members of the *Bari* family bear promoters that are able to drive the transcription of a reporter gene in non-insect cells as distant as human, yeast, and bacterial cells[97]. Notably, other members of the same superfamily isolated in other organisms share this considerably unusual feature[98], suggesting an evolutionarily conserved aptitude of these TEs to function in unrelated host genomes. The promoters of the *Tc1/mariner* elements that showed this feature was named “blurry promoters” to recall that they do not “sharply” activate transcription within a single or few species[97].

The identification of TE-related CRSs displaying an extended activity over multiple distantly related species also involved insulators and silencers. The insulator identified in the 5’UTR of the *ZAM* retrotransposon in Drosophila showed the ability to block the enhancer-promoter communication in human HEK293 cells[99], while the Drosophila silencer identified in the *Mos1* transposon can suppress transcription in human cells [85].

Where do the special features of TE-derived CRSs come from? An evolutionary perspective

Why some TEs and some viruses share the unique feature of CRSs that function across multiple organisms is still an unanswered question. Our current knowledge in Biology and Molecular Biology covers less than 20% of living organisms[100], which is mostly focused on few experimental model systems. The more we will look at non-model organisms, the more we will probably find surprisingly novel biological phenomena.

Similarly, we still have not fully understood how TEs evolve, an important issue that can let us decipher. TEs are ancient genomic components[63], and their ancient origin could explain why they are found virtually in all living organisms. While the evolution of TEs has parallelized that of the genomes in which they reside, TEs also gained independent evolutionary trajectories due to their ability to jump from a species to another, through horizontal transfer (HT).
HT is a non-mendelian way of transmission of the genetic material. DNA can be passed from a species to another, through a direct or indirect transfer mediated by one or more vectors (either uni- or multi-cellular organisms or viruses). Virtually, all types of DNA sequences (coding and non-coding) can be transferred horizontally and the participation of vector organisms in the HT process allows overtaking the ecological and geographical barriers existing among species involved in HT. It has been proposed that the invasion of new genomes grants TEs to escape extinction[101]. With this premises, a key question in HTT concerns how and to what extent the success of TEs in a new genome depend on their endogenous CRSs functioning. Indeed, the expression the transposition-related gene(s) in a new genome that evolved its CRSs/TFs independently from the genome of origin, appears a critical issue.
In this respect, it has been recently shown that the promoters of *Tc1/mariner* transposons are somewhat able to break the barriers existing between the transcriptional processes in different Kingdoms[97,98]. This peculiarity has been linked to the ability of these TEs to spread into new, unrelated, species' genome via horizontal transfer compared to other eukaryotic mobile sequences[102]. The *Tc1/mariner* "blurry promoter" would allow colonizing easily the new genome overcoming the transcriptional barriers that the divergent evolution between organisms has built. It is therefore possible that other TEs have evolved similar CRSs with the same evolutionary purpose. Further and deeper investigations concerning the TE-linked CRSs are required to identify such peculiarity.

TE-derived CRSs currently used in EVs

Excluding vectors expressing transposition related proteins as part of transposition systems, several EVs make use of TE-derived regulatory sequences. Table 3 reports a list of such vectors. The pCoBlast vector has been successfully commercialized as a part of expression kits (Promega), highlighting the feasibility of TE regulatory sequences to be implemented in successful expression systems. Although the commercial success of EVs based on TE-related CRSs seems to be limited to the pCoBlast vector, many non-trademarked vectors that ground the expression of the resistance genes on the *copia* promoter can be found in the Addgene vector repository. Vector repositories (such as Addgene, Riken, DGRC, DNASU Plasmid Repository) are specialized vector banks that help in storing, distributing, and sharing recombinant vectors, a service solicited by the exponential growth of recombinant vectors design and construction and the necessity to make new tools immediately available to other researchers. The birth of gene synthesis services has extremely increased the need for such nonprofit repository.
In addition to the *copia* promoter, many human TE-derived promoters have been shown to drive the expression of reporter genes in recombinant vectors. Several HERV-type promoters have been identified and experimentally characterized so far [103-105], and some show a significant cell-type specificity[103].

Table 3. Non-commercial EVs that make use of TE CRSs.

Name	Source	Description	Reference
pDrBB2	AG #18946	copia-hygromycin resistance cassette	[106]
pCoPURO	RK RDB08531	copia-puromycin resistance cassette.	[107]
Copia-Renilla	AG #38093	copia-renilla luciferase cassette	[108]
pRB31	AG #72868	copia-Puromycin resistance cassette	[109]
pRB32	AG #72863	copia-Blasticidin resistance cassette	[109]
pSK32	AG #72854	copia-Puromycin resistance cassette	[109]
pRB39	AG #72870	copia-Puromycin resistance cassette	[109]
pRB25	AG #72853	copia-Puromycin resistance cassette	[109]
pRB40	AG #72865	copia-Blasticidin resistance cassette	[109]
pRB38	AG #72864	copia-Blasticidin resistance cassette	[109]
pSK24	AG #72852	copia-Puromycin resistance cassette	[109]
pRB37	AG #72869	copia-Puromycin resistance cassette	[109]
pRB34	AG #72862	copia-Blasticidin resistance cassette	[109]
pSK15	AG #72855	copia-Blasticidin resistance cassette	[109]
pKF297	AG #74774	copia-Puromycin resistance cassette	[109]
pRB33	AG #72867	copia-Puromycin resistance cassette	[109]
pSK23	AG #72851	copia-Puromycin resistance cassette	[109]
pRB35	AG #72866	copia-Puromycin resistance cassette	[109]
pSK41	AG #74886	copia-Puromycin resistance cassette	[109]
pRB36	AG #72861	copia-Blasticidin resistance cassette	[109]
pKF296	AG #74773	copia-Blasticidin resistance cassette	[109]
pBID-UAS-series	AG #35199	2 Gypsy insulators flanks the cloned gene	[110]
AByG	AG #111083	Gypsy insulator CRISPR/Cas construct	[111]
pJFRC2-INS	AG #26215	Gypsy insulators flanks the cloned gene	[112]
pGL3-Ba1p	NA	Multi-species blurry promoter	[97]
pGL3-Ba3p	NA	Multi-species blurry promoter	[97]
pGL3/SBp	NA	Multi-species blurry promoter	[98]
pGL3/Hsmar1p	NA	Multi-species blurry promoter	[98]
pGL3/UnGE	NA	Multi-species blurry promoter	(Palazzo & Marsano unpublished)
pGL3/PBP	NA	Multi-species blurry promoter	(Palazzo & Marsano unpublished)

Table legend. AG: Addgene; RK Riken; NA; not deposited in vector repositories.

A brief survey of the EVs market and its related patents.

According to the most outstanding market research and consulting companies, the global recombinant DNA technology market is expected to reach up to 850 billion USD by 2025. The primary end-users of the EVs market are Pharmaceutical and Biotechnology Companies, Contract Research Organizations (CROs), Contract Manufacturing Organizations (CMOs), and Academic Research Institutes, which continuously contribute to the expansion of this kind of market. This huge success is owed to a large number of applications of such techniques in the widest fields of Applied Sciences, such as genetically modified organisms (microorganisms, crops, and animals), biofuels production[113,114], Biopharma, gene therapy, and basic research. The latter fields have been tremendously boosted in the last years by the introduction of precision gene and genome editing tools[115].

Patenting is directly linked to the market expansion since patents are exceptional means to establish commercial partnerships and exploit the market.

Protection of the intellectual property of newly discovered CRSs, especially those associated with unique transcriptional features, is witnessed by the plethora of patent applications and publications. On average, 197 patents/year have been published over the last 10 years (2010-2019). Considering promoter-related patents only, the major pulse come from Asian Countries (1444 patents), followed by North American Countries (286), and other Countries (Figure 3).

Notably, 13 patents apply to TE-derived regulatory sequences, including 8 patents related to TE-promoters, 2 to insulators, 1 enhancer, and 1 to IRES (Table 4).

It is noteworthy that two patents in this list (EP2772539 and US AN16/715,451) claim that a new feature discovered in two fly transposons can be used to generate EVs with a multi-platform application, while another patent (US7064246B2) claims that the conjugation of previously known CRSs with TE-associated features is useful to obtain new and more performing EVs.

Given their peculiar features, TE-derived CRSs could give an additional and important impulse to the market of EVs and expression systems.

TABLE 4 Patents related to TE-derived CRSs.

PATENT NUMBER	STATUS	YEAR	DESCRIPTION
WO2001005986A2	Issued	2001	DNA transposon promoter
WO2000053789A3	Issued	2001	HERV promoter
WO0222839A2	Issued	2002	IRES of Errantiviruses
JP2002291473A	Issued	2002	Plant Retrotransposon promoter
WO03070958A1	Issued	2003	Insulator in a retroviral gene therapy vector
US7064246B2	Issued	2006	Transposase binding motif fused to a promoter to modulate gene expression
EP1887083	Issued	2012	Retrotransposon-derived insulator
CN104450708A	Issued	2015	Retrotransposon promoter
CN103952404	Issued	2014	MITE transposon enhancer
EP2443255A2	Issued	2017	LTR used as a promoter
EP2772539	Issued	2019	TE “blurry” promoters
US AN16/715,451	Filed	2020	TE “blurry” promoter
US6395549	Issued	2002	LTR CRSs

Concluding remarks and future challenges: could a new generation of EVs be inspired by TE-derived cis-regulatory sequences?

Basic research in non-model systems constantly calls for new expression methods, which rely on the development of new molecular (i.e. vectors) and technological innovations (i.e. transfection and transformation methods).

TE-derived CRSs are promising sources to be considered as a starting point to obtain engineered promoters with improved performances, such as the possibility to perform gene expression in multiple organisms using a single construct.

But is a universal EV conceivable? Such a vector should contain universal CRSs and should be able to drive and control efficiently the transcription in virtually every cellular environment, irrespective of the taxonomic level of the starting and the destination organisms. Given the above-described features of the transcription process in eukaryotes and prokaryotes, we can easily deduce that a reasonable answer to the above question is no. To date, at least three different types of CRSs (promoters, insulators, and silencers) identified in TEs carry out their function in non-native cells, suggesting that TE-derived CRSs represent a promising starting point to develop a "one-for-all" EV. However, much effort is needed to test them in additional cellular systems and tag them as the "magic sequences". Furthermore, the discovery of the blurry promoters teaches us that like sifting out sand on a riverside in search of gold nuggets, digging into a pile of apparently junk annotated and unannotated sequences could probably lead to surprising discoveries.

The systematic search of TE-related promoters and their functional tests in promiscuous conditions, i.e. test CRSs in non-host cell systems, will provide an invaluable source of information. Further investigations on CRSs should also involve poorly studied TE families such as *Helitron* and *Mavericks*. Also, previous descriptive studies of the TE complement in non-model organisms deserve further investigations especially if novel structural variants of TEs have been described[116].

The genome sequencing projects are an alternative and invaluable source of CRSs-focused information. The development of efficient genome sequencing methods, such as those producing ultra-long reads[117], will soon aid in deciphering the sequence of a historically hard-to-sequence genomic compartments, such as centromeric and pericentromeric heterochromatin, where new transcriptional regulation sequences can be characterized. It is indeed well known from studies in model organisms that heterochromatin is not transcriptionally silent[118] and that CRSs that regulate the expression of heterochromatic genes can have peculiar features[119,120].

Finally, a combination of new and known CRSs could be a good halfway to obtain a universal vector useful to express the gene of interest in the experimental system of choice. The availability of gene synthesis services, at affordable costs, would help to generate new combinations of CRSs and easily testing them in various model systems.

An ideal EV is depicted in Figure 4A. Such a vector would give the possibility to stably or transiently express our genes, thanks to the TIRs bracketing the whole expression cassette enabling its integration upon the co-expression of a compatible transposase. This system will be maintained in the repressed status until proper promoter induction. The insulator sequences protect the transcriptional unit from position effect when it integrates into the genome. One or more genes can be expressed in equimolar amounts, in a way dictated by the action of both the promoter and the enhancer. This vector will work irrespective of the cells that are being transfected/transformed, due to the universal nature of the CRSs used. Another possible use of the TE-related CRSs in EVs is related to the expression of selection

cassettes, such as drug-resistance genes. As an example, a blurry promoter will enable the selection of the same EV in many transfected/transformed cells types using a single resistance cassette, thus expanding the possibility to conduct experiments in multiple experimental systems without the need to change the vector backbone (Figure 4B).

The development of an EV with the above-described features will straightforward the research workflow in many laboratories, since it will eliminate the need of cloning and sub-cloning steps required to switch from an expression platform to another. Furthermore, since TE-derived sequences are regarded as "safe" due to their non-viral origin they can find application in gene therapy trials[67].

In conclusion, we propose that multidisciplinary approaches, based on evolutionary, functional, cellular, and molecular biology studies should be undertaken to improve existing expression systems. These improvements will accelerate the timeline of testing, both in the Bio-Pharma industry and in small research laboratories, with a great reduction of the costs associated with the cellular platform changes.

FIGURE LEGENDS

Figure 1. Regulatory sequences and their organization in prokaryotes (A), in eukaryotes (B), and in a generic EVs (C)

Figure 2. Overview of TEs in eukaryotes (A) and in prokaryotes (B).
Symbols used are explained in the box. TNP: transposase gene; RES: resolvase gene; βLAC; β lactamase gene; GAG-POL-ENV: retroviral-like genes; ZN: Zn-finger protein coding gene; REP: replicase gene; HEL: helicase gene

Figure 3. Overview of the patented promoter sequences. (A) Number of patents issued per year. (B) Number of patents distributed by Country. Data sourced from Espacenet (https://worldwide.espacenet.com).

Figure 4. An ideal expression vector containing TE-derived CRSs.
A) Organization of an ideal, host-unbiased, expression cassette to express sequences of interest.
B) Possible use of special CRSs to obtain host-independent selection systems

REFERENCES

1 Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Current Opinion in Biotechnology* **13**, 345-351, doi:https://doi.org/10.1016/S0958-1669(02)00328-2 (2002).

2 Derouazi, M. *et al.* High-yield production of secreted active proteins by the Pseudomonas aeruginosa type III secretion system. *Appl Environ Microbiol* **74**, 3601-3604, doi:10.1128/AEM.02576-07 (2008).

3 Hunter, P. The prospects for recombinant proteins from transgenic animals. *EMBO Rep* **20**, e48757, doi:10.15252/embr.201948757 (2019).

4 Tavano, O. L., Berenguer-Murcia, A., Secundo, F. & Fernandez-Lafuente, R. Biotechnological Applications of Proteases in Food Technology. *Comprehensive Reviews in Food Science and Food Safety* **17**, 412-436, doi:10.1111/1541-4337.12326 (2018).

5 Maurer, K. H. Detergent proteases. *Curr Opin Biotechnol* **15**, 330-334, doi:10.1016/j.copbio.2004.06.005 (2004).

6 Hunt, J. P., Yang, S. O., Wilding, K. M. & Bundy, B. C. The growing impact of lyophilized cell-free protein expression systems. *Bioengineered* **8**, 325-330, doi:10.1080/21655979.2016.1241925 (2017).

7 Oliver, M. J. Why we need GMO crops in agriculture. *Mo Med* **111**, 492-507 (2014).

8 Zhang, C., Wohlhueter, R. & Zhang, H. Genetically modified foods: A critical review of their promise and problems. *Food Science and Human Wellness* **5**, 116-123, doi:https://doi.org/10.1016/j.fshw.2016.04.002 (2016).

9 Kamle, M., Kumar, P., Patra, J. K. & Bajpai, V. K. Current perspectives on genetically modified crops and detection methods. *3 Biotech* **7**, 219-219, doi:10.1007/s13205-017-0809-3 (2017).

10 Schmeer, M. & Schleef, M. Pharmaceutical grade large-scale plasmid DNA manufacturing process. *Methods Mol Biol* **1143**, 219-240, doi:10.1007/978-1-4939-0410-5_14 (2014).

11 Bakker, N. A. M. *et al.* Small-scale GMP production of plasmid DNA using a simplified and fully disposable production method. *Journal of Biotechnology: X* **2**, 100007, doi:https://doi.org/10.1016/j.btecx.2019.100007 (2019).

12 Samuels, S. *et al.* HPV16 E7 DNA tattooing: safety, immunogenicity, and clinical response in patients with HPV-positive vulvar intraepithelial neoplasia. *Cancer Immunol Immunother* **66**, 1163-1173, doi:10.1007/s00262-017-2006-y (2017).

13 Browning, D. F. & Busby, S. J. Local and global regulation of transcription initiation in bacteria. *Nat Rev Microbiol* **14**, 638-650, doi:10.1038/nrmicro.2016.103 (2016).

14 Rosenberg, L. E. & Rosenberg, D. D. in *Human Genes and Genomes* (eds Leon E. Rosenberg & Diane Drobnis Rosenberg) 97-116 (Academic Press, 2012).

15 Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**, 621-637, doi:10.1038/s41580-018-0028-8 (2018).

16 Ottoz, D. S. M. & Rudolf, F. Constitutive and Regulated Promoters in Yeast: How to Design and Make Use of Promoters in S. cerevisiae. *Synthetic Biology*, 107-130, doi:doi:10.1002/9783527688104.ch6

10.1002/9783527688104.ch6 (2018).
17 Panyukov, V. V. & Ozoline, O. N. Promoters of Escherichia coli versus Promoter Islands: Function and Structure Comparison. *PLOS ONE* **8**, e62601, doi:10.1371/journal.pone.0062601 (2013).
18 Andersson, R. & Sandelin, A. Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* **21**, 71-87, doi:10.1038/s41576-019-0173-8 (2020).
19 Kallunki, T., Barisic, M., Jaattela, M. & Liu, B. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* **8**, doi:10.3390/cells8080796 (2019).
20 Doshi, A., Sadeghi, F., Varadarajan, N. & Cirino, P. C. Small-molecule inducible transcriptional control in mammalian cells. *Critical Reviews in Biotechnology*, 1-20, doi:10.1080/07388551.2020.1808583 (2020).
21 Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401 (1993).
22 Potter, C. J., Tasic, B., Russler, E. V., Liang, L. & Luo, L. The Q system: a repressible binary system for transgene expression, lineage tracing, and mosaic analysis. *Cell* **141**, 536-548, doi:10.1016/j.cell.2010.02.025 (2010).
23 McGuire, S. E., Roman, G. & Davis, R. L. Gene expression systems in Drosophila: a synthesis of time and space. *Trends Genet* **20**, 384-391, doi:10.1016/j.tig.2004.06.012 (2004).
24 Lodish, H. *Molecular cell biology*. (2016).
25 Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-1735, doi:10.1093/hmg/ddg180 (2003).
26 Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110-1122, doi:10.1016/j.molcel.2019.05.032 (2019).
27 Lewis, E. B. The Theory and Application of a New Method of Detecting Chromosomal Rearrangements in Drosophila melanogaster. *The American Naturalist* **88**, 225-239, doi:10.1086/281833 (1954).
28 Geyer, P. K., Green, M. M. & Corces, V. G. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *EMBO J* **9**, 2247-2256 (1990).
29 Foecking, M. K. & Hofstetter, H. Powerful and versatile enhancer-promoter unit for mammalian expression vectors. *Gene* **45**, 101-105, doi:10.1016/0378-1119(86)90137-x (1986).
30 Hoppler, S., Brown Jd Fau - Moon, R. T. & Moon, R. T. Expression of a dominant-negative Wnt blocks induction of MyoD in Xenopus embryos. (1996).
31 Hieber, V., Dai X Fau - Foreman, M., Foreman M Fau - Goldman, D. & Goldman, D. Induction of alpha1-tubulin gene expression during development and regeneration of the fish central nervous system. doi:10.1002/(sici)1097-4695(19981115)37:3<429::aid-neu8>3.0.co;2-n (1998).
32 Brand Ah Fau - Breeden, L., Breeden L Fau - Abraham, J., Abraham J Fau - Sternglanz, R., Sternglanz R Fau - Nasmyth, K. & Nasmyth, K. Characterization of a "silencer" in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer.

33 Ogbourne, S. & Antalis, T. M. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* **331 (Pt 1)**, 1-14, doi:10.1042/bj3310001 (1998).
34 Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**, 29-59, doi:10.1146/annurev.genom.7.080505.115623 (2006).
35 Das, A. T., Tenenbaum, L. & Berkhout, B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr Gene Ther* **16**, 156-167, doi:10.2174/1566523216666160524144041 (2016).
36 Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428-434, doi:10.1038/nature13379 (2014).
37 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).
38 Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194-1211, doi:10.1016/j.cell.2009.06.001 (2009).
39 West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev* **16**, 271-288, doi:10.1101/gad.954702 (2002).
40 Rodnina, M. V. *et al.* Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res* **48**, 1056-1067, doi:10.1093/nar/gkz783 (2020).
41 Helinski, D. R. Plasmids as vectors for gene cloning. *Basic Life Sci* **9**, 19-49, doi:10.1007/978-1-4684-0880-5_4 (1977).
42 Roy, A. L. & Singer, D. S. Core promoters in transcription: old problem, new insights. *Trends Biochem Sci* **40**, 165-171, doi:10.1016/j.tibs.2015.01.007 (2015).
43 Anish, R., Hossain, M. B., Jacobson, R. H. & Takada, S. Characterization of Transcription from TATA-Less Promoters: Identification of a New Core Promoter Element XCPE2 and Analysis of Factor Requirements. *PLOS ONE* **4**, e5103, doi:10.1371/journal.pone.0005103 (2009).
44 Gagniuc, P. & Ionescu-Tirgoviste, C. Eukaryotic genomes may exhibit up to 10 generic classes of gene promoters. *BMC Genomics* **13**, 512, doi:10.1186/1471-2164-13-512
1471-2164-13-512 [pii] (2012).
45 Odell, J. T., Nagy, F. & Chua, N. H. Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter. *Nature* **313**, 810-812, doi:10.1038/313810a0 (1985).
46 Fromm, M., Taylor, L. P. & Walbot, V. Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc Natl Acad Sci U S A* **82**, 5824-5828, doi:10.1073/pnas.82.17.5824 (1985).
47 Seternes, T., Tonheim, T. C., Myhr, A. I. & Dalmo, R. A. A plant 35S CaMV promoter induces long-term expression of luciferase in Atlantic salmon. *Sci Rep* **6**, 25096, doi:10.1038/srep25096 (2016).
48 Qin, J. Y. *et al.* Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLOS ONE* **5**, e10611, doi:10.1371/journal.pone.0010611 (2010).
49 de Villiers, J., Olson, L., Tyndall, C. & Schaffner, W. Transcriptional 'enhancers' from SV40 and polyoma virus show a cell type preference. *Nucleic Acids Res* **10**, 7965-7976, doi:10.1093/nar/10.24.7965 (1982).

50 Ruvinsky, I. & Ruvkun, G. Functional tests of enhancer conservation between distantly related species. *Development* **130**, 5133-5142, doi:10.1242/dev.00711 (2003).
51 Lai, Y. T. *et al.* Enhancer identification and activity evaluation in the red flour beetle, Tribolium castaneum. *Development* **145**, doi:10.1242/dev.160663 (2018).
52 Lin, Y., Meng, F., Fang, C., Zhu, B. & Jiang, J. Rapid validation of transcriptional enhancers using agrobacterium-mediated transient assay. *Plant Methods* **15**, 21, doi:10.1186/s13007-019-0407-y (2019).
53 Umarov, R. K. & Solovyev, V. V. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *PLoS One* **12**, e0171410, doi:10.1371/journal.pone.0171410 (2017).
54 Nguyen, T. A. *et al.* High-throughput functional comparison of promoter and enhancer activities. *Genome Res* **26**, 1023-1033, doi:10.1101/gr.204834.116 (2016).
55 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).
56 Cusanovich, D. A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e1318, doi:10.1016/j.cell.2018.06.052 (2018).
57 Erwin, J. A., Marchetto, M. C. & Gage, F. H. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat Rev Neurosci* **15**, 497-506, doi:10.1038/nrn3730 (2014).
58 Biemont, C. & Vieira, C. Genetics: junk DNA as an evolutionary force. *Nature* **443**, 521-524, doi:10.1038/443521a (2006).
59 Chenais, B., Caruso, A., Hiard, S. & Casse, N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* **509**, 7-15, doi:10.1016/j.gene.2012.07.042 (2012).
60 Hedges, R. W. & Jacob, A. E. Transposition of ampicillin resistance from RP4 to other replicons. *Molecular and General Genetics MGG* **132**, 31-40, doi:10.1007/BF00268228 (1974).
61 Shapiro, J. A. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. *J Mol Biol* **40**, 93-105, doi:https://doi.org/10.1016/0022-2836(69)90298-8 (1969).
62 Mc Clintock, B. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**, 344-355, doi:10.1073/pnas.36.6.344 (1950).
63 Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**, 331-368, doi:10.1146/annurev.genet.40.110405.090448 (2007).
64 Ivics, Z. & Izsvák, Z. The expanding universe of transposon technologies for gene and cell engineering. *Mob DNA* **1**, 25, doi:10.1186/1759-8753-1-25 (2010).
65 Ivics, Z. & Izsvak, Z. Transposons for gene therapy! *Curr Gene Ther* **6**, 593-607, doi:10.2174/156652306778520647 (2006).
66 VandenDriessche, T., Ivics, Z., Izsvak, Z. & Chuah, M. K. Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood* **114**, 1461-1468, doi:10.1182/blood-2009-04-210427 (2009).

67 Hudecek, M. *et al.* Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. *Critical Reviews in Biochemistry and Molecular Biology* **52**, 355-380, doi:10.1080/10409238.2017.1304354 (2017).
68 Chandler, M. Prokaryotic DNA Transposons: Classes and Mechanism. *eLS*, 1-16, doi:doi:10.1002/9780470015902.a0000590.pub2
10.1002/9780470015902.a0000590.pub2 (2017).
69 Kapitonov, V. V. & Jurka, J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* **9**, 411-412; author reply 414, doi:10.1038/nrg2165-c1 (2008).
70 Li, Y. & Dooner, H. K. Excision of Helitron transposons in maize. *Genetics* **182**, 399-402, doi:10.1534/genetics.109.101527 (2009).
71 Eickbush, T. H. & Eickbush, D. G. Transposable Elements: Evolution. *eLS*, doi:doi:10.1038/npg.els.0005130
10.1038/npg.els.0005130 (2006).
72 Bennetzen, J. L. Transposable element contributions to plant gene and genome evolution. *Plant Mol Biol* **42**, 251-269, doi:10.1023/A:1006344508454 (2000).
73 Carbone, L. *et al.* Gibbon genome and the fast karyotype evolution of small apes. *Nature* **513**, 195-201, doi:10.1038/nature13679 (2014).
74 Meyer, T. J. *et al.* The Flow of the Gibbon LAVA Element Is Facilitated by the LINE-1 Retrotransposition Machinery. *Genome Biol Evol* **8**, 3209-3225, doi:10.1093/gbe/evw224 (2016).
75 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18**, 71-86, doi:10.1038/nrg.2016.139 (2017).
76 Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42**, 631-634, doi:10.1038/ng.600 (2010).
77 Moschetti, R., Palazzo, A., Lorusso, P., Viggiano, L. & Marsano, R. M. "What You Need, Baby, I Got It": Transposable Elements as Suppliers of Cis-Operating Sequences in Drosophila. *Biology (Basel)* **9**, doi:10.3390/biology9020025 (2020).
78 Sundaram, V. & Wysocka, J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* **375**, 20190347-20190347, doi:10.1098/rstb.2019.0347 (2020).
79 Vandecraen, J., Chandler, M., Aertsen, A. & Van Houdt, R. The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* **43**, 709-730, doi:10.1080/1040841X.2017.1303661 (2017).
80 Notwell, J. H., Chung, T., Heavner, W. & Bejerano, G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun* **6**, 6644, doi:10.1038/ncomms7644 (2015).
81 Conte, C., Dastugue, B. & Vaury, C. Coupling of enhancer and insulator properties identified in two retrotransposons modulates their mutagenic impact on nearby genes. *Mol Cell Biol* **22**, 1767-1777 (2002).
82 Todd, C. D., Deniz, O., Taylor, D. & Branco, M. R. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *Elife* **8**, doi:10.7554/eLife.44344 (2019).
83 Gause, M., Morcillo, P. & Dorsett, D. Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins.

Mol Cell Biol **21**, 4807-4817, doi:10.1128/MCB.21.14.4807-4817.2001 (2001).
84 Cai, H. N. & Levine, M. The gypsy insulator can function as a promoter-specific silencer in the Drosophila embryo. *Embo J* **16**, 1732-1741, doi:10.1093/emboj/16.7.1732 (1997).
85 Bire, S. *et al.* Mariner Transposons Contain a Silencer: Possible Role of the Polycomb Repressive Complex 2. *PLoS Genet* **12**, e1005902, doi:10.1371/journal.pgen.1005902 (2016).
86 Liu, N. *et al.* Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228-232, doi:10.1038/nature25179 (2018).
87 Robbez-Masson, L. *et al.* The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res* **28**, 836-845, doi:10.1101/gr.228171.117 (2018).
88 Wang, J. *et al.* MIR retrotransposon sequences provide insulators to the human genome. *Proc Natl Acad Sci U S A* **112**, E4428-4437, doi:10.1073/pnas.1507253112 (2015).
89 Lunyak, V. V. *et al.* Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**, 248-251, doi:10.1126/science.1140871 (2007).
90 Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335-348, doi:10.1016/j.cell.2011.11.058 (2012).
91 Chandler, M. & Fayet, O. Translational frameshifting in the control of transposition in bacteria. *Mol Microbiol* **7**, 497-503, doi:10.1111/j.1365-2958.1993.tb01140.x (1993).
92 Palazzo, A. *et al.* Functional Characterization of the Bari1 Transposition System. *PLOS ONE* **8**, e79385, doi:10.1371/journal.pone.0079385 (2013).
93 Palazzo, A., Moschetti, R., Caizzi, R. & Marsano, R. M. The Drosophila mojavensis Bari3 transposon: distribution and functional characterization. *Mob DNA* **5**, 21, doi:10.1186/1759-8753-5-21
1759-8753-5-21 [pii] (2014).
94 Palazzo, A., Lovero, D., D'Addabbo, P., Caizzi, R. & Marsano, R. M. Identification of Bari Transposons in 23 Sequenced Drosophila Genomes Reveals Novel Structural Variants, MITEs and Horizontal Transfer. *Plos One* **11**, doi:ARTN e0156014
10.1371/journal.pone.0156014 (2016).
95 Yuan, Y. W. & Wessler, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc Natl Acad Sci U S A* **108**, 7884-7889, doi:10.1073/pnas.1104208108
1104208108 [pii] (2011).
96 Tellier, M., Bouuaert, C. C. & Chalmers, R. Mariner and the ITm Superfamily of Transposons. *Microbiol Spectr* **3**, MDNA3-0033-2014, doi:10.1128/microbiolspec.MDNA3-0033-2014 (2015).
97 Palazzo, A., Caizzi, R., Viggiano, L. & Marsano, R. M. Does the Promoter Constitute a Barrier in the Horizontal Transposon Transfer Process? Insight from Bari Transposons. *Genome Biol Evol* **9**, 1637-1645, doi:10.1093/gbe/evx122 (2017).

98 Palazzo, A. *et al.* Transcriptionally promiscuous "blurry" promoters in Tc1/mariner transposons allow transcription in distantly related genomes. *Mob DNA* **10**, 13, doi:10.1186/s13100-019-0155-6 (2019).
99 Minervini, C. F. *et al.* Evidences for insulator activity of the 5'UTR of the Drosophila melanogaster LTR-retrotransposon ZAM. *Mol Genet Genomics* **283**, 503-509, doi:10.1007/s00438-010-0529-4 (2010).
100 Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G. & Worm, B. How many species are there on Earth and in the ocean? *PLoS Biol* **9**, e1001127, doi:10.1371/journal.pbio.1001127 (2011).
101 Hartl, D. L., Lozovskaya, E. R., Nurminsky, D. I. & Lohe, A. R. What restricts the activity of mariner-like transposable elements. *Trends Genet* **13**, 197-201 (1997).
102 Gilbert, C. & Feschotte, C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev* **49**, 15-24, doi:10.1016/j.gde.2018.02.007 (2018).
103 Schön, U. *et al.* Human endogenous retroviral long terminal repeat sequences as cell type-specific promoters in retroviral vectors. *J Virol* **83**, 12643-12650, doi:10.1128/JVI.00858-09 (2009).
104 Fuchs, N. V. *et al.* Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3. *J Virol* **85**, 3436-3448, doi:10.1128/JVI.02539-10 (2011).
105 Montesion, M., Williams, Z. H., Subramanian, R. P., Kuperwasser, C. & Coffin, J. M. Promoter expression of HERV-K (HML-2) provirus-derived sequences is related to LTR sequence variation and polymorphic transcription factor binding sites. *Retrovirology* **15**, 57, doi:10.1186/s12977-018-0441-2 (2018).
106 Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782, doi:10.1534/genetics.166.4.1775 (2004).
107 Iwaki, T., Figuera, M., Ploplis, V. A. & Castellino, F. J. Rapid selection of Drosophila S2 cells with the puromycin resistance gene. *Biotechniques* **35**, 482-484, 486, doi:10.2144/03353bm08 (2003).
108 Lum, L. *et al.* Identification of Hedgehog pathway components by RNAi in Drosophila cultured cells. *Science* **299**, 2039-2045, doi:10.1126/science.1081403 (2003).
109 Kunzelmann, S., Bottcher, R., Schmidts, I. & Forstemann, K. A Comprehensive Toolbox for Genome Editing in Cultured Drosophila melanogaster Cells. *G3 (Bethesda)* **6**, 1777-1785, doi:10.1534/g3.116.028241 (2016).
110 Wang, J. W., Beck, E. S. & McCabe, B. D. A modular toolset for recombination transgenesis and neurogenetic analysis of Drosophila. *PLoS One* **7**, e42102, doi:10.1371/journal.pone.0042102 (2012).
111 Buchman, A. & Akbari, O. S. Site-specific transgenesis of the Drosophila melanogaster Y-chromosome using CRISPR/Cas9. *Insect Mol Biol* **28**, 65-73, doi:10.1111/imb.12528 (2019).
112 Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186**, 735-755, doi:10.1534/genetics.110.119917 (2010).

113 Lin, H., Wang, Q., Shen, Q., Zhan, J. & Zhao, Y. Genetic engineering of microorganisms for biodiesel production. *Bioengineered* **4**, 292-304, doi:10.4161/bioe.23114 (2013).
114 Sticklen, M. B. Plant genetic engineering for biofuel production: towards affordable cellulosic ethanol. *Nat Rev Genet* **9**, 433-443, doi:10.1038/nrg2336 (2008).
115 Tyagi, S., Choudhary, R., Das, A., Won, S. Y. & Shukla, P. CRISPR-Cas9 system: a genome-editing tool with endless possibilities. *J Biotechnol*, doi:https://doi.org/10.1016/j.jbiotec.2020.05.008 (2020).
116 Marsano, R. M. *et al.* Mosquitoes LTR retrotransposons: a deeper view into the genomic sequence of Culex quinquefasciatus. *PLoS One* **7**, e30770, doi:10.1371/journal.pone.0030770 (2012).
117 Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338-345, doi:10.1038/nbt.4060 (2018).
118 Marsano, R. M., Giordano, E., Messina, G. & Dimitri, P. A New Portrait of Constitutive Heterochromatin: Lessons from Drosophila melanogaster. *Trends Genet* **35**, 615-631, doi:10.1016/j.tig.2019.06.002 (2019).
119 Yasuhara, J. C., DeCrease, C. H. & Wakimoto, B. T. Evolution of heterochromatic genes of Drosophila. *Proc Natl Acad Sci U S A* **102**, 10958-10963, doi:10.1073/pnas.0503424102 (2005).
120 Caizzi, R. *et al.* Comparative Genomic Analyses Provide New Insights into the Evolutionary Dynamics of Heterochromatin in Drosophila. *PLoS Genet* **12**, e1006212, doi:10.1371/journal.pgen.1006212 (2016).