TRANCERs: Engineering enhancers into autonomous tissue-specific expression cassettes

Elisa K. Barrow Molina[1], Alastair L. Smith[1], Paco López-Cuevas[2], Matthew Nicholls[1], Richard M. White[2], Thomas A. Milne[1*] & Jim R. Hughes[1*]

[1] Weatherall Institute of Molecular Medicine, University of Oxford, The John Radcliffe Hospital, Headington, Oxford, OX3 9DS, UK
[2] Nuffield Department of Medicine, Ludwig Institute for Cancer Research, OX3 7DQ, University of Oxford, Oxford, UK

Abstract

The process by which a foreign gene is introduced into a cell and translated into a functional protein is referred to as transgene expression. Understanding and optimising transgene expression is crucial for a wide variety of applications which range from investigating basic biology, synthetic biology, recombinant protein production to gene therapy. However, precise control over the timing and location of transgene expression remains a key limitation, particularly in therapeutic contexts where specificity and compactness are critical.

To address limitations of traditional promoter-based transgene construct approaches, we developed TRANCERs (TRanscriptionally Autonomous enhaNCERs), a modular and compact system that transforms cell type-specific enhancers into standalone sources of productive transcription. To accomplish this, we leveraged basic facets of enhancer function as well as known and novel principles of transcript stability and nuclear export.

We demonstrate that TRANCERs can drive lineage-restricted expression of both reporter and therapeutic genes across multiple mammalian cell types both *in vitro* and *in vivo*, and that this expression pattern faithfully mirrors the activity of the enhancer element used. Enhancer elements are the most cell type specific element in the mammalian genome and are used to drive development, differentiation and cellular responses to extrinsic signals. Therefore, the TRANCER system allows for the generation of specific stable expression constructs for all conceivable situations for which a native enhancer exists with the required activity pattern. Furthermore, we show that the TRANCER system is compatible with previously published machine learning derived artificial enhancer sequences suggesting the possibility even for specific bespoke expression patterns that do not exist in nature. This, combined with a significantly reduced vector footprint ideal for size-constrained delivery systems such as lentiviral vectors, means that the TRANCER system represent a significant step forward in our ability to control gene expression for both basic biology and therapeutic applications.

Introduction

While promoters are known to be crucial for transcriptional initiation, they are generally insufficient for establishing the complex tissue and condition-specific expression patterns required in somatic cells of a multicellular organism. Additional control elements termed enhancers are required to provide such control most genes [1]. Enhancers are non-coding cis-regulatory sequences that direct the pattern, level and timing of gene expression by promoting or enhancing the expression of their target promoter [2].

Promoters have been shown to drive mainly unidirectional transcription and are found at the start of genes to establish the transcriptional start site (TSS). Unlike promoters, enhancers are

distributed unpredictably around the genes they control; they can be located both upstream or downstream of the promoter they interact with and at variable distances[3]. Enhancers are typically located within 250kb of their target genes, but they can be located up to 1.5 Mb away of the target interacting promoter. Additionally, enhancers can be found in the regions of the genome between genes (intergenic) as well as within the noncoding intronic regions of genes (intragenic) [2,4]. While most mammalian enhancers are also transcriptionally active, their product, known as enhancer RNA (eRNA), are considered weak ephemeral transcripts that generally only span around 300 bps. eRNAs are generally bidirectional, short-lived, non-polyadenylated, low in abundance, 5' capped, unspliced, non-coding and nuclear retained [5–8]. The function of eRNAs remains unclear. It is unknown whether the process of transcription and eRNA production itself is important or just a byproduct of enhancer activity [9–12]. Nevertheless, there is a clear link between an enhancer's transcription and its activity which means that eRNA production has the same precise tissue and context specificity of the source enhancer [13,14].

Beyond both being transcriptional sources enhancers and promoters share many features, such as association with open chromatin, high levels of chromatin modification and RNA polymerase loading. However, differences between these elements exist and, of central importance to our technology, they differ in their level of functional autonomy. The loading of transcription factors (TFs) on promoters is highly dependent on their interactions with enhancers, which is critical for regulating the level of transcription, whereas the loading TFs on enhancers is inherent in its sequence and is not dependent on other elements[15–17]. While autonomous cell type-specific promoters have been reported, they are rare. In contrast, there are estimated to be millions of cell type-specific enhancers in the human genome controlling development, differentiation and cellular responses to extrinsic signals[15]. This ubiquitous use of enhancers combined with their intrinsic autonomy therefore suggests that if their transcription could be captured in a productive manner, then compact promoters could be generated for all conceivable biological niches and stimuli which activate specific enhancer elements.

Studies investigating the properties and interactions between enhancers and promoters have generally relied on transient transfection approaches that allow for large-scale analysis [18–22]. However, these approaches ignore the normal chromatin context of the genome which is a known important property affecting transcriptional initiation at promoters as well as enhancer activity [23]. Under a chromatised context, enhancer elements situated in the body of genes (intragenic) can generate highly expressed isoforms of the host gene (termed multi-exonic or meRNAs) even when the promoter of the gene has been deleted or is naturally repressed. Thus, the enhancer elements alone can act as alternative promoters producing long non-coding RNA (lncRNA) that are abundant and mature (polyadenylated and spliced) [9]. Since then, some groups have provided evidence of enhancers that form polyadenylated and stable eRNAs. Currently, there is an active debate regarding the classification of stable eRNAs as some argue they could be classified as lnRNAs [9]. In previous work, we have also shown that when a 325 base pair (bp) sequence containing the major erythroid-specific α-globin R2 enhancer element was inserted into a neutral region in chromosome X (ChrX in male mouse embryonic stem cells (mESC)) with no pre-existing function, the enhancer acted as a potent source of transcription in erythroid cells and generated a nascent transcribed domain of ~70 kb and stable transcripts easily detectable in PolyA selected RNA-seq experiments [24].

The link between these observations is the presence of downstream sequences, including splice and polyadenylation (polyA) sites. In the context of meRNAs and stable eRNAs, these correspond to the downstream splice and polyA sites of the host gene. The novel enhancer transcript in the ChrX locus from the globin enhancer employs downstream cryptic splice and polyA sites to produce stable, spliced isoforms. Additionally, the chromatin signature associated with this well characterised enhancer sequence switched from that of an enhancer in its native context to a

promoter enriched for H3k4me3 when moved to chrX [24]. Such results further blur the traditional distinctions between enhancers and promoters which remain to be completely understood [25]. This lack of absolute differences between enhancers and promoters and the number of shared features between them suggest that their function may depend, at least in part, on genomic context. This is further supported by previous work which showed that ~50% of erythroid long non-coding RNAs (lncRNAs) come from enhancer elements and use poorly conserved splice and polyadenylation sites [9]. Taking these observations together, we here hypothesised that eRNAs are not weak but rather get rapidly degraded and that through the addition of downstream sequences we might be able to convert eRNAs into potent sources of productive transcription with the same specificity of the enhancer element.

To produce a desired RNA or protein in a specific cell type or situation, the current approach is to combine well-characterised enhancer sequences with minimal promoters. These elements are integrated into an expression vector, which serves as the vehicle for driving the expression of the target gene. This approach can often result in what is called "leaky" expression, where the RNA or protein is expressed when not desired [26–28] . An alternative approach to tissue or cell type-specific expression is to try and specifically deliver the expression system to a particular cell type, which is extremely challenging and often prone to off-target problems [29,30].

Another technical problem is that such tissue-specific expression vectors are large and difficult to package into viral delivery vectors. This severely limits the size of the payload that can be expressed and thus, affects their applicability [31]. Additionally, the predictability of the effectiveness of any enhancer and promoter combination is poor, and this practically limits the available repertoire of enhancers and promoters [28,32] . Similarly, promoters can also interact with other enhancers in the genome and cause off target effects.

To try and address all of these limitations, we have generated a new platform technology which we call "Transcriptionally Autonomous Enhancers" (or TRANCERs for short), where we can utilise the inherent transcription associated with enhancer elements and show that they can be converted to small (~250-300 base pairs), highly expressed, autonomous, highly cell type/state specific autonomous transcriptional promoters. We have demonstrated the robustness of TRANCERs to act as tissue and cell-type specific genes in different mammalian cell lines, organoid systems and *in vivo* in zebrafish.

Results

Enhancer transcription can be stabilised and made to produce protein coding transcripts via the addition of a range of cis acting sequences.

For the creation of TRANCERs we hypothesised that enhancer transcription is not ephemeral or weak, but instead it is highly unstable. We further hypothesised that this transcription could be stabilised to achieve functionally useful levels of RNA expression and made to engage with the protein translation machinery by the addition of naturally occurring cis acting sequences. We chose mouse embryonic stem cells (mESC) as our initial model system due to their high levels of characterisation, genetic tractability and their inherent ability to be differentiated into highly complex multicellular organoids.

To identify enhancer sequences active in mES cells, we used eRNA signals as a proxy for enhancer activity, a well-established approach in the field [33–35]. For this we used nascent RNA-seq analysis and generated TT-seq in mES to allow for the efficient detection of short-lived RNA species to obtain a more accurate representation of enhancer activity and active transcription in

the cells[36]. We also generated a map of the positions of enhancers, promoters and boundary elements using open chromatin data (ATAC-seq) and ChIP-seq data for key chromatin modifications and bound proteins (H3k4me3, H3k4me1, K3k27ac and CTCF). Integrating these data gave us a comprehensive map on highly active enhancers in mES cells. We further refined our list of enhancers by focusing on intergenic enhancers to reduce noise and confounding signals from promoters and gene bodies. The top TT-seq eRNA hits with a classical enhancer signature were then compared with the literature to select previously validated elements that meet the classical definition of an enhancer either via genetic or chromatin profiling. The list of the enhancer elements used is shown in Supplementary Table 1. Note that, enhancer names were adopted directly from existing literature, where each enhancer region is assigned to the nearest annotated gene. However, linking enhancers to their target genes is non-trivial. Enhancers are known to form long-range interactions with their target promoters often spanning several hundred of kilobases [3]. As a result, some of the enhancer–gene assignments in our initial compilation were likely inaccurate. For example, an enhancer previously linked to *Tmprss13* showed robust activity in TT-seq data despite *Tmprss13* being silent in mESCs, suggesting a misattribution. To resolve such discrepancies, we performed Micro-Capture-C (MCC) (Supplementary Figure 1), which conclusively demonstrated that this enhancer (chr9:45260481–45261718) interacts with *Fxyd6*. Accordingly, we will hereafter refer to this element as the *Fxyd6* enhancer, reflecting its validated target gene.

Having identified suitable enhancer sequences, a base TRANCER construct was designed containing: an enhancer (varying sizes, Supplementary Table 1), an artificial splice site that we will refer to as 'intron' (294 bp), a mini Neon Green (mNG) as a reporter (711 bp) with a 5' Kozak sequence, polyadenylation signal from the bovine growth hormone gene (bGH) (225 bp) and a LoxP floxed puromycin resistance cassette under an EF1-α core promoter (857 bp) used for antibiotic selection. These sequences were flanked with homology arms targeting the neutral region on chrX used in previously published experiments [24]. The scrambled negative control used is a reversed non-complementary version of the Nanog enhancer used to preserve general nucleotide content characteristics, while destroying the TF binding motifs present in the native sequence.

All assembled enhancer constructs drove readily detectable mNG expression, albeit at varying levels. Notably, the inclusion of canonical stabilisation sequences (splice and polyadenylation sequences) was sufficient to confer promoter-like activity to enhancer elements, eliminating the need for any other promoter elements (Figure 1.A). To exclude the possibility that the puromycin resistance cassette contributed to mNG transcription, we removed this region using the Cre–loxP system [37]. As shown in Figure 1B, mNG expression remained unchanged, indicating that the additional promoter driving puromycin expression did not influence reporter activity and the enhancer sequence was the only potential surce of transcription in the construct.

Downstream of the mNG reporter, the construct carried a 3' untranslated region (UTR) with a canonical bGH polyadenylation signal (Figure 1.A and B). UTRs are increasingly recognised as key regulators of RNA maturation, stability, localisation, and translation [38]. To probe the contribution of 3' end stabilisation to reporter expression and to explore alternative elements which may also act as stabilisation sequences, we replaced this poly(A) based UTR with a hairpin-forming sequence from the long non-coding RNA MALAT1 [39]. MALAT1 terminates in a highly conserved AU-rich triple-helical structure that protects the transcript from degradation. Although noncoding, MALAT1 accumulates to levels comparable to housekeeping genes [40–43]. Its transcriptional maturation is mediated by RNase P cleavage and similar looped structures are found in other lncRNAs [39,41,44]. In these constructs we used the minimal conserved MALAT1 loop (78bp), hereafter termed the Hairpin Stabilisation Sequence (HSS). In addition to protecting against degradation, the HSS has previously been reported to enhance translation in the

cytoplasm, suggesting potential advantages over canonical poly(A) for transgene expression [42]. Figure 1.C shows that the HSS can indeed function as a 3' stabilisation domain replacing the need for a polyA signal, however, there was no evidence of any notable increase in protein expression suggesting enhanced translation. As a control we removed of the 3' stabilisation sequence (either HSS or polyA, Figure 1.C.IV) and showed that this completely abolished mNG expression, confirming the requirement of at least one of these classes of elements for reporter activity.

Canonical RNA splicing is the eukaryotic process that converts newly transcribed precursor messenger RNA (pre-mRNA) into mature mRNA. This occurs in the nucleus, where introns are removed and exons joined to form a continuous coding sequence [45]. While splicing is thought to have evolved to expand proteomic diversity by generating alternative isoforms from a single gene, here we focus on its further role in regulating protein production [46]. Intronic sequences also link transcription to RNA export. The transport/export (TREX) complex couples splicing and transcription to mRNA export [47,48]. Thus, splicing is directly tied to mRNA export and its cytoplasmic translation into protein. To test the contribution of the synthetic splice site (labelled as "intron" in Figure 1) to TRANCER expression, we compared constructs with and without the intron. Removing the intron abolished mNG protein production (data not shown). However, qPCR analysis showed robust transcript accumulation despite the absence of detectable protein (Figure 1D). This result suggests that unspliced transcripts are retained in the nucleus, similar to lncRNAs. The lack of functional splice sites likely prevents access to the splicing-dependent export pathway, triggering nuclear retention and stabilization rather than translation. This aligns with the evolutionary bias in eukaryotes toward exporting and translating only properly spliced mRNAs, maintaining expression fidelity and suppressing aberrant gene activity [47,48].

[47,48]While the link between splicing and export is well supported by research, many intronless genes bypass the need for splicing through specialised export elements. Among the best characterised are cytoplasmic accumulation regions (CARs). In 2013, a minimal 10 bp motif, CAR-E (CCAGTTCCTG; Figure 1.E), was identified within CARs of naturally intronless genes such as *HSPB3*, *c-Jun*, *IFNz1*, and *IFNβ1*. CAR-E recruits the TREX export machinery, and tandem repeats of the motif (1, 10, or 16 copies) were shown to enhance trafficking efficiency in a copy-number-dependent manner [49]. To restore mNG expression in the intronless TRANCER construct, 3 to 5 concatenated copies of the CAR-E sequence (30-50bp) were inserted immediately downstream of the mNG stop codon at the 3' end adjacent to the PolyA signal in the intronless construct (Figure 1.E). Flow cytometry analysis of monoclonal cell populations clearly demonstrated restoration of mNG expression following the addition of the CAR-E. Notably, the construct containing 5 tandem CAR-E copies resulted in an almost complete re-establishment of expression (Figure 1.E). As a control, a critical guanine residue within the CAR-E consensus motif was mutated to a thymine (G to T). Contrary to our expectation, we observed that expression remained high. This surprising result suggests that the precise CAR-E sequence motif itself, which is derived from only a handful of genes, may not be the complete determinant for TREX complex interaction and efficient export. Instead, the successful export and subsequent translation, could potentially be driven by the higher-order folding or overall structural presentation arising from the multiple copies of this motif.

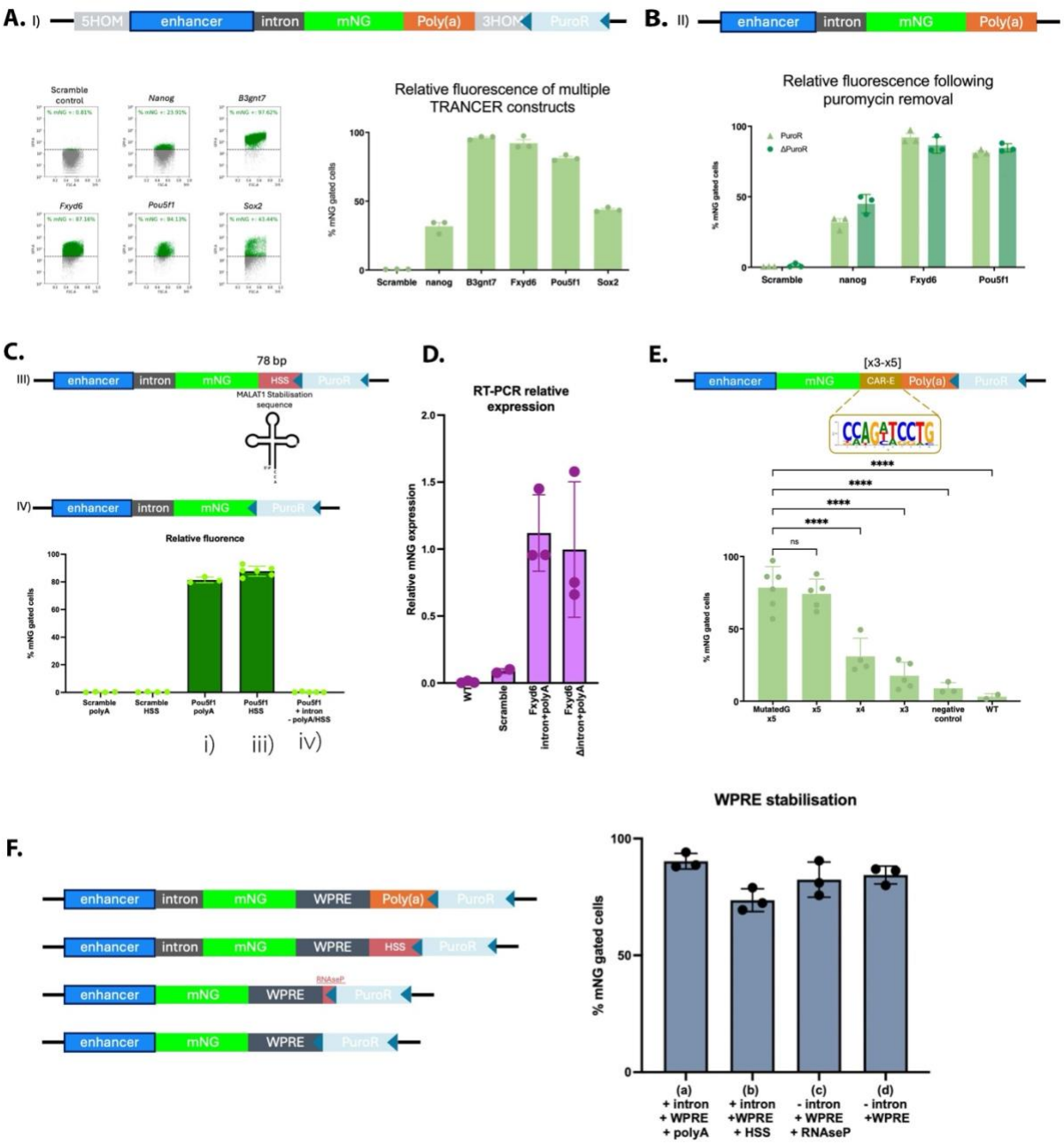

Figure 1. Generation and characterization of TRANCER constructs using multiple approaches.
(A) *base TRANCER constructs (intron + polyA).* (top) Schematic of the TRANCER construct inserted as a single copy into a neutral locus on ChrX. The construct includes a synthetic intron (dark grey box), polyadenylation site (orange box), and an interchangeable enhancer element (blue box) that drives specific expression patterns. A floxed puromycin resistance cassette enables selection. Optional 5′ and 3′ homology arms (grey boxes) allow for CRISPR-mediated targeted integration if required. (bottom) Flow cytometry (Attune NxT) analysis of Neon Green (mNG) reporter expression, reflecting TRANCER-driven protein production. Green and grey populations denote mNG-positive and -negative cells, respectively. Enhancer identity is indicated above each plot. The "Scramble control" corresponds to a Nanog enhancer sequence with disrupted transcription factor binding sites but preserved base composition. Left: bar graph quantifying mNG-positive cells (%); error bars represent SEM.
(B) *Effect of puromycin cassette excision on mNG expression.* (top) Schematic illustrating removal of the floxed puromycin resistance cassette by Cre recombinase. (bottom) Bar plot summarizing flow cytometry data (three biological replicates) comparing TRANCER constructs with (light green triangles) or without (dark green circles) the puromycin cassette. Expression levels remain unchanged, indicating that cassette removal does not affect mNG expression.
(C) *Replacement of the polyA site with a Hairpin Stabilization Sequence (HSS).* (top) Schematic showing substitution of the 225 bp polyA signal with a 78 bp HSS derived from the *MALAT1* non-coding RNA. A control construct lacking both polyA and HSS sequences was also generated. (bottom) Flow cytometry quantification of mNG fluorescence for each configuration, assessing RNA stabilization efficiency.

(D) *TRANCER transcript localization and nuclear retention.* qPCR analysis confirming transcript expression and nuclear accumulation without cytoplasmic export, consistent with absent protein production (as in panel C). These data suggest loss of TREX pathway engagement and nuclear retention of the transcript.
(E) *Rescue of intronless TRANCER expression by CAR-E motifs.* Three to five tandem copies of the 10 bp CAR-E consensus motif were inserted into an intronless TRANCER containing a 3′ polyA site. mNG expression was restored in a copy number–dependent manner, indicating that CAR-E elements can substitute for the intron in promoting expression.
(F) *Inclusion of* Woodchuck Posttranscriptional Regulatory Element (WPRE) *restores expression in intron- and polyA-deficient TRANCERs.* (left) Schematic overview of the constructs tested, including: (1) TRANCER with intron and polyA plus WPRE; (2) HSS-containing construct with WPRE; (3) construct lacking the intron plus WPRE; (4) construct lacking polyA but containing an RNAse P cleavage site plus WPRE; and (5) construct lacking both TREX engagement and 3′ stabilization elements in the presence of the WPRE. WPRE inclusion preserves mNG expression even in the absence of intron and polyA sequences. (right) Bar plot quantifying the mNG-positive cells for all constructs.

Many viruses have evolved sequences that bypass canonical splicing or polyadenylation to achieve heterologous gene expression [50–54]. An example of such an element is the Woodchuck Hepatitis Virus Post-transcriptional Regulatory Element (WPRE), a 591 bp cis-acting RNA sequence (nt 1093–1684 of the WHV genome)[55]. The WPRE is thought to enhance nuclear export and cytoplasmic accumulation of mRNA, leading to markedly higher expression of encoded proteins [56,57]. Its incorporation into therapeutic vectors is now common for this reason, with clinical applications across multiple indications [58–60]. Although the mechanism remains incompletely defined, WPRE is thought to synergise with polyA signals to maximise transcript stability and export. To test whether it was also able to interact synergistically with the PolyA signal within a TRANCER construct, we introduced the clinically accepted WPREmut6 sequence, hereafter referred to as WPRE, into the base TRANCER construct, which also contains an intron and a poly(A) tail. Surprisingly, negative control constructs in this experiment which lacked a splice site but contained a WPRE showed the same amount of mNG expression as those which contained both a splice site and a WPRE. The precise biology of such viral elements is poorly understood and suggested that they may have evolved alternative sequence for promoting RNA stability and export prompting a closer examination of WPRE function in the context of TRANCERs. Strikingly, we found that the WPRE could substitute for all canonical elements. When present, the WPRE rendered both the splice site and the polyA tail dispensable (Figure 1.F), suggesting that the WPRE alone is sufficient to stabilise transcripts, promote nuclear export, and enable translation through a yet undefined mechanism.

Here, we have demonstrated that we can stabilise eRNAs creating productive transcription in a new technology we have called TRANCERs. Overall, the requirements for expression can be split into two main functional elements: (1) interaction with the nuclear-to-cytoplasmic export machinery and (2) 3’ end stabilisation of the transcript.

Investigating the cell and tissue specific expression of TRANCERs in organoids.

As stated, we chose mESCs in part for capacity to differentiate into diverse cell types and tissues, which would provide a convenient multi-cellular model to study TRANCER specificity [61]. For this study, we used *in vitro* differentiation into embryoid bodies (EBs), a previously established organoid approach. EBs are 3D aggregates of pluripotent stem cells that differentiate to give rise to derivatives of all three germ layers: ectoderm, mesoderm, and endoderm [62].

During EB differentiation, primitive embryonic erythroid cells emerge and provide a tractable system to study enhancer activation dynamics of the most well characterised mammalian enhancer sequences without the need for *in vivo* models [62,63]. For this study, we used the well-characterised R2 enhancer of the α-globin locus, which accounts for ~50% of total enhancer activity of the α-globin genes in their native context [64]. A base TRANCER construct (intron + polyA

signal) driven by R2 was introduced into the neutral region at chrX (Figure 2.A). Fluorescent mNG reporters were unsuitable due to high haemoglobin autofluorescence, so we substituted this with a truncated human CD19 receptor (Addgene #123661) which can be detected with APC-conjugated antibodies using flow cytometry (Figure 2.B).

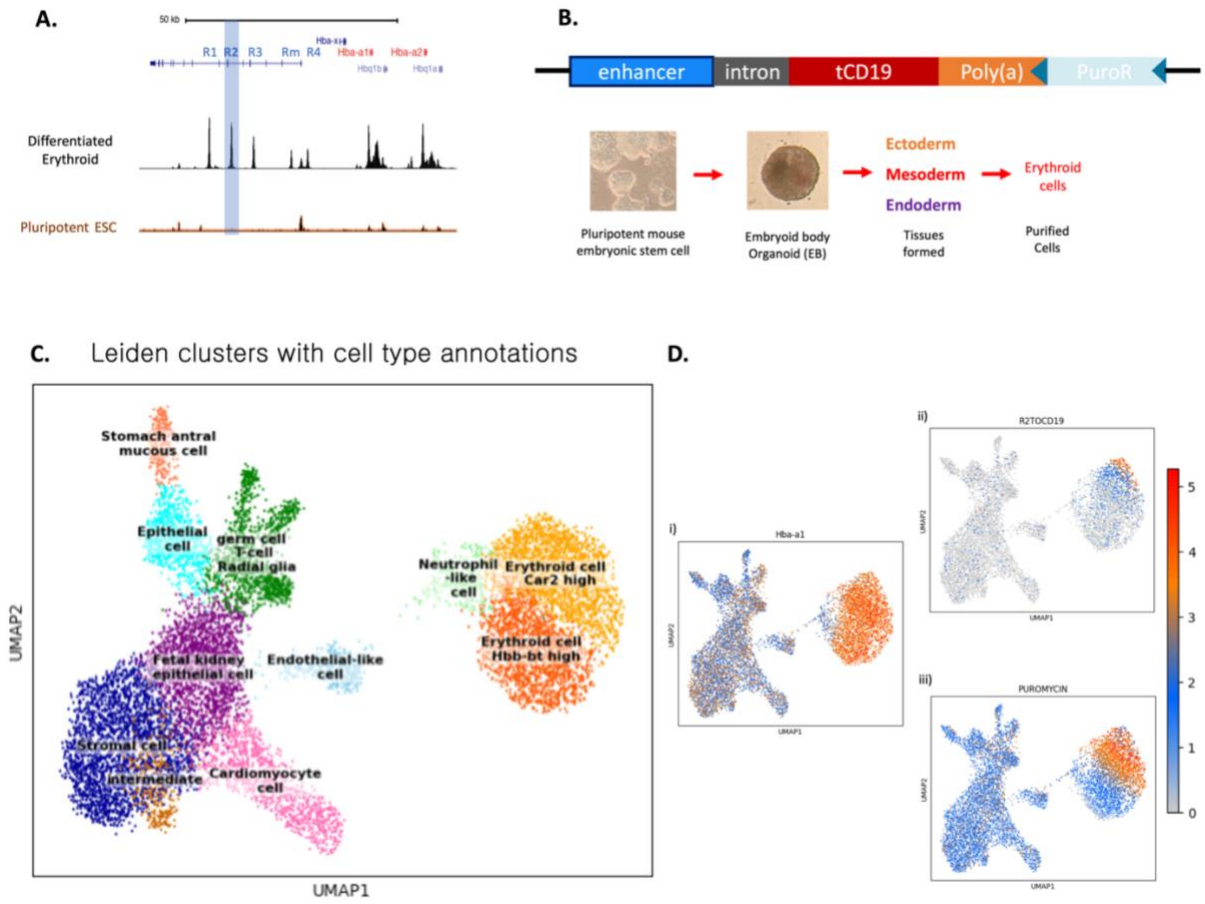

Figure 2. CD19 expression is restricted to erythroid cells in an EB model.
(A) ATAC-seq data from differentiated mESC showing that open chromatin accessibility at the R2 enhancer (highlighted in bold and indicated by the blue bar) is restricted to erythroid cells.
(B) (top) Schematic of a TRANCER construct containing the erythroid-specific R2 enhancer element as the transcriptional driver, with the mNG reporter replaced by the coding sequence of human *CD19*. (bottom) Stages of erythroid cell generation from mESC in the formation of erythroid cells. Experimental workflow proceeds left to right: formation of embryoid body (EB) organoids, can be followed by immunopurification of red blood cell precursors from the heterogeneous population of ectoderm, mesoderm, and endoderm lineages produced in this system, enabling erythroid enrichment.
(C) UMAP projection of full EBs containing the R2–CD19 TRANCER construct, with annotated cell types corresponding to those identified in wild-type (WT) EBs.
(D) UMAP visualizations derived from multiome data demonstrating that the R2–tCD19 TRANCER construct is specifically expressed in erythroid cells, as evidenced by its colocalization with *Hba-a1* expression. (i) Expression of the erythroid marker gene *Hba-a1*, showing spatial overlap with *Cd19*-expressing cells. (ii) UMAP of the same dataset, where darker colors represent higher read counts of the inserted R2–tCD19 TRANCER sequence (from the *CD19* gene to the puroR cassette). (iii) UMAP representing read counts from *CD19* to the polyA tail.
(E) Pseudobulk tracks derived from WT germ cell and erythroid clusters compared with the *CD19* (panel D.iii) data from TRANCER-containing EBs. The ATAC-seq accessibility signature is identical between the erythroid cluster and *CD19*-expressing cells, confirming erythroid-specific activation of the R2 enhancer. Note that multimapping issues resulted in R2 reads being excluded from the track.

Due to the high levels of cellular complexity of EB organoids, we used single-cell 10X Multiome to collect chromatin accessibility and gene expression to demonstrate specific CD19 expression

in the EB derived erythroid cells. Since erythroid cells constitute only a minor fraction of embryoid bodies (EBs), CD19+ cells where enriched using FACS to make up around 20% of stained cells. This allowed for robust analysis of both the open chromatin landscape and transcriptome of the CD19+ cells in the context of all the other major constituent cell types within the EBs. Cell type annotations were done using cellxgene (Figure 2.C). The UMAP in Figure 2.D.i, coloured by *Hba-α1* expression, marks the erythroid cluster in agreement with standard cell-type annotation. UMAPs in Figure 2.D.ii-iii are coloured according to the expression of the CD19 alone (ii) or the full cassette including the puromycin resistance gene (iii). The broader expression in iii shows constitutive expression in all cells driven by the EF1α core promoter, however, increased expression is seen in the erythroid-labelled cluster likely boosted by enhancer-promoter crosstalk or read-through transcription from the R2-containing TRANCER. As expected, only a small fraction of EB-derived cells was CD19 positive, reflecting the limited representation of terminally differentiated RBC within the mixed germ-layer composition of EBs.

Within the erythroid cluster, computational annotation distinguishes "Car2 high" (carbonic anhydrase 2) and "Hbb-bt high" (Beta globin high) subclusters (Figure 2.B), representing late- versus earlier-stage erythropoiesis, respectively. Strikingly, R2-driven CD19 expression was restricted to the more differentiated "Car2 high" cells, despite the canonical role of R2 in early α-globin activation suggests even greater cellular specificity than previously suspected. To confirm lineage specificity, we pseudobulked the ATAC signal from CD19+ cells within the multiome data and compared with to bulk erythroid and germ layer clusters (Figure 2.E). The chromatin profile of the CD19+ TRANCER active cells matched the canonical erythroid open chromatin profile, confirming the cell type in which the R2 TRANCER was active was indeed erythroid and so matches the R2 enhancers native activity. However, while the R2 peak itself had to be excluded from this analysis due to multi-mapping of endogenous and integrated copies, the other erythroid specific enhancers, that with R2 make up the α-globin super-enhancer can be seen to be clearly active (Figure 2.E).

Overall, this experiment confirms cell-type specific expression of CD19 in a TRANCER construct driven by the well-characterised R2 enhancer. Additionally, we also selected other enhancers specific to different distinct clusters annotated by our EB multiome analysis, including those annotated as stomach Mucosal, Endothelial and Cardiomyocyte cells. In each case each show a different expression pattern using confocal imaging (Figure 3). Most striking of these was the TRANCER construct containing a cardiomyocyte enhancer. EB organoids at late stages of development become contractile and begin to beat rhythmically due to the development of functional cardiac tissues. This cardiac structure can be clearly seen as rings within the embryoid bodies (Figure 3.C). Together, these data show that TRANCERs can drive highly selective expression within narrow cellular subpopulations, achieving sharper resolution than native markers such as *Hba-α1* in complex multicellular contexts.

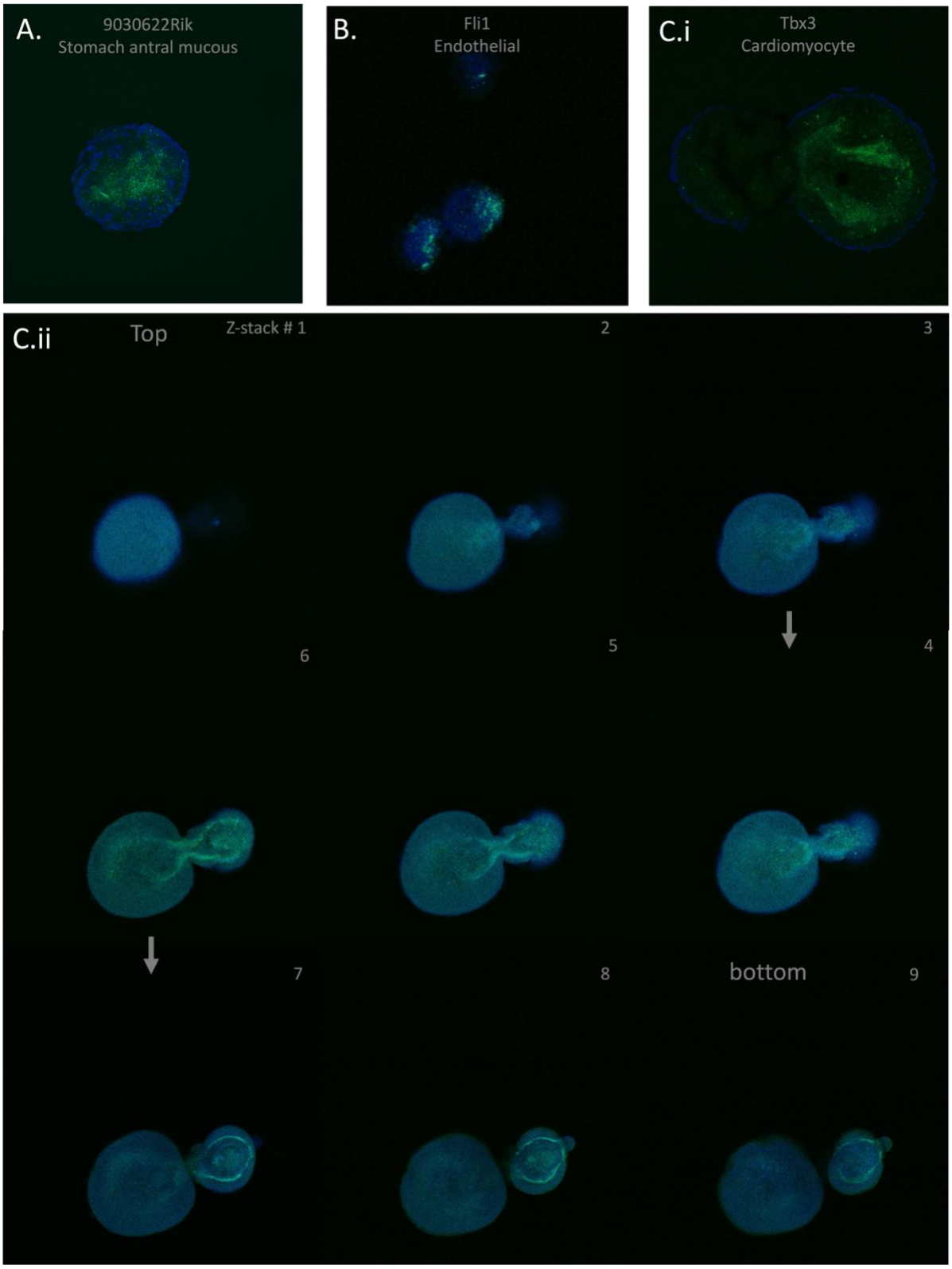

Figure 3. **Confocal images of fully formed EBs where different TRANCER constructs have been stably integrated into chrX LP.**
(A) Orthogonal projection image of a representative clone expressing a stomach enhancer. The image shows a diffused yet specific mNG signal in the EB.
(B) Endothelial enhancer showing expression in three different EBs, localized to one pole of the EB. **(C)** Cardiomyocyte enhancer. In both examples, two EBs are found together, with the outline stained using DAPI.(i) Single slice from a z-stack displaying mNG in a ring-like formation corresponding to the cardiac tissue region within the EB. (ii) Sequential images from a z-stack (read left to right following the arrows; numbers indicate slice positions; a video is available

via the QR code). A progressive increase in mNG signal is observed across slices, consistent with the ring-like structure seen previously. All selected enhancers showed distinct expression patterns.

In vivo TRANCER activity in Zebrafish

To demonstrate that TRANCERs work in a live animal model we chose zebrafish due to their unique suitability for whole animal visualisation of marker gene expression. Due to the highly dynamic structure of the heart we again chose cardiomyocytes as our target cell type. To select enhancers uniquely active in the zebrafish heart myocardium, we used publicly available scATAC-seq data (10x Multiome zebrafish resource) [65]. Our top candidate (chr20:18114062–18114563; 555 bp), was chosen due to it high levels of open chromatin and highly specific activity in cardiomyocytes. This sequence was synthesised and cloned into a base TRANCER construct (intron + polyA) using pDestTol2 as the entry vector.

For *in vivo* delivery, constructs were introduced into zebrafish embryos via yolk sac microinjection using the Tol2 system [66]. This approach resulted in selective TRANCER expression in the myocardium of zebrafish embryos (Figure 4.B), confirming tissue-specific activity *in vivo*. It is important to note some ectopic expression being observed around the eye of the fish likely due to F0 lines being formed.

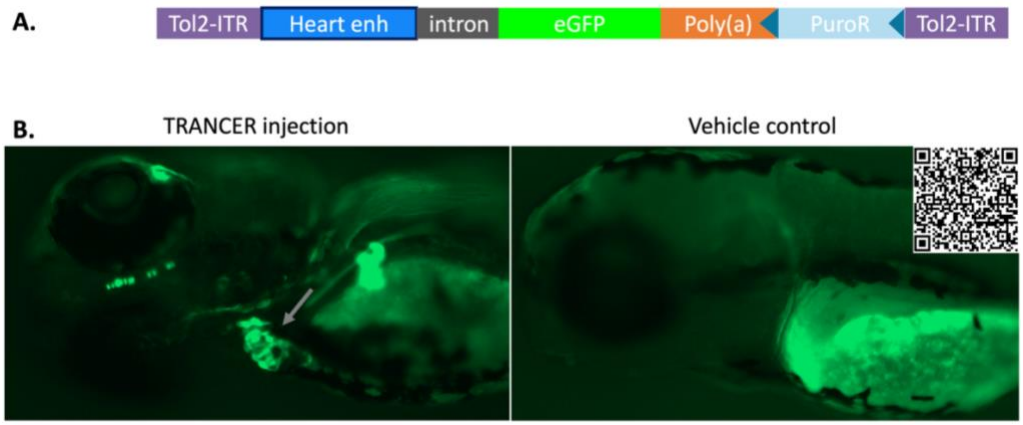

Figure 4. Expression of a heart myocardium TRANCER in zebrafish.
(A) Schematic representation of the construct used in this experiment. The TRANCER donor, driven by a heart-specific enhancer, is flanked by *Tol2* inverted terminal repeats (ITRs). The reporter used in this construct is *eGFP*
(B) Fluorescent microscopy images showing tissue-specific *eGFP* expression localized to the heart myocardium (indicated by grey arrows, left). The right panel shows a negative control embryo that did not receive the injection. A video showing the heartbeat is accessible via the provided QR code. Note: mild background autofluorescence is visible in the yolk sac, a common occurrence attributed to the natural composition of yolk lipids and proteins.

TRANCERs are compatible with synthetic enhancers predicted using AI.

Recent advances in AI and high-throughput reporter assays have accelerated the design of synthetic enhancers to precisely tune gene expression for clinical and biotechnological applications [67]. Gosai *et al.* (2024) recently used deep neural networks and massively parallel reporter assays to generate and test thousands of novel enhancers across three human cell lines: K562 (erythroid precursors), HepG2 (hepatocytes), and SK-N-SH (neuroblasts) [68]. From this dataset, we selected the top two synthetic enhancers specific for K562 or HepG2 cells. These were tested for TRANCER activity in K562 cells with the HepG2-derived enhancers serving as negative controls.

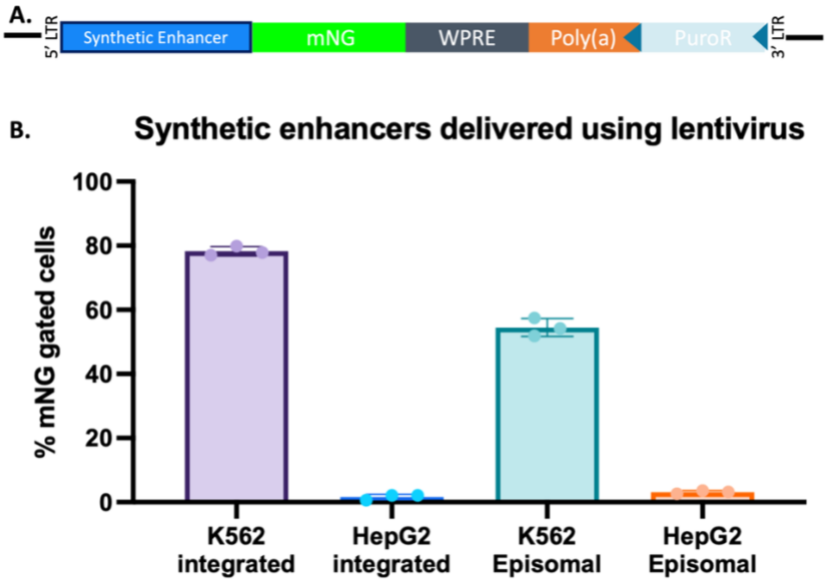

Figure 5. Lentiviral delivery of synthetic TRANCER constructs.
(A) Diagram of the lentivirus TRANCER construct used in this experiment.
(B) Demonstration of TRANCER delivery using top-performing synthetic enhancers [68]. This panel shows that TRANCER constructs can be efficiently delivered and expressed in K562 cells via lentiviral transduction. Both integrated and episomal versions of the lentiviral TRANCER exhibit measurable reporter activity, confirming that lentiviral systems provide a viable platform for TRANCER delivery and functional expression. Data represent $n = 3$ biological replicates; error bars indicate SEM.

TRANCER constructs were built with a WPRE and bGH polyA in a lentivirus vector (Figure 4.A) and delivered by transduction in the presence of polybrene. Both integrating and integrase-deficient (D64V) lentivirus were used. These experiments demonstrate that TRANCERs are also compatible with AI derived synthetic enhancers (Figure 5.B), enabling compact, tissue-specific transgene expression programs that can, in principle, be tailored to any cell type or even combinations of cell types or stimuli.

TRANCERs are intrinsically bidirectional and can be used to express two functional transcripts from the same construct with applications to cell specific genome editing.

The ability to express proteins or RNA species with high cellular specificity is fundamental for both gene therapy/editing applications and basic research. As a platform technology, TRANCERs have the potential to enable a broad range of applications. Here, we showcase how this technology can be used in gene editing applications.

In many contexts, it is desirable to express two transcripts simultaneously and in a coordinated manner within the same cell, particularly where two molecules are needed to form a functional complex. Unlike predominantly unidirectional promoters' enhancers typically produce eRNAs bidirectionally. To determine whether TRANCERs retain their bidirectional transcriptional capability, we developed a TRANCER design that would test if they could express two protein simultaneously. The essential design was to express to different coloured fluorophores (mNG and mCherry) from the same enhancer element, one from the 5' end and one from the 3' end of the enhancer sequence (Figure 6.A). We took advantage of the multiple elements we had characterised as TRANCER components to use two distinct stabilisation strategies to minimise self-homology for ease of construct assembly, utilising a WPRE element alone for mCherry and a PolyA plus intron combination for mNG (Figure 6.A). This construct was used to generate a

stable clonal population following targeted integration of the construct into the neutral locus on chrX in mESC [24]. FACS analysis of these cells shows both fluorophores are highly active, predominantly at high levels in the same cells (Figure 6.A lower panel).

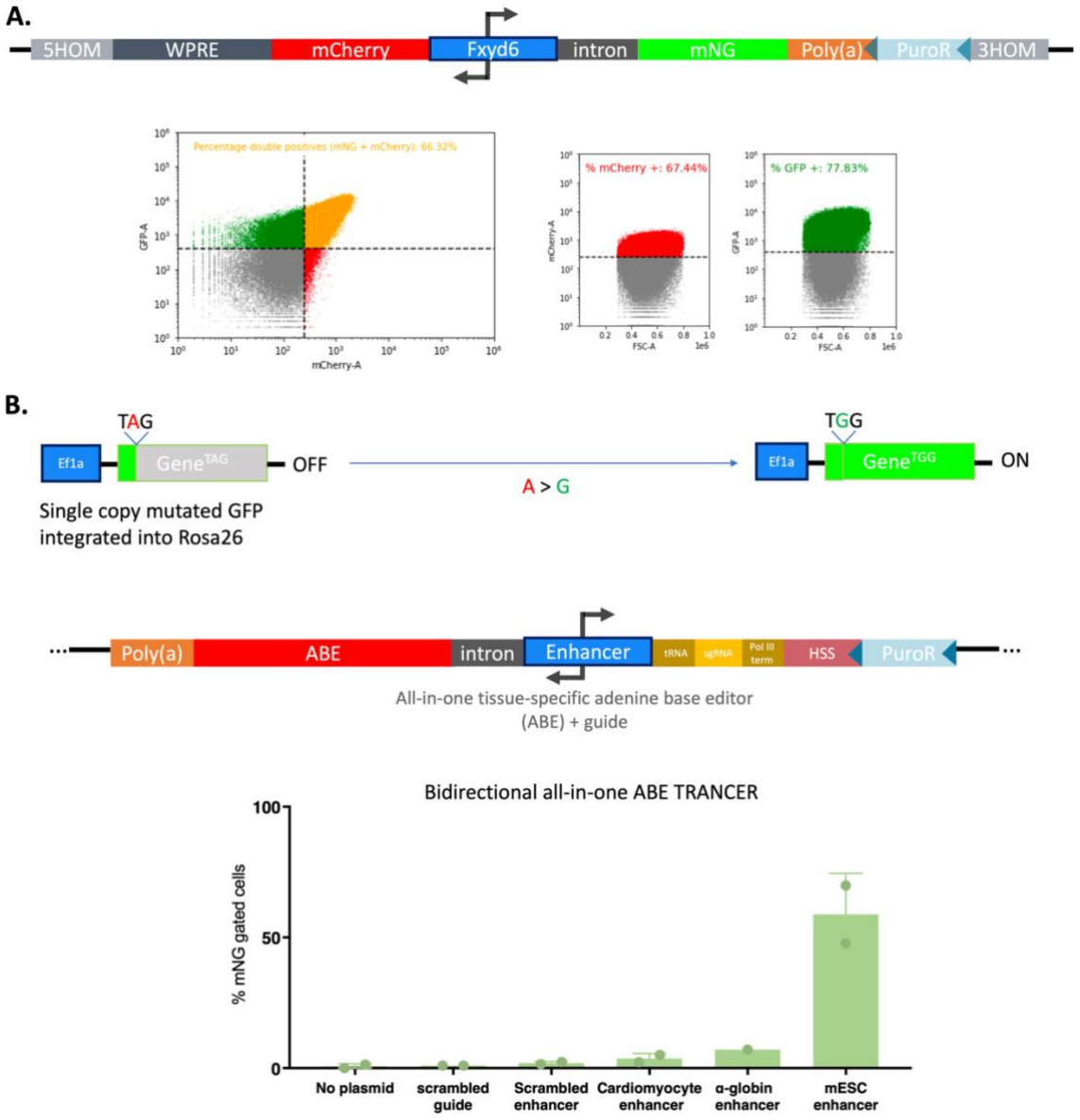

Figure 6. **Bidirectional TRANCER constructs enable simultaneous dual-fluorophore expression and tissue-specific base editing.**
(A) Bidirectional TRANCER expressing mCherry and mNG. (top) Schematic of the TRANCER construct integrated into chrX. From left to right, the construct expresses *mCherry*, stabilized by the inclusion of a WPRE and driven by the mESC-specific *Fxyd6* enhancer. In the opposite direction, *mNeonGreen* (mNG) expression is stabilized and exported via an intron and a polyadenylation (polyA) sequence. (bottom) Flow cytometry analysis of stably integrated clonal mESCs showing simultaneous bidirectional expression of *mCherry* and *mNG*. Double-positive populations correspond to cells co-expressing both fluorophores. Independent single-fluorophore expression profiles are shown to the right.

(B) *All-in-one bidirectional TRANCER design for tissue-specific base editing.* (top) Schematic of the *GFP*(TAG) reporter integrated as a single copy at the *Rosa26* locus. The premature stop codon (TAG) abolishes GFP translation; base editing that converts A→G restores the wild-type codon (*TGG*), enabling GFP expression detectable by flow cytometry. (middle) Diagram of the bidirectional lentiviral TRANCER construct used for targeted base editing. One arm expresses a catalytically inactive Cas9 (dCas9)–based editor stabilized by an intron and polyA sequence. The opposite arm expresses the corresponding sgRNA, whose transcription is primarily enhancer-driven and supported by a Pol III full-length tRNA promoter. Background activity is minimized using a Pol III terminator sequence, and transcript stability is enhanced by a 3′ Pol II Hairpin Stabilization Sequence (HSS). A floxed puromycin resistance cassette enables selection.

(bottom) Bar plot summarizing flow cytometry data from cells transduced with different TRANCER constructs. From left to right: untransduced control, scrambled sgRNA, scrambled enhancer, cardiomyocyte-specific enhancer (*Tbx3*, Figure 3), erythroid enhancer (*R2*, Figure 2), and mESC enhancer (*Fxyd6*). Only the *Fxyd6*-driven construct restores GFP expression, demonstrating tissue-specific base editing activity.

An obviously situation where two transcripts need to be expressed to form a functional complex is CRISPR editing where both the Cas9 nuclease protein and guide RNA are needed to form the active complex. We therefore tested whether the bidirectional capabilities of TRANCERs was capable of expressing both the coding Cas9 transcript and the non-coding guide RNA and form a functional CRISPR editing complex. However, rather than employing first-generation CRISPR-tools, we decided to use base editing (BE) which mitigate the efficiency and toxicity limitations associated with the original CRISPR/Cas9 system. Specifically, we selected adenine base editors (ABEs), as these have been shown to confer improved cell survival and higher editing precision [69].

Following the GFP reporter design described by Katti et al. (2020), we engineered a mouse embryonic stem cell (mESC) line carrying a GFP transgene with a premature stop codon (TAG) integrated into the Rosa26 locus[70]. This mutation prevents GFP translation, rendering the cells non-fluorescent. The GFP(TAG) transgene was placed under the control of a constitutive EF1α promoter. Correction of the TAG codon by an ABE to generate a TGG (Trp) codon restores GFP expression. The TRANCER construct was designed taking advantage of the bidirectional nature of enhancers and expressing both a sgRNA and ABE respectively. Different enhancers were used as negative controls, including no plasmid, a scrambled guide, a previously used scrambled Nanog enhancer (Figure 1), a cardiomyocyte enhancer (Figure 3.C), and the R2 alpha globin enhancer (Figure 2). The Fxyd6 mESC enhancer was the only TRANCER capable of successfully repairing the premature stop codon and restoration of GFP fluorescence (Figure 6B) and did so at similar efficiency as previous published conventional editing experiments (reference).

Discussion

We set out to address long-standing limitations in transgene design relevant to both basic research and translational applications, particularly the need for expression systems that maintain high cell-type or state specificity, minimise cargo size, and remain compatible with viral and non-viral delivery strategies. TRANCERs represent a versatile platform capable of meeting these demands.

Enhancer RNAs (eRNAs) have historically been dismissed as unstable by-products of transcriptional noise. However, accumulating evidence demonstrates a tight association between enhancer transcription, enhancer activity, and gene regulation [13]. Importantly, eRNAs display highly cell-type- and context-specific expression patterns, yet mechanistic insights into their stabilisation remain scarce. Existing evidence for stable eRNAs has largely been observational and has not been translated into transgene applications [9,24,33,34].

Here we show that by incorporating stabiliser motifs (e.g. polyA or HSS) and nuclear export elements (e.g. CARs or introns), enhancer sequences can substitute for promoters and support robust and highly specific productive transcription that mirrors the activity patterns of the enhancer element used. The stabiliser motif protects transcripts from exosome-mediated degradation, while TREX engagement via splicing or cytoplasmic accumulation signals ensures nuclear export. This dual mechanism enables native or synthetic enhancers to generate stable,

properly processed RNAs that can be translated into functional protein products, as demonstrated by expression of reporter cargos such as mNG, tCD19 and Cas9.

We systematically evaluated combinations of stabilisation and export elements. In mESC enhancers, splice sites engaged TREX while polyadenylation stabilised cargo transcripts. We subsequently discovered that a compact cloverleaf motif (HSS) could substitute for polyadenylation without loss of protein output, and that TREX engagement could be achieved without splicing through tandem repeats of a minimal Cytoplasmic Accumulation Region (CAR). This show that the essential TRANCER elements can be reduced to an extremely small vector footprint of ~128 bp, comparable in size to a minimal CMV promoter yet requiring no additional stabilisation sequence. By contrast, conventional CMV-containing constructs remain dependent on enhancer elements to promote specificity and suffer from mis-expression. This size advantage could be transformative for delivery of large cargos, particularly where viral packaging capacity is limiting. Finally, we demonstrated that the WPRE element can substitute for both stabilisation and export functions. Although larger (589 bp) than other TRANCER components, WPRE offers a single-insert solution suitable for certain contexts, such as integration into endogenous loci. Future work may enable functional minimisation of this element to further streamline constructs.

Additionally, here we show TRANCERs' stage and tissue specificity conclusively through organoid differentiation of mESCs (Figure 2 and 3), as well as *in vivo* using zebrafish (Figure 3). In summary, TRANCERs provide a modular framework for engineering enhancer-derived transcription into a programmable tool for transgene design. Depending on the application, constructs can be tailored for compactness, cytoplasmic translation, or nuclear confinement. This versatility positions TRANCERs as a broadly applicable technology for basic biology, gene therapy, and genome engineering.

Materials and methods

Cell Culture

Mouse embryonic stem cells (E14TG2a.IV; 129/Ola background) were maintained feeder-free on 0.1% gelatin-coated plates in GMEM supplemented with 10% FBS, 1× MEM-NEAA, 2 mM L-glutamine, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol, 1000 U/mL LIF, and 100 U/mL penicillin–streptomycin at 37 °C and 5% CO_2. Cells were passaged using 0.05% trypsin–EDTA or Accutase and reseeded into fresh medium.

K562 human myelogenous leukemia cells were cultured in RPMI-1640 supplemented with 10% FBS and 2 mM L-glutamine and passaged by dilution into pre-warmed medium.

Cloning and Vector Assembly

All constructs were generated using Gibson HiFi assembly according to the manufacturer's instructions. DNA inserts were either synthesised as gBlocks (IDT) or PCR-amplified from genomic DNA using Q5 High-Fidelity polymerase with gene-specific primers. PCR products and entry vectors were purified prior to assembly. Entry vectors were sourced from in-house stocks (Milne group or MRC WIMM Genome Engineering Core, courtesy of Christian Babbs) and prepared either by restriction digestion or PCR amplification.

Stable cell lines were generated via CRISPR–Cas9-mediated homology-directed repair (HDR). Guide RNAs and 500 bp homology arms were designed using the IDT Alt-R tool unless otherwise specified.

Following assembly, constructs were transformed into *E. coli* DH5α, selected on antibiotic-containing LB agar, and cultured overnight at 37 °C. Plasmid DNA was isolated using standard miniprep procedures, screened by diagnostic digestion, and verified by Sanger sequencing.

Fluorescence-Activated Cell Sorting (FACS) and Flow Cytometry

FACS was used to identify and isolate successfully transfected cells. Live cells were resuspended in FACS buffer (PBS with 10% FBS) and sorted with assistance from Kevin Clark (MRC WIMM Flow Cytometry Facility).

Flow cytometry was performed on an Attune NxT flow cytometer to assess TRANCER signal expression and intensity in either 96-well plates or individual tubes. Wild-type, non-expressing cells were used for gating, and cellular debris and doublets were excluded based on forward (FSC) and side scatter (SSC) parameters. Data were analyzed using FlowCytometryTools for gating and PRISM software for plotting replicate results.

Differentiation of mESCs into Embryoid Bodies (EBs)

Embryoid body (EB) cultures were generated as previously described Francis *et al.*, (2022) [63]. Forty-eight hours prior to differentiation, mESCs were passaged into Iscove's Modified Dulbecco's Medium (IMDM) supplemented with LIF. To initiate differentiation (day 0), cells were dissociated with trypsin and resuspended in freshly prepared differentiation medium consisting of 84.5 mL IMDM, 16.5 mL heat-inactivated ΔFBS, 5.5 mL PFHM II, 1 mL GlutaMAX, 1 mL transferrin, 1 mL amino acids, and 0.0026% (v/v) monothioglycerol (MTG).

Single-cell suspensions were plated in 10 cm non-adherent dishes at 3×10^3 cells per dish and cultured for seven days with daily gentle agitation to prevent EB attachment. EBs were harvested and dissociated in 0.25% trypsin for 5 min at 37 °C, and resulting single cells were analyzed by FACS to detect mNG^+ populations.

Lentivirus Production and Transduction

Lentiviral particles were produced in HEK293T cells using a three-plasmid second-generation packaging system. The packaging plasmid psPAX2 (Addgene #12260) or integrase-deficient psPAX2-D64V (Addgene #63586), the envelope plasmid pMD2.G (Addgene #12259), and the transfer vector Lenti MCP-LSD1 Hygro (Addgene #138457) were co-transfected using Lipofectamine 2000. Viral supernatants were collected 48 and 72 h post-transfection, filtered (0.45 µm), flash-frozen, and stored at −80 °C. Insert sequences were cloned into the transfer vector via NheI/KpnI digestion and Gibson HiFi assembly, and constructs were verified by Sanger sequencing.

For transduction of suspension cells, spinfection was performed at 1000 × g for 90 min at 32 °C in the presence of 8 µg/mL polybrene, followed by antibiotic selection (puromycin or blasticidin) 24 h later. For mESC transduction, Retronectin-coated plates were used to enhance infection. Viral particles were bound to plates by centrifugation at 2000 × g for 2 h at 32 °C, after which 1×10^5 target cells were added and spinfected at 1000 × g for 1 h. Cells were subsequently cultured at 37 °C for recovery and downstream assays.

Single-Cell Multiome (ATAC + RNA) Sequencing

mESCs were differentiated into embryoid bodies (EBs) as described above. FACS was performed on wild-type and R2-CD19 lines, collecting ~1×10^5 cells per sample. $CD19^+$ cells were enriched using APC anti-human CD19 antibody (BioLegend, #302211).

Nuclei isolation was performed according to the 10x Genomics Chromium Single Cell Multiome ATAC + Gene Expression protocol (CG000365, Rev C; low-cell input workflow). Briefly, cells were washed and pelleted (300 × g, 5 min, 4 °C), resuspended in PBS + 0.04% BSA, and lysed in chilled Lysis Buffer. Nuclei were subsequently washed and resuspended in Diluted Nuclei Buffer as per the manufacturer's instructions. The resulting nuclei suspensions were immediately used for 10x Genomics library preparation and sequencing.

Bioinformatic Analysis

All sequencing datasets were processed using the *Seqnado* pipeline (https://github.com/alsmith151/SeqNado) with default parameters and aligned to either the mm39 or hg38 genome builds.

ChIP-seq, ChIPmentation, and ATAC-seq reads were trimmed with *Trim Galore*, aligned using *Bowtie2*, and visualized as normalized bigwig files via *deepTools* on the UCSC Genome Browser. Peaks were called using *LanceOtron*.

RNA-seq, TT-seq, and POINT-seq data were aligned with *STAR*, quantified using *featureCounts*, and visualized with *deepTools* for strand-specific transcription profiles.

Micro-Capture-C data were aligned with *Bowtie2* and filtered for valid interaction pairs before visualization on UCSC.

Single-cell Multiome (ATAC + RNA) data were processed with *Cell Ranger ARC* using a custom mm39 reference including the R2-tCD19 locus. Downstream analysis was performed using *muon*, *ArchR*, and *Seurat*. Quality control included ambient RNA removal (*CellBender*), doublet exclusion (*Scrublet* and *ArchR*), and filtering for cells with >1000 detected genes. Dimensionality reduction and integration of RNA and ATAC modalities were performed with UMAP and *muon*, and cell types were annotated using *cellxgene* transfer learning models.

References

1. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res* 16, 1 (2006).
2. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science (1979)* 361, 1341–1345 (2018).
3. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* 14, 288 (2013).
4. 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, 1725–35 (2003).
5. de Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol* 8, (2010).
6. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (1979)* 339, 950–953 (2013).
7. Andersson, R. *et al.* Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun* 5, (2014).
8. Arnold, P. R., Wells, A. D. & Li, X. C. Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front Cell Dev Biol* 7, 503820 (2020).
9. Kowalczyk, M. S. *et al.* Intragenic enhancers act as alternative promoters. *Mol Cell* 45, 447–458 (2012).
10. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol* 14, 103–105 (2007).
11. Natoli, G. & Andrau, J. C. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* 46, 1–19 (2012).
12. Melo, C. A. *et al.* eRNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Mol Cell* 49, 524–535 (2013).
13. Vernimmen, D., Gobbi, M. De, Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* 26, 2041 (2007).
14. Plant, K. E., Routledge, S. J. E. & Proudfoot, N. J. Intergenic transcription in the human beta-globin gene cluster. *Mol Cell Biol* 21, 6507–6514 (2001).
15. Van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol* 35, 145 (2016).
16. Arnold, M. & Stengel, K. R. Emerging insights into enhancer biology and function. *Transcription* 14, 68 (2023).
17. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16, 144 (2015).
18. Banerji, J., Rusconi, S. & Schaffner, W. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308 (1981).
19. Moreau, P. *et al.* The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9, (1981).
20. Gao, T. *et al.* EnhancerAtlas: a resource for enhancer annotation and analysis in 105 human cell/tissue types. *Bioinformatics* 32, 3543 (2016).

21. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature 2014 507:7493* 507, 455–461 (2014).
22. Kyle Satterstrom, F. & Haigis, M. C. Luciferase-based reporter to monitor the transcriptional activity of the SIRT3 promoter. *Methods Enzymol* 543, 141 (2014).
23. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* https://doi.org/10.1016/j.cell.2018.01.029 (2018) doi:10.1016/j.cell.2018.01.029.
24. Georgiades, E. *et al.* Active regulatory elements recruit cohesin to establish cell specific chromatin domains. *Scientific Reports 2025 15:1* 15, 1–16 (2025).
25. Core, L. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *nature.comLJ Core, AL Martins, CG Danko, CT Waters, A Siepel, JT LisNature genetics, 2014•nature.com* https://www.nature.com/articles/ng.3142.
26. Bhatt, B., García-Díaz, P. & Foight, G. W. Synthetic transcription factor engineering for cell and gene therapy. *Trends Biotechnol* 42, 449–463 (2024).
27. Hernandez-Garcia, C. M. & Finer, J. J. Identification and validation of promoters and cis-acting regulatory elements. *Plant Science* 217–218, 109–119 (2014).
28. Kim, N.-G. *et al.* Minimization of gene editing off-target effects by tissue restriction of expression. https://doi.org/10.1101/2025.01.21.634017 doi:10.1101/2025.01.21.634017.
29. Cavazza, A. *et al.* Advanced delivery systems for gene editing: A comprehensive review from the GenE-HumDi COST Action Working Group. *Mol Ther Nucleic Acids* 36, 102457 (2025).
30. Driscoll, J. *et al.* Using aptamers for targeted delivery of RNA therapies. *Molecular Therapy* 33, 1344–1367 (2025).
31. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia 2018 32:7* 32, 1529–1541 (2018).
32. Kim, H., Kim, M., Im, S. K. & Fang, S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res* 34, 147–159 (2018).
33. Zhang, Z. *et al.* Transcriptional landscape and clinical utility of enhancer RNAs for eRNA-targeted therapy in cancer. *Nature Communications 2019 10:1* 10, 1–12 (2019).
34. Fukaya, T. Enhancer dynamics: Unraveling the mechanism of transcriptional bursting. *Sci Adv* 9, (2023).
35. Song, C. *et al.* eRNAbase: a comprehensive database for decoding the regulatory eRNAs in human and mouse. *Nucleic Acids Res* 52, D81–D91 (2024).
36. Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science (1979)* 352, 1225–1228 (2016).
37. Sauer, B. Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae. *Mol Cell Biol* 7, 2087 (1987).
38. Zhao, W., Blagev, D., Pollack, J. L. & Erle, D. J. Toward a Systematic Understanding of mRNA 3' Untranslated Regions. *Proc Am Thorac Soc* 8, 163 (2011).
39. Brown, J. A., Valenstein, M. L., Yario, T. A., Tycowski, K. T. & Steitz, J. A. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENβ noncoding RNAs. *Proc Natl Acad Sci U S A* 109, 19202–19207 (2012).
40. Kirsebom, L. A. RNase P RNA mediated cleavage: substrate recognition and catalysis. *Biochimie* 89, 1183–1194 (2007).

41. Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained non-coding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 135, 919 (2008).
42. Wilusz, J. E. *et al.* A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly(A) tails. *Genes Dev* 26, 2392 (2012).
43. Zhang, B. *et al.* The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep* 2, 111–123 (2012).
44. Wilusz, J. E. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. *Biochim Biophys Acta* 1859, 128 (2016).
45. Brown, T. A. Synthesis and Processing of RNA. https://www.ncbi.nlm.nih.gov/books/NBK21132/ (2002).
46. Wang, Y. *et al.* Mechanism of alternative splicing and its regulation. *Biomed Rep* 3, 152 (2015).
47. Okamura, M., Inose, H. & Masuda, S. RNA Export through the NPC in Eukaryotes. *Genes 2015, Vol. 6, Pages 124-149* 6, 124–149 (2015).
48. Khan, M., Hou, S., Chen, M. & Lei, H. Mechanisms of RNA export and nuclear retention. *Wiley Interdiscip Rev RNA* 14, e1755 (2023).
49. Lei, H., Zhai, B., Yin, S., Gygi, S. & Reed, R. Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. *Nucleic Acids Res* 41, 2517 (2013).
50. Guang, S. & Mertz, J. E. Pre-mRNA processing enhancer (PPE) elements from intronless genes play additional roles in mRNA biogenesis than do ones from intron-containing genes. *Nucleic Acids Res* 33, 2215–2226 (2005).
51. Mouzannar, K., Schauer, A. & Liang, T. J. The Post-Transcriptional Regulatory Element of Hepatitis B Virus: From Discovery to Therapy. *Viruses 2024, Vol. 16, Page 528* 16, 528 (2024).
52. Rausch, J. W. & le Grice, S. F. J. HIV Rev Assembly on the Rev Response Element (RRE): A Structural Perspective. *Viruses 2015, Vol. 7, Pages 3053-3075* 7, 3053–3075 (2015).
53. Pasquinelli, A. E. *et al.* The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J* 16, 7500 (1997).
54. Gales, J. P., Kubina, J., Geldreich, A. & Dimitrova, M. Strength in Diversity: Nuclear Export of Viral RNAs. *Viruses* 12, (2020).
55. Hope, T. J., Zufferey, R., TronoJohn, D. & Donello, E. RNA export element. (1997).
56. Mastroyiannopoulos, N. P., Feldman, M. L., Uney, J. B., Mahadevan, M. S. & Phylactou, L. A. Woodchuck post-transcriptional element induces nuclear export of myotonic dystrophy 3' untranslated region transcripts. *EMBO Rep* 6, 458 (2005).
57. Schambach, A. *et al.* Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. *Gene Therapy 2006 13:7* 13, 641–645 (2005).
58. Zufferey, R., Donello, J. E., Trono, D. & Hope, T. J. Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element Enhances Expression of Transgenes Delivered by Retroviral Vectors. *J Virol* 73, 2886 (1999).

59. Brun, S., Faucon-Biguet, N. & Mallet, J. Optimization of transgene expression at the posttranscriptional level in neural cells: implications for gene therapy. *Molecular Therapy* 7, 782–789 (2003).
60. Ye, L. *et al.* Enhancement of plasmid-mediated stable gene expression by woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) in human embryonic kidney (HEK293) cells. *Afr J Biotechnol* 11, 10792–10798 (2012).
61. Briggs, J. A. *et al.* Mouse embryonic stem cells can differentiate via multiple paths to the same state. *Elife* 6, (2017).
62. Oudelaar, A. M., Beagrie, R. A., Kassouf, M. T. & Higgs, D. R. The mouse alpha-globin cluster: a paradigm for studying genome regulation and organization. *Curr Opin Genet Dev* 67, 18–24 (2021).
63. Francis, H. S. *et al.* Scalable in vitro production of defined mouse erythroblasts. *PLoS One* 17, (2022).
64. Blayney, J. W. *et al.* Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* 186, 5826-5839.e18 (2023).
65. Lange, M. *et al.* A multimodal zebrafish developmental atlas reveals the state-transition dynamics of late-vertebrate pluripotent axial progenitors. *Cell* 187, 6742-6759.e17 (2024).
66. Urasaki, A., Morvan, G. & Kawakami, K. Functional Dissection of the Tol2 Transposable Element Identified the Minimal cis-Sequence and a Highly Repetitive Sequence in the Subterminal Region Essential for Transposition. *Genetics* 174, 639 (2006).
67. Taskiran, I. I. *et al.* Cell-type-directed design of synthetic enhancers. *Nature 2023 626:7997* 626, 212–220 (2023).
68. Gosai, S. J. *et al.* Machine-guided design of cell-type-targeting cis-regulatory elements. *Nature 2024 634:8036* 634, 1211–1220 (2024).
69. Gaudelli, N. M. *et al.* Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature 2017 551:7681* 551, 464–471 (2017).
70. Katti, A. *et al.* GO: a functional reporter system to identify and enrich base editing activity. *Nucleic Acids Res* 48, 2841–2852 (2020).