A *piggyBac*-based reporter system for scalable *in vitro* and *in vivo* analysis of 3′ untranslated region-mediated gene regulation

Arindam Chaudhury, Natee Kongchan, Jon P. Gengler, Vakul Mohanty, Audrey E. Christiansen, Joseph M. Fachini, James F. Martin and Joel R. Neilson[*]

Department of Molecular Physiology and Biophysics and Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

ABSTRACT

Regulation of messenger ribonucleic acid (mRNA) subcellular localization, stability and translation is a central aspect of gene expression. Much of this control is mediated via recognition of mRNA 3′ untranslated regions (UTRs) by microRNAs (miRNAs) and RNA-binding proteins. The gold standard approach to assess the regulation imparted by a transcript's 3′ UTR is to fuse the UTR to a reporter coding sequence and assess the relative expression of this reporter as compared to a control. Yet, transient transfection approaches or the use of highly active viral promoter elements may overwhelm a cell's post-transcriptional regulatory machinery in this context. To circumvent this issue, we have developed and validated a novel, scalable *piggyBac*-based vector for analysis of 3′ UTR-mediated regulation *in vitro* and *in vivo*. The vector delivers three independent transcription units to the target genome—a selection cassette, a turboGFP control reporter and an experimental reporter expressed under the control of a 3′ UTR of interest. The pBUTR (*piggyBac*-based 3′ *Un*Translated *R*egion reporter) vector performs robustly as a siRNA/miRNA sensor, in established *in vitro* models of post-transcriptional regulation, and in both arrayed and pooled screening approaches. The vector is robustly expressed as a transgene during murine embryogenesis, highlighting its potential usefulness for revealing post-transcriptional regulation in an *in vivo* setting.

INTRODUCTION

Coordinated regulation of gene expression is fundamentally important for cellular division, differentiation and response to environmental cues. While the field of proteomics is rapidly advancing, the most broadly utilized practice in assessing coordinated regulation of gene expression is via analysis of messenger ribonucleic acid (mRNA) steady-state expression using either microarray [1] or Next Generation Sequencing approaches [2–4]. Both of these approaches are exceptionally powerful, providing the ability to simultaneously monitor increases and decreases of the individual gene products comprising the transcriptome. Yet, the information provided by either approach is limited in that it does not effectively reveal whether a given mRNA, irrespective of relative representation, is being actively used by the cell's translational machinery.

Indeed, paired transciptomic and proteomic analyses have revealed varying but significant degrees of discordance between the relative expression of a given mRNA species and the protein encoded by these transcripts [5–9]. The results of these studies are generally consistent with the described role of post-transcriptional regulation of gene expression in virtually every physiological context [10–13]. Interestingly however, a recent study [9] suggested that mRNA translation rates are the major (∼55%) contributor to protein expression in murine fibroblasts. While mRNA degradation rates were found to contribute to protein expression levels to a much lower extent in this study (∼5%), the results are consistent with a model where the aggregate regulation of mRNA stability and translation plays a significant, if not dominant, role in protein expression.

Much of the control of gene expression at the mRNA level is thought to be conferred via *cis*-regulatory elements in the non-coding 3′ untranslated region (UTR) of a given mRNA [11,14,15]. In various physiological states, these *cis*-regulatory elements may be recognized by microRNAs

(miRNAs) or RNA-binding proteins dictating transcript localization, stability and/or translation. In some systems, 3′ UTR identity is largely sufficient to confer correct temporospatial gene expression *in vivo* (14). Recently, much interest has been accorded to both alternative splicing (16) and alternative cleavage and polyadenylation (17), both of which may alter 3′ UTR identity and thus visibility of related gene products to the post-transcriptional regulatory machinery. Given that mutations within the UTRs of certain genes can significantly impact human health, for example muscular dystrophy and schizophrenia (18–20), it is of great interest to determine if and how genomic variations within the 3′ UTR, uncovered via genome-wide association studies and other high-throughput screening approaches, impact the pathology of the disease or phenotype with which they are associated. To this end, a scalable and robust reporter system for modeling these variations is desirable.

The most common strategy for studying the impact of 3′ UTR identity on protein expression is to fuse a 3′ UTR of interest to a non-native reporter gene. Comparison of the relative expression of this reporter to a second (control) reporter can then isolate differences in post-transcriptional regulation between the two reporters. Transient transfection approaches are widely used for this type of study, but are inherently unsuited for analysis over time (21) and may obscure or under-represent endogenous regulation due to saturation of the cell's regulatory machinery. The latter concern must also be taken into consideration when transient transduction of cells with a viral vector (e.g. adenovirus) is employed. While utilization of a single-copy genomic integrant circumvents many of these issues, the use of retro- or lentiviral vector systems for monitoring 3′ UTR-based regulation also suffers from several drawbacks. For example the difficulty in inclusion of a completely distinct control reporter, compulsory inclusion of irrelevant vector-derived sequence and the potential presence of commonly used stability elements [e.g. the woodchuck hepatitis virus post-translationally regulated element (WPRE)] (22) confounds native post-transcriptional control, and retro- and lentiviral long terminal repeat elements may be silenced over time by the cell both *in vitro* and *in vivo* (23–27).

To circumvent some of these limitations and facilitate higher-throughput analysis of 3′ UTR-mediated control of gene expression *in vitro* and *in vivo*, we have engineered and validated a novel, scalable *piggyBac* transposon-based reporter system, *pBUTR* (*piggyBac*-based 3′ *Un*Translated *R*egion reporter). Originally isolated from the genome of the cabbage looper moth *Trichoplusia ni* (28), the *piggyBac* transposon has a large cargo size (29), is highly active in many cell types (30,31) and has been shown to effect long-term expression in mammalian cells *in vivo* (32). The *pBUTR* vector system is comprised of three independent transcription units—a G418 selection cassette, a control turbo-green fluorescent protein (tGFP) reporter cassette and a Gateway® (33) recombineering cassette under the control of the Ubiquitin C (UbC) promoter. Upon recombination of turboRFP (tRFP), a 3′ UTR of interest, and a barcoded minimal polyadenylation site into this cassette, a bi-fluorescent reporter vector is produced that can be employed in both *in vitro* and *in vivo* model systems. Here

we assess the performance of the *pBUTR* vector/reporter in the context of synthetic RNA interference (RNAi)-based siRNA/miRNA sensor activity, established models of post-transcriptional regulation by miRNAs and RNA binding proteins, arrayed and pooled screening approaches and in the context of murine embryogenesis. The reporter performs robustly in each of the scenarios tested and has the potential to be a valuable tool for prospective characterization of the impact of 3′ UTR identity on gene regulation and function in both cell-based and animal studies.

MATERIALS AND METHODS

pBUTR vector construction

The *piggyBac* (*pB*) transposon, pTpB, a generous gift from Dr Matthew H. Wilson (34), was used as the backbone of the destination vector. The bovine growth hormone (Bgh) polyadenylation site was amplified from pUbC-KBPA-iFGFR1-F2A-Luc2-E (a kind gift of Dr Jeffrey M. Rosen hereafter referred to as pUbC) using oligos containing XmaI and AgeI sites and subcloned into the AgeI site (located at the 3′ end of the G418 cassette) of the pTpB vector. The 5′-*attR1* flanked *Cm^R*/*ccdB* cassette was amplified from pDEST17 (Life Technologies, Carlsbad, CA, USA) with oligonucleotides containing PstI and EcoRI sites and subcloned into the PstI and EcoRI sites of pUbC. A synthetically generated attR5 was subcloned into the EcoRI and XhoI sites of the *modified* pUbC. The complete *attR1-Cm^R*/*ccdB-attR5* cassette was subsequently subcloned from the modified pUbC into the pTpB transposon using SacII sites. The *Pgk* promoter was amplified from the pL45 vector and inserted into pTurboGFP with Nde1 and EcoRI. The chimeric intron from Rr-Luc-6xCXCR4 was amplified and subcloned into the EcoRI and AgeI sites of pTurboGFP. Finally, the *Pgk*-chimeric intron-turbo GFP-SV40 polyadenylation cassette was amplified from the modified pTurboGFP plasmid into the SacII and BamHI sites of the pTpB vector.

Entry vector construction

pDONR 223 *attP2r-attP4* was a kind gift of Kenneth Scott (Baylor College of Medicine). To engineer the pDONR *attP4r*/*P5* entry vector, the *SalI* fragment was excised from pDONR 223 *attP2r-attP4*, leaving only the *attP4* site in the vector backbone. Separately, the *EcoRV*/*SalI* fragment containing the *Cm^R*/*ccdB* cassette was subcloned into a synthetic plasmid containing an *XbaI-EcoRV-Xho1-attP5-NheI* insert in the pIDTSMART vector (IDT DNA Technologies, Coralville, IA, USA). Finally, the *XbaI*/*NheI* fragment was subcloned from the modified pIDTSMART vector into the initial pDONR223 construct from which the *SalI* fragment had been excised.

Construction of donor vectors

The Turbo-RFP donor plasmid was generated by polymerase chain reaction (PCR) amplification of the coding sequence of the turbo-red fluorescent protein (tRFP) from its commercially available vector (Evrogen, Farmingdale, NY, USA) using primers containing the *attB1* and

attB2 Gateway® recombination sequences. Donor plasmids with siRNA/miRNA sensor elements or various 3′ UTRs were generated by PCR amplification of synthetic oligonucleotides or UTR elements using primers containing the Gateway® *attB2r and attB4* recombination sequences. The minimal polyadenylation/barcode element donor plasmids were generated by amplification of an oligonucleotide (*attR4_mPA_barcode_attL5 oligo*) using primers containing Gateway® *attB4r* and *attB5* recombination sequences. Barcodes within the *attR4_mPA_barcode_attL5* oligonucleotide were generated using a sequence of mixed bases corresponding to the nucleotide frequency of bases following endogenous polyadenylation sites in the human genome. Following blood pressure recombination and transformation in Top10 competent cells, entry clones were screened through colony PCR using M13 forward and reverse primers and all putative positive clones were sequence verified. Relevant oligonucleotide sequences are listed in Supplementary Table S1.

Construction of expression reporters

Complete expression reporters were generated via four part recombineering using the destination vector with equimolar amounts of the three donor plasmids—the tRFP entry clone, the donor plasmid containing the 3′ UTR of interest and the pool of donor plasmids containing the minimal polyadenylation signal (35) and barcode—using Gateway LR Clonase II enzyme mix (Life Technologies, Carlsbad, CA, USA). Recombination reactions were transformed into One Shot Mach-1 competent cells (Life Technologies, Carlsbad, CA, USA), which were plated on LB-Agar containing both ampicillin and kanamycin. All properly recombined expression vectors were initially identified via colony PCR and subsequently sequence verified.

Cell culture and treatment

Hela, MCF7, MCF10A and U937 cell lines were obtained from ATCC (Manassas, VA, USA). MCF10A cells were cultured in Dulbecco's modified Eagle's medium (DMEM)/F12 medium (Life Technologies, Carlsbad, CA, USA), containing 5% horse serum (Life Technologies, Carlsbad, CA, USA), 0.01-mg/ml bovine insulin (Cell Applications, San Diego, CA, USA), 0.5-μg/ml hydrocortisone (Sigma-Aldrich, St. Louis, MO, USA), 100-ng/ml cholera toxin (Sigma-Aldrich, St. Louis, MO, USA), 20-ng/ml human EGF (Peprotech, Rocky Hill, NJ, USA) and 100-U/ml penicillin and 0.1-mg/ml streptomycin (Life Technologies, Carlsbad, CA, USA). MCF7 and Hela cells were cultured in DMEM medium, containing 10% fetal bovine serum (FBS) (Lonza, Walkersville, MD, USA) and 100-U/ml penicillin and 0.1-mg/ml streptomycin. U937 cells were cultured in RPMI1640 (ATCC, Manassas, VA, USA), containing 10% FBS and 100-U/ml penicillin and 0.1-mg/ml streptomycin. Cells were kept at 37°C under a humidified atmosphere of 5% CO_2. Where indicated, MCF7 and MCF10A cells were treated with TGFβ1 (R&D Systems, Minneapolis, MN, USA) at a final concentration of 5 ng/ml for 72 h. U937 cells were treated with lipopolysaccharide (LPS) from *Escherichia coli* 026:B6 (Sigma-Aldrich, St. Louis, MO, USA) at a final concentration of 1 μg/ml for 24 h.

Transfection and stable clone generation

Hela cells (2×10^5) were transfected using Lipofectamine-2000 (Life Technologies, Carlsbad, CA, USA). MCF7 (4×10^4), MCF10A (4×10^4) and U937 (10^5) cells were transfected using Lipofectamine LTX (Life Technologies, Carlsbad, CA, USA), each according to the manufacturer's instructions. Plasmids containing transposase (*pCMV-HA-m7pB*) and transposon (respective *pBUTR* vector) were used at a ratio of 1:2. Forty-eight hours after transfection, cells were split 1:10 and subsequently selected with G418 (1000 μg/ml for MCF7, MCF10A and Hela and 250 μg/ml for U937 cells) (Teknova, Hollister, CA, USA) for ~2 weeks. Indicated miRNA mimics and inhibitors (Life Technologies, Carlsbad, CA, USA) were used at 30 nM, whereas *ZFP36* siRNA (Life Technologies, Carlsbad, CA, USA) was used at 300 nM. Transiently transfected cells were harvested and analyzed at 24 h post transfection.

Flow cytometry and cell sorting

Expression of tGFP, tRFP, E-Cadherin and CD86 were determined by flow cytometry using a FACSCalibur system (BD Biosciences). For assessing E-Cadherin expression, MCF7 and MCF10A cells were scraped in calcium and magnesium free phosphate buffered saline (PBS) (Life Technologies, Carlsbad, CA, USA) containing 1-mM ethylenediaminetetraacetic acid (EDTA), pelleted at 129 g for 3 min and stained with allophycocyanin (APC) CD324 (E-Cadherin) antibody (clone 67A4) (BioLegend, San Diego, CA, USA). For CD86 protein expression, U937 cells were pelleted at 290 g for 5 min, then stained with APC CD86 antibody (BioLegend, San Diego, CA, USA). Hela cells were harvested by trypsinization. All cells were resuspended in PBS containing 2% FBS and 1-mM EDTA and counterstained with 10-μg/ml propidium iodide (PI) (Roche Diagnostics, Indianapolis, IN, USA) to allow exclusion of dead cells. At least 30 000 events were collected for each analysis. Data were analyzed using FlowJo version 9 (Tree Star, Ashland, OR, USA).

To obtain reference points for setting the flow cytometry gates for cell sorting, single dye positive and negative control samples were prepared for each of the fluorescent signals (tGFP, tRFP, E-cadherin and PI) used. The top 10% of tRFP positive population in indicated MCF10A cells (which were all GFP positive) were sorted by a BD FACS Aria II cell sorter (BD Biosciences) using 100-μm filter and 20-psi nozzle pressure. The cells were collected in FBS and immediately processed for genomic deoxyribonucleic acid (DNA) isolation.

Genomic DNA isolation, library preparation and limited next generation sequencing

Genomic DNA was isolated from indicated populations of MCF10A cells using overnight lysis (100-mM NaCl, 20-mM Tris, pH 7.6, 10-mM EDTA, pH 8.0, 0.5% odium dodecyl sulphate and 0.5-mg/ml proteinase K) at 55°C

before salting out with 60% volume saturated NaCl and ethanol precipitation. Barcoded PCR primer pairs (Supplementary Table S1) were designed and used to amplify the 3′-UTR-intrinsic unique barcode elements in the indicated populations. An Ion Torrent adapter-ligated library was made following the manufacturer's Ion Plus Fragment Library Kit (Life Technologies, Carlsbad, CA, USA) protocol (#4471252, Revision 3.0). The resulting libraries were purified using AMPure beads (Agencourt, Beckman Coulter, Brea, CA, USA) and the concentration was determined using Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA). Sample emulsion PCR < emulsion breaking and enrichment were performed using the Ion PGM Template OT2 200 Kit (#4480974, Revision 5.0) following manufacturer's instructions. The samples were prepared for sequencing using the Ion PGM 200 Sequencing Kit (#4474004, Revision C) and the complete samples were loaded on an Ion 314 chip and sequenced on the PGM. Data from the PGM runs were processed initially using the bam2fastq to generate the fastq files and custom Perl scripts were used to trim adapter sequences, filter and to determine the percent representation of the different barcodes in the indicated populations.

Preparation of whole cell lysates and immunoblot analysis

Cells were lysed in buffer containing 25-mM Tris-HCl pH 7.4, 150-mM NaCl, 1-mM EDTA, 1% NP-40 and 5% glycerol containing complete, Mini protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN, USA). Ten micrograms of whole cell lysate was resolved on a Nu-PAGE 4–20% gel (Life Technologies, Carlsbad, CA, USA), transferred to an Immobilon PVDF membrane (Millipore, Billerica, MA, USA) and probed with E-cadherin antibody (clone 24E10) (Cell Signaling Technology, Danvers, MA, USA) and N-cadherin antibody (Cell Signaling Technology, Danvers, MA, USA). The blot was subsequently stripped and re-probed for Hsp90 (clone C45G5) (Cell Signaling, Danvers, MA, USA) to confirm equal loading. The blots were imaged using enhanced chemiluminescence (ECL) Plus western blotting substrate (Pierce, Rockford, IL, USA) and HyBlot chemiluminescence (CL) autoradiography film (Denville Scientific, Metuchen, NJ, USA).

Generation and injection of *pBUTR-ZEB1* expressing embryonic stem cells (ESCs)

All experiments were performed in accordance with the guidelines of the Baylor College of Medicine Institutional Animal Care and Use Committee. V6.5 ESCs (36) derived from F1 hybrid strain (C57BL/6 × 129/Sv) were a kind gift from Dr Rudolf Jaenisch. ESCs were cultured at 37°C in 5% CO_2 in complete ES medium composed of DMEM (Millipore, Billerica, MA, USA), 15% FBS (Hyclone, Rochester, NY, USA), 1000-U/ml leukemia inhibitory factor (LIF) (ESGRO, Millipore, Billerica, MA, USA), 1% β-mercaptoethanol (Millipore, Billerica, MA, USA), 1% non-essential amino acids (Millipore, Billerica, MA, USA), 1% L-glutamine (Life Technologies, Carlsbad, CA, USA), 0.5% penicillin/streptomycin (Life Technologies, Carlsbad, CA).

V6.5 ESCs (5×10^6 cells) in suspension in 1X PBS were nucleofected using a Bio-Rad GenePulser (Hercules, CA, USA) with 1 μg each of *pBUTR-ZEB1* wild-type (wt) vector and *pCMV-HA-m7pB* transposase (38). Post-nucleofection, cells were plated onto 100-mm culture dishes with a feeder layer. After selection with G418 (300 μg/ml) for 8 days, resulting ESC colonies were verified for tGFP and tRFP expression, expanded and frozen. *pBUTR-ZEB1* wt ES cell clones were injected into 2-N (3.5-days post coitus) C57BL/6 blastocysts and subsequently transferred to the uterine horns of 2.5-days post coitus pseudopregnant Imprinting Control Region (ICR) recipient female mice.

### Embryo harvest, tissue preparation and imaging

Pregnant females were sacrificed by carbon dioxide asphyxiation at day 11.5 postcoitus (e11.5). Embryos were dissected and fixed in 4% paraformaldehyde for 1 h, cryoprotected in 15% and 30% sucrose in PBS and embedded in optimal cutting temperature (OCT) compound (Sakura Finetek, Torrance, CA, USA) before cryosectioning at 10 μM. The sections were mounted on SuperFrost Plus slides (Thermos Fisher Scientific, Houston, TX, USA) using Vectashield (Vector Labs, Burlingame, CA, USA).

Images were obtained using a Zeiss LSM 510 META confocal laser scanning microscope (Carl Zeiss MicroImaging, Thornwood, NY, USA). Each tissue section was initially centered manually using a Carl Zeiss EC Plan-Neofluar 10x/0.3 objective. Sections were tile scanned using a 5×8 grid pattern (at 898.24 μm^2/grid) allowing for a resolution of 512×512 pixels per field at a depth of 28.29 μm.

RESULTS

Vector design and construction

We considered that an ideal vector for monitoring post-transcriptional regulation at the mRNA level would contain three independent transcription units, including a selection marker, a control expression cassette and the experimental expression cassette. Further considerations included a desire for ubiquitously expressed cellular (rather than viral) promoter elements to reduce the risk of saturation of the post-transcriptional regulatory machinery, and the inclusion of chimeric introns in each of the two reporter elements to ensure reasonable expression levels in the absence of non-endogenous and potentially confounding RNA stability elements. We thus modified the previously described *pTpB* vector [(34)—a kind gift from Dr Matthew Wilson, Baylor College of Medicine] to conform to this architecture (Figure 1). Briefly, the promoter and coding sequence of the pre-existing *Neo/G418* selection cassette in this vector were retained and terminated by introducing a Bgh polyadenylation signal. 3′ to this selection cassette, we inserted an Ubiquitin C (UbC) promoter element upstream of a modified Gateway® selection cassette in which the standard chloramphenicol resistance marker (*Cm^R*) and *ccdB* bacterial suicide gene were flanked by *attR1* and *attR5* recombination sites. Finally, a tGFP cassette driven by the murine phosphoglycerate kinase 1 (*Pgk*) promoter was inserted downstream of the former two elements such that it
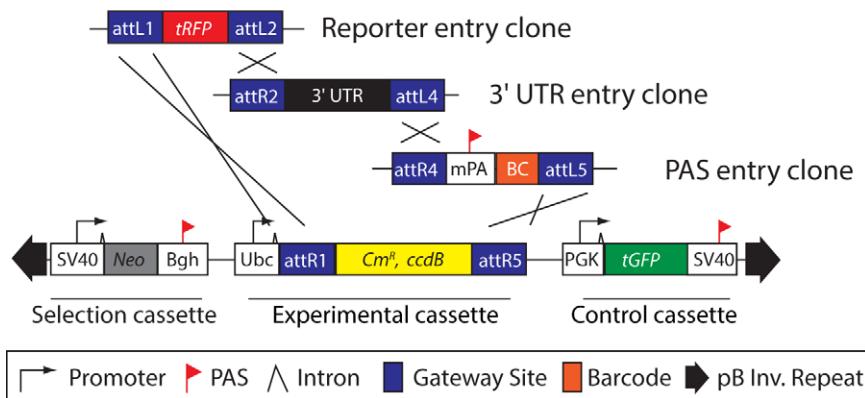
Figure 1. Schematic representation of the *pBUTR* vector backbone. attXN, Gateway® recombination site; tRFP, turboRFP; UTR, untranslated region; mPA, minimum polyadenylation signal (35); BC, 24-nt barcode; PAS, polyadenylation signal; SV40 (left), SV40 early promoter region. Neo, neomycin-resistance gene; Bgh, bovine growth hormone polyadenylation signal; UbC, Ubiquitin C promoter element; Cm^R, chloramphenicol-resistance gene; PGK, murine phosphoglycerate kinase 1 promoter; tGFP, turboGFP; SV40 (right) SV40 late polyadenylation signal. Features not to scale.

was terminated by SV40 late polyadenylation signal present in the original parent vector.

The unconventional configuration of the *attR* sites in the completed *pBUTR* (*piggyBac* 3′ *UnTranslated Region* reporter) destination vector was designed to take advantage of existing local open reading frame (ORF) entry clone libraries existing in an *attL1/L2* format. The *pBUTR* destination vector is functionalized by four-part Gateway® recombineering using an *attL1/L2*-flanked coding sequence of interest, an *attR2/attL4*-flanked 3′ UTR element and an *attR4/attL5*-flanked minimal polyadenylation sequence (35) followed by a unique 24-nucleotide barcode. The composition of the barcode, generated via mixed nucleotide synthesis, was informed by the average nucleotide composition of the 24 base pairs following the G/U-rich region of native polyadenylation sequences in the human genome. The inclusion of unique barcode elements with the minimal polyadenylation signal was made to allow analyses within pooled cell populations via flow cytometry and cell sorting.

Validation of *pBUTR* functionality using synthetic miRNA sensors and response to RNAi-mediated targeting

We first assessed the performance of the *pBUTR* vector as a siRNA/miRNA sensor, recombineering tRFP under the control of three distinct 3′ UTR elements. Each element contained a tandem duplicate of synthetic sequence perfectly complementary to the broadly used synthetic *CXCR4* siRNA (37), the mature human *miR-17* miRNA or the mature human *miR-124* miRNA. Each of the constructs, along with the *pCMV-HA-m7pB* transposase (38,39), was individually transfected into the Hela cell line. Stable transfectants were isolated by G418 selection and tGFP expression was verified by flow cytometry. Hela cells are known to express *miR-17* (40); however, *miR-124* expression is largely limited to neuronal lineages (41). As expected, tRFP expression (monitored by flow cytometry) in the Hela cells stably transfected with the *miR-17* sensor (mock treatment, Figure 2C) was significantly reduced as compared to Hela cells stably transfected with either the sensor to the synthetic *CXCR4* siRNA (mock treatment, Figure 2A) or the *miR-124* sensor (mock treatment, Figure 2B). Transient transfection of the

cells carrying the *CXCR4* sensor with the *CXCR4* siRNA, or cells carrying the *miR-124* sensor with a synthetic *miR-124* mimic, resulted in a marked decrease in tRFP expression (Figure 2A and B). In contrast, transient transfection of the cells carrying the *miR*-17 sensor with an *miR-17* 'antagomir' (42) resulted in a robust increase in tRFP expression (Figure 2C). No change in tGFP protein expression was observed in any of the examined conditions (Figure 2). Taken together, these experiments confirm the functionality of the *pBUTR* vector system, including the $G418^R$ selection marker, the tGFP control expression cassette and the recombineered tRFP experimental expression cassette. The results of the experiments suggest that the *pBUTR* vector is a valid reagent for *in situ* monitoring of miRNA or siRNA activity via synthetic, perfectly complementary target sites.

Monitoring endogenous 3′ UTR-mediated post-transcriptional regulation by miRNAs using the *pBUTR* system

Given the validation of the functionality of our bi-fluorescent *pBUTR* reporter system in the context of synthetic miRNA/siRNA sensor elements, we next set out to assess the performance of the reporter system in monitoring well-characterized models of post-transcriptional regulation by miRNAs and RNA-binding proteins.

We first assessed miRNA-mediated repression in a cell-based model of epithelial to mesenchymal transition (EMT). The E-cadherin transcriptional repressors ZEB1 (also known as δEF1) and ZEB2 (also known as SIP1) play established roles in EMT and tumor metastasis (43). The mRNA transcripts of both of these gene products are characterized by multiple, validated *miR-200* family recognition elements in their respective 3′ UTRs (43). Cells with an epithelial phenotype express high relative levels of the *miR-200b* miRNA, which enforces post-transcriptional repression of the *ZEB1* and *ZEB2* mRNA transcripts. However, as cells undergo EMT, for example in response to transforming grown factor-beta (TGF-β), relative levels of *miR-200b* are reduced, allowing increased expression of ZEB1 and ZEB2 proteins and transcriptional repression of the *CDH1* (E-cadherin) gene.
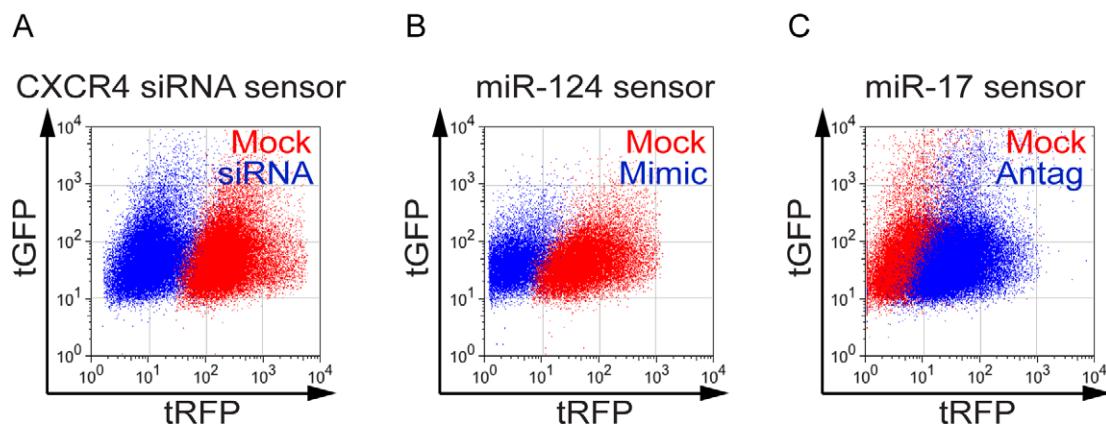
Figure 2. Use of the *pBUTR* system as a siRNA or miRNA sensor. Flow cytometric analysis of Hela cells stably transfected with *pBUTR* recombineered to contain tRFP fused to synthetic 3′ UTR elements in which two tandem sequences perfectly complementary to the indicated siRNA/miRNA are present. (**A**) Response of Hela cells carrying a *CXCR4* sensor to transiently transfected *CXCR4* siRNA (blue) or mock transfection (red). (**B**) Response of Hela cells carrying an *miR-124* sensor to transiently transfected *miR-124* mimic (blue) or mock transfection (red). (**C**) Response of Hela cells carrying a tRFP *miR-17* sensor to transiently transfected anti-*miR-17* antagomir (blue) or mock transfection (red). All data are representative of a minimum of three individual experiments performed 24 h post-transient transfection on G418-selected cells. tRFP, turboRFP; UTR, untranslated region; Antag, antagomir.

Previously described (41) wt and mutant (where each *miR200b*-binding site has been ablated via site-directed mutagenesis) *ZEB1* and *ZEB2* 3′ UTR elements were recombineered into the *pBUTR* destination vector so as to confer regulation upon tRFP expression in the assembled reporter. Both the human non-transformed mammary epithelial cell line MCF10A and the human breast adenocarcinoma cell line MCF7 were stably transfected with each of the four resulting *pBUTR* reporters. Following G418 selection, cells were treated with TGF-β or vehicle for 72 h. As expected, MCF10A cells switched from polarized, tightly packed discoid epithelial cells to highly motile fibroblastic or mesenchymal phenotype, characteristics of distinct morphological changes associated with EMT (44,45), while MCF7 cells, which are refractory to TGF-β-mediated EMT (44,46), maintained epithelial morphology (Figure 3A). MCF-10A-specific EMT was further verified in the stably transfected cell lines via immunoblot, which demonstrated a reduction in E-cadherin protein expression concomitant with an induction of the mesenchymal cell marker N-cadherin in TGF-β-treated MCF10, but not MCF7, cells (Figure 3B).

We next employed multicolor flow cytometry to assess the expression of the wt and mutant *ZEB* reporters in each of the stably transfected cell lines. As expected, treatment of MCF-10A cells with TGF-β resulted in decreased surface levels of E-cadherin protein. In cells stably transfected with the wt *ZEB1* and *ZEB2* reporters, these decreased levels coincided with marked increases in tRFP fluorescence. The levels of tRFP fluorescence in these cells, as assessed via median fluorescence intensity, were similar to those observed in MCF-10A cells transfected with mutant ZEB1 and ZEB2 reporters. In the latter cells, although decreased surface levels of E-cadherin were observed upon TGF-β treatment, tRFP levels remained constant (Figure 3C). TurboGFP fluorescence levels were essentially unchanged in all experimental conditions (Supplementary Figure S1A).

In contrast to the stably transfected MCF10A cells, surface E-cadherin levels in TGF-β-treated MCF7 cells were

effectively identical when compared to mock treated controls (Figure 3D). tRFP fluorescence in these cells was significantly elevated in MCF7 cells carrying the mutant, as compared to the wt, *ZEB1* and *ZEB2* reporters. Once again, levels of tGFP fluorescence derived from the control reporter cassette were largely consistent (Supplementary Figure S1B).

Taken together, these experiments are consistent with well-characterized models (43) in which the *miR-200* family binding sites in the *ZEB1* and *ZEB2* mRNA transcripts mediate negative post-transcriptional regulation of these transcripts by *miR-200b* in the epithelial, but not the mesenchymal state. Nonetheless, to further support this conclusion we examined the response of the reporters to transient transfection of anti-*miR-200b* antagomirs and *miR-200b* mimics in mock treated MCF7 and MCF10A and TGF-β-treated MCF10A cells. Anti-*miR-200b* transfection resulted in increased relative tRFP fluorescence derived from wt, but not mutant, *ZEB1* and *ZEB2* reporters in both untreated MCF7 and MCF10A (Figure 4A and B and Supplementary Figure S2A and B). In contrast, transient transfection of *miR-200b* mimic into TGF-β-treated MCF10A cells resulted in decreased relative tRFP fluorescence derived from wt, but not mutant, *ZEB1* and *ZEB2* reporters (Figure 4C and Supplementary Figure S2C), confirming that *miR-200b* activity was both necessary and sufficient for post-transcriptional regulation of the *ZEB1* and *ZEB2* reporters in this model of EMT.

Performance of *pBUTR* as a reporter of 3′ UTR-mediated post-transcriptional regulation by RNA-binding proteins

Given the performance of the *pBUTR* reporter in the context of monitoring defined miRNA targets in a physiologically relevant context, we next assessed whether it performed similarly in the context of regulation by RNA-binding proteins. One of the best-characterized examples of this type of regulation is observed within myeloid cells of the immune system, where tumor necrosis factor alpha (*TNF-*
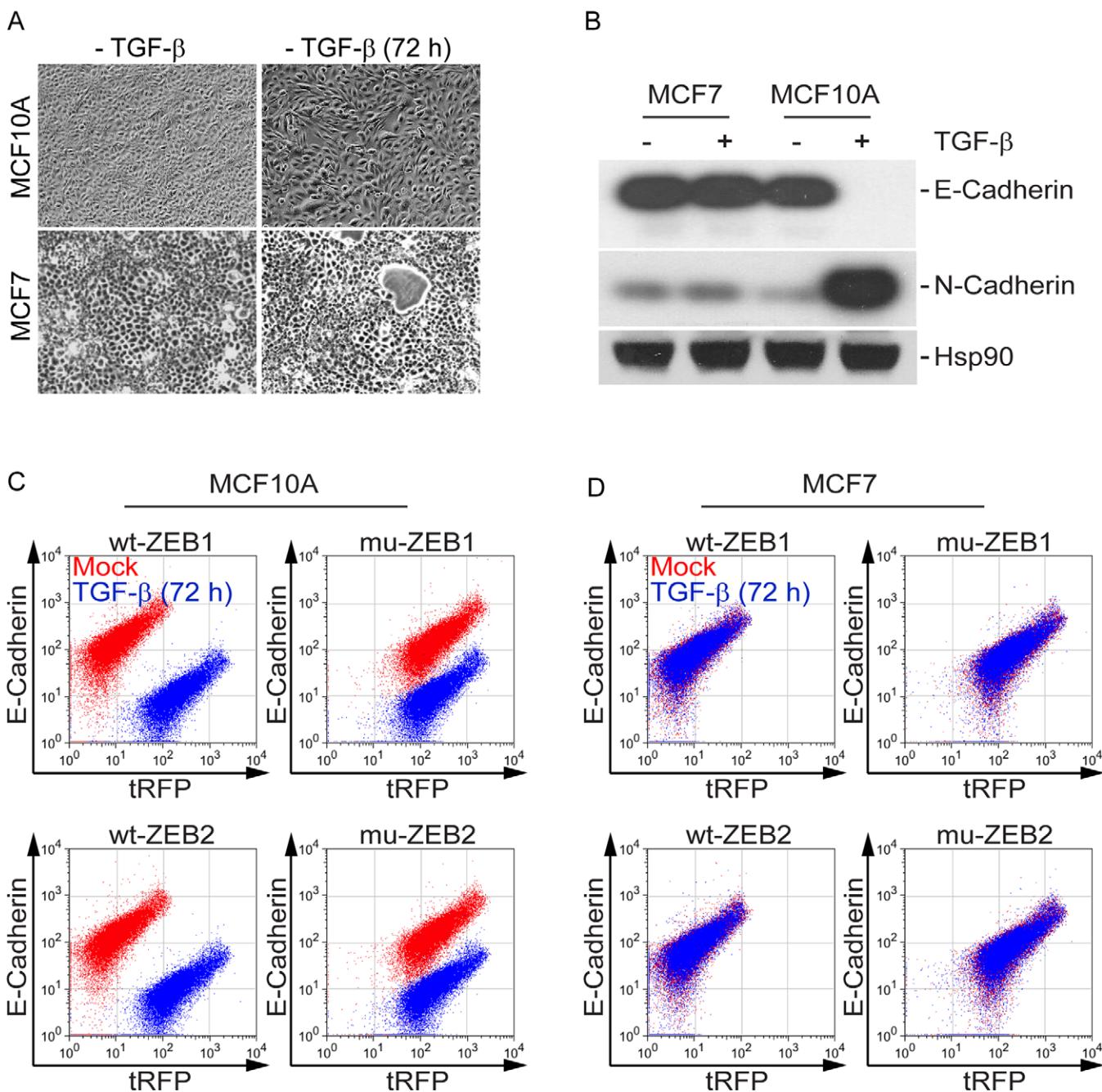
Figure 3. Monitoring endogenous miRNA-mediated regulation *in vitro*. (**A**) Phase contrast micrographs of untreated or TGF-β-treated MCF10A (top panels) and MCF7 (bottom panels) cells. Images were obtained at ×10 magnification. (**B**) Immunoblot (IB) analysis of E-cadherin (upper panel) and N-cadherin (middle panel) protein levels in MCF7 and MCF10A cells treated with TGF-β for 72 h. The blot was stripped and re-probed for Hsp90 (bottom panel) as a loading control. (**C**) and (**D**) Flow cytometric analysis of mock treated (red) and TGF-β-treated (blue) MCF10A (C) and MCF7 (D) cells transfected with vector recombineered to contain tRFP under the control of the wild-type (wt) *ZEB1* (top panels) or *ZEB2* (bottom panels) 3′-UTR, which is responsive to *miR-200*, or a mutant (mu) 3′-UTR in which the *miR-200* recognition elements are deleted (43). All data are representative of a minimum of three individual experiments. UTR: untranslated region.

α) mRNA stability is reduced in response to signaling by the Toll-like Receptor 4 (TLR4) receptor (47). While stimulation of TLR4 by LPS stimulates transcription of the *TNF-α* mRNA, at the same time levels of the tristetraprolin (also known as ZFP36) protein are increased, which destabilizes the *TNF-α* mRNA transcript via recognition and binding of

adenosine/uracil-rich element (ARE) located in the 3′ UTR of the transcript (47).

We recombineered tRFP into the *pBUTR* vector under the control of wt (wt-*TNF-α*) or mutant (Δ-*TNF-α*, in which the ARE was deleted via site-directed mutagenesis) TNF-α 3′ UTR elements. These vectors were used to transfect the human monocytic macrophage cell line U937,
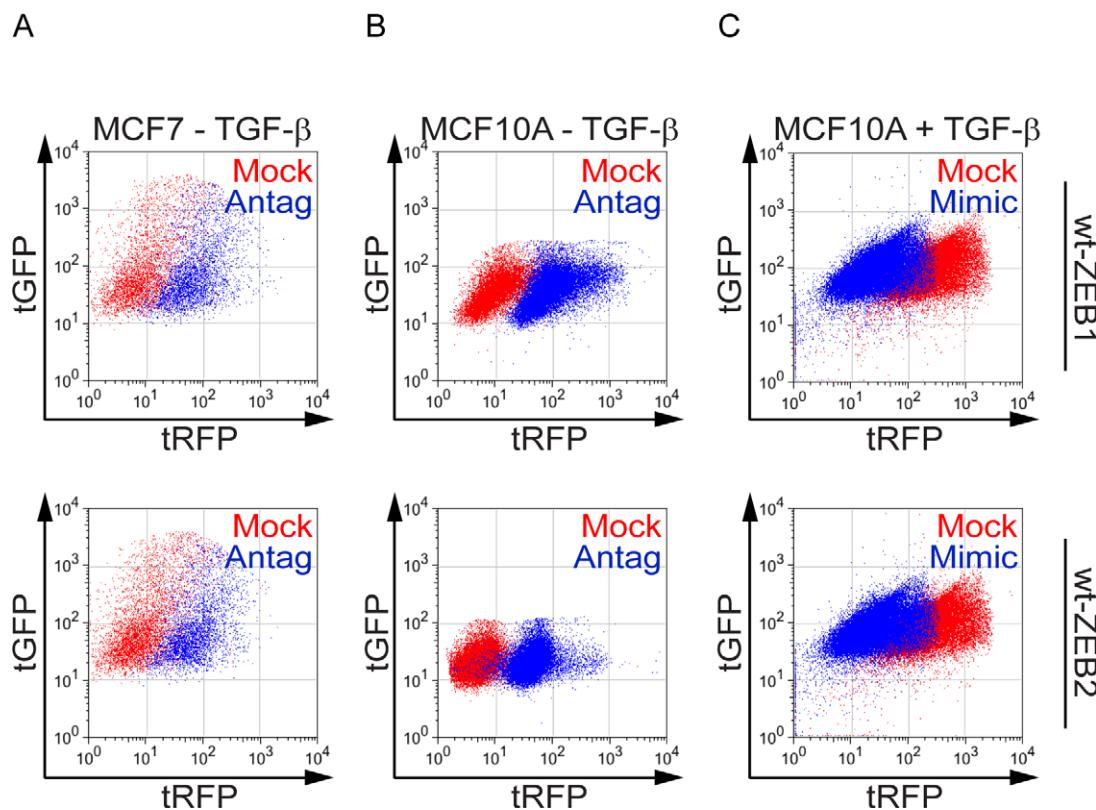
Figure 4. Differential regulation of tRFP expression in the *pBUTR ZEB1* and *ZEB2* 3′ UTR reporters was mediated by *miR200b* expression levels associated with EMT. (**A**) and (**B**) Flow cytometric analysis 24 h post-transient transfection of anti-*miR-200b* antagomir (blue) in untreated MCF7 (A) and MCF10A (B) cells transfected with vector recombineered to contain tRFP under the control of the wild-type (wt) *ZEB1* (top panels) or *ZEB2* (bottom panels) 3′ UTR. (**C**) Flow cytometric analysis 24 h post-transient transfection of anti-*miR-200b* mimic (blue) in TGF-β-treated (72 h) MCF10A cells transfected with vector recombineered to contain tRFP under the control of the wild-type (wt) *ZEB1* (top panels) or *ZEB2* (bottom panels) 3′ UTR. All data are representative of a minimum of three individual experiments. tRFP, turboRFP; UTR, untranslated region.

and stable transfectants were obtained via G418 selection. Stably transfected cells were mock-treated or treated with LPS for 24 h and then analyzed by flow cytometry using CD86 protein induction as a marker for activation (Figure 5A). Unstimulated U937 cells stably transfected with the wt-*TNF-α* (Figure 5B) and Δ-*TNF-α* (Figure 5C) reporters were characterized by robust tRFP fluorescence. As expected, tRFP fluorescence was markedly decreased in LPS-treated U937 cells stably transfected with the wt-*TNF-α* reporter (Figure 5B). In contrast, no reduction in tRFP fluorescence was observed in LPS-treated U937 cells stably transfected with the Δ-*TNF-α* reporter (Figure 5C). Once again, tGFP fluorescence was essentially identical in each case (Figure 5B and C).

To further confirm that the decrease in wt-TNF-α reporter stability and hence tRFP fluorescence was due to ZFP36 protein activity, U937 cells harboring the wt or mutant TNF-α 3′ UTR elements were transiently transfected with siRNAs targeting *ZFP36* before being treated with LPS. Silencing of *ZFP36* gene expression (Figure 5D) completely attenuated reduction in tRFP fluorescence in LPS-treated U937 cells stably transfected with the wt-TNF-α reporter (Figure 5E). No effect on tRFP fluorescence, as expected, was observed when *ZFP36* expression was silenced in LPS-treated U937 cells stably transfected with the Δ-TNF-α reporter (Figure 5F). Taken together, our ex-

periments are consistent with the notion that the *pBUTR* vector/reporter system is well suited for monitoring known or novel instances of 3′ UTR-mediated post-transcriptional regulation, whether by miRNAs or RNA-binding proteins, in a broad spectrum of *in vitro* models of cellular physiology.

Performance of *pBUTR* in a pooled screening approach to monitor endogenous 3′ UTR-mediated post-transcriptional regulation by miRNAs

The *pBUTR* vector was functionalized with Gateway® technology to allow high-dimensionality screening and validation applications. Given that Gateway® recombineering is scalable—meaning multiple individual 3′ UTR elements can be cloned into the vector in bulk—an inclusive, aggregate set of 3′ UTRs of interest can be rapidly generated to determine whether each or any of these 3′ UTRs confers any of the observed translational regulation in a given physiological context. As a proof of concept, we elected to assess 3′ UTR-mediated responsiveness to TGF-β treatment and EMT in MCF10A cells for 11 distinct genes (*TAPBPL*, *HIST1H4K*, *CCNL1*, *PLK4*, *PDZK1IP1*, *WBP4*, *SYS1*, *RNF6*, *SUV39H2*, *ZEB1* and *ZEB2*). The eleven reporter constructs were built in bulk, sequence verified and individually stably transfected into MCF-10A cells, which were subsequently selected with G418.
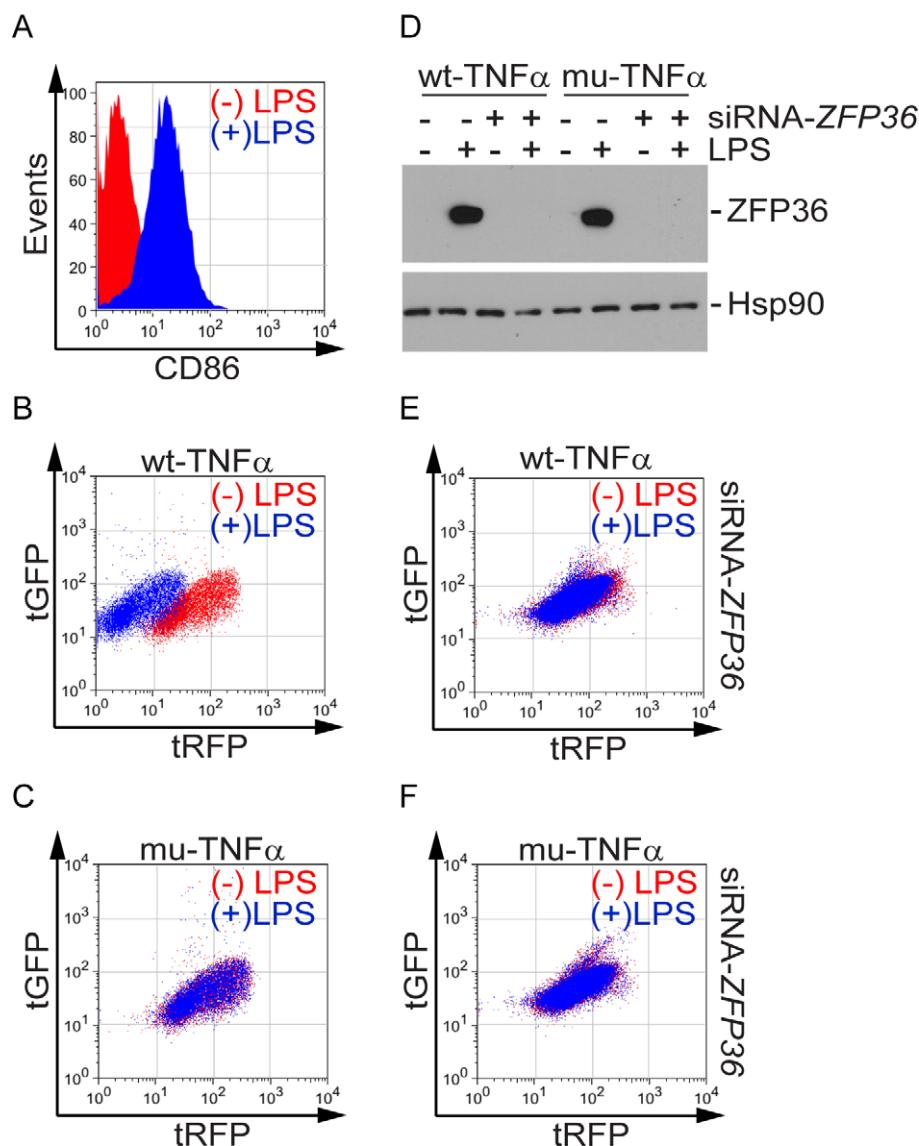
Figure 5. *pBUTR* system can be used to study preferential regulation of mRNA stability by *cis*-elements. (**A**) Flow cytometric analysis of U937 cells transfected with vector recombineered to contain tRFP under the control of the wild-type (wt) TNF-α 3′-UTR for CD86 protein expression (positive activation marker) post-LPS stimulation for 24 h. (**B**) and (**C**) Flow cytometric analysis of untreated (red) or 24-h LPS-treated (blue) U937 cells transfected with vector recombineered to contain tRFP under the control of the wild-type (wt) 3′-UTR (B) or mutant (mu) TNF-α 3′-UTR in which the AU-rich elements are deleted (C). (**D**) Immunoblot (IB) analysis of ZFP36 (upper panel) protein levels in U937 cells harboring wt or mutant TNF-α 3′-UTR reporters transiently transfected with siRNA-*ZFP36* post-LPS stimulation for 24 h. The blot was stripped and re-probed for Hsp90 (bottom panel) as a loading control. (**E**) and (**F**) Flow cytometric analysis of untreated (red) or 24-h LPS-treated (blue) U937 cells harboring wt (E) or mutant (F) TNF-α 3′-UTR reporters transiently transfected with siRNA-*ZFP36* vector. All data are representative of a minimum of three individual experiments. tRFP, turboRFP; UTR, untranslated region; LPS, lipopolysaccharide.

We first assessed 3′ UTR-mediated responsiveness to TGF-β signaling and EMT in an arrayed format. Each stably transfected line was treated with TGF-β or vehicle for 72 h. EMT was monitored both morphologically and by loss of E-cadherin expression via flow cytometry. As expected, cells stably transfected with *ZEB1* and *ZEB2* 3′ UTR reporter constructs were marked by an increase in tRFP expression in TGF-β-treated cells. In contrast, no increase in relative fluorescence was observed in the other nine gene products (Figure 6A).

We next tested our ability to replicate these results in the context of a pooled screening approach. MCF-10A cells

stably transfected with each of the 11 individual reporters were mixed together such that each reporter was equivalently represented within the population. This mixture of cells was divided to two pools, and each of these pools was treated with TGF-β or vehicle, respectively, for 72 h. Once again, EMT was monitored both morphologically and by loss of E-cadherin expression via flow cytometry. Ten percent of each pool of cells was then collected for genomic DNA isolation. The remaining 90% of each population of cells were sorted via flow cytometry to obtain the subpopulation (10%) of cells characterized by the highest tRFP expression. Genomic DNA was isolated from these tRFP[hi]
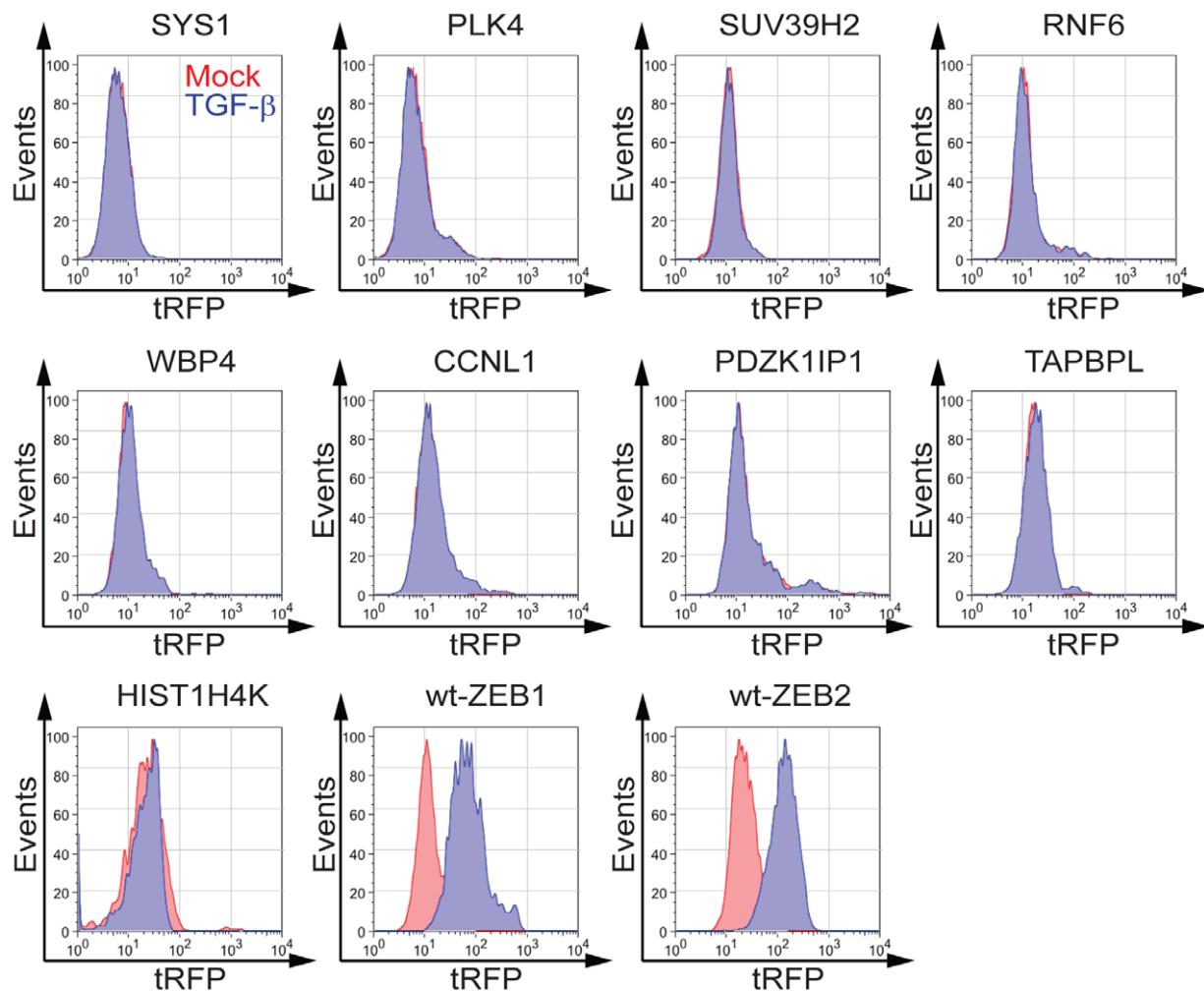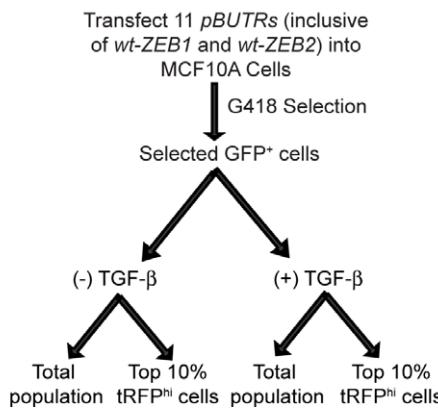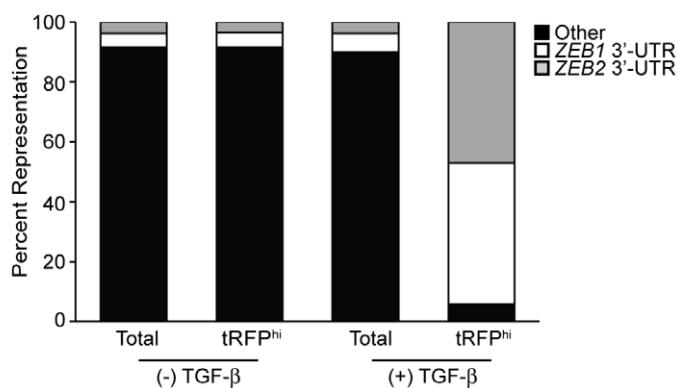
Figure 6. *pBUTR* system is scalable and is amenable to arrayed and pooled screens to study endogenous post-transcriptional gene regulation. (**A**) 3′ UTR-mediated responsiveness to TGF-β signaling and EMT in an arrayed format, Flow cytometric analysis of mock treated (red) and TGF-β-treated (blue) MCF10A cells transfected with vector recombineered to contain tRFP under the control of the indicated 3′-UTRs. (**B**) Schematic representation of the experimental design to test the *pBUTR* system in a pooled screen. (**C**) Comparison of relative *pBUTR* barcode sequence abundance from each of the four populations [untreated, untreated tRFP^{hi}, treated (TGF- β for 72 h) and treated tRFP^{hi} (TGF- β for 72 h) MCF10A cells] via limited Next Generation Sequencing. tRFP, turboRFP; UTR, untranslated region.

populations. We next amplified the pBUTR barcode sequences from each of the four populations (untreated, untreated tRFP[hi], treated and treated tRFP[hi]) and assessed the relative abundance of the barcodes in each of these four populations via limited Next Generation Sequencing (Figure 6B).

Comparison of relative representation of barcodes in the top 10% sorted RFP positive cells in vehicle and TGF-β-treated MCF10A cells revealed an enrichment for *ZEB1* and *ZEB2* 3′ UTRs in the sorted cells post-TGF-β treatment (6.25% *ZEB1* and 3.78% *ZEB2* in vehicle treated compared to 47.06% *ZEB1* and 47.06% *ZEB2* in TGF-β treated) (Figure 6C). This enrichment was not due to a transcriptional upregulation of *ZEB1* and *ZEB2*, as evident by similar barcode abundance of *ZEB1* and *ZEB2* along with the other nine 3′ UTRs in the total populations isolated from the vehicle and TGF-β-treated MCF10A cells (Figure 6C). Taken together, these results showed that the *pBUTR* system is both scalable when performed in an arrayed format and amenable to pooled screening approaches via flow cytometry-based cell sorting and barcode sequencing.

In vivo monitoring of post-transcriptional regulation using the *pBUTR* vector

Cell-line-based models are valuable tools that aid in the understanding of gene regulation events underlying cellular physiology. However, it is often desirable to assess the applicability of the findings derived from these models via transgenesis *in vivo*, whether in the context of mammalian development or disease. Transposon-based systems for transgenesis have some degree of value for such an endeavor, since there is currently little evidence that they are subject to the same silencing mechanisms hampering the use of retro- and lentiviral vectors for this purpose (23–25). We thus undertook a proof-of-concept approach to assess the potential usefulness of the *pBUTR* reporter vector for revealing post-transcriptional regulation in an *in vivo* setting.

The expression patterns of the *ZEB1* gene product during murine embryonic development have been described (48). We stably transfected murine ESCs with our *pBUTR* vector in which tRFP is expressed from the *Ubc* promoter under the control of the wt *ZEB1* 3′ UTR. G418-resistant ESCs were isolated and used for blastocyst injections, and chimeric embryos were harvested at the equivalent of 11.5 days post-coitus (e11.5) for histological analysis of reporter expression. Sagittal sections of the e11.5 embryos revealed relatively even expression of tGFP throughout the embryo (Supplementary Figure S3A). Strikingly, while low levels of tRFP fluorescence were observable throughout the embryo, we observed markedly increased fluorescence in many regions known to be populated by migrating mesenchymal cells derived from the neural crest (Figure 7 and Supplementary Figure S3). Increased tRFP fluorescence was observed in the nasal process, in the maxillary and mandibular arches, in the area of the trigeminal ganglion and in the rhombic lip. Additional increases in fluorescence were observed in the medial and lateral myotome, as well as within non-condensed regions in the distal portion of the limb bud (Figure 7). These areas largely overlap previously described domains of ZEB1 expression at this embryonic stage (48),

which would be consistent with a model in which the *ZEB1* 3′ UTR is sufficient to confer correct temporospatial expression of the tRFP reporter during murine development. While our analysis is not complete enough to draw any firm conclusions in this regard, the data do highlight the potential of the pBUTR reporter for *in vivo* analysis of 3′ UTR-mediated post-transcriptional regulation.

DISCUSSION

We have developed and tested a *piggyBac*-based vector for monitoring 3′ UTR-mediated post-transcriptional regulation of gene expression in *in vitro* and *in vivo* settings. The vector facilitates delivery of three independent transcription units to the target genome—a selection cassette, a constitutively expressed tGFP control reporter and a constitutively expressed experimental reporter, purposed here for monitoring the regulation conferred to a tRFP reporter by distinct 3′ UTRs of interest.

In the context of a stably integrated 3′ UTR reporter system, the *pBUTR* vector has several advantages over competing approaches. The vector is able to deliver a comparatively large payload to the genome of essentially any cell type amenable to transfection, and recent advances utilizing chimeric transposase/ZFN constructs facilitate site-specific targeting of *pB*-based vectors to discrete sites within the genome (49). As compared to retro/lentiviral delivery systems, *pBUTR* virtually eliminates the hazards of exposure to potentially infectious agents derived from packaged vector or *in vitro* recombination with endogenous retroviral elements. Even more importantly from an experimental standpoint, delivery of completely non-overlapping control and experimental transcription units in the context of a lentiviral or retroviral vector poses several problems. Architecture consisting of two reporters separated by 2A peptide sequences, internal ribosome entry sites or even multiple transcriptional promoters will result in shared 3′ UTR identity among the reporters or significant inclusion of irrelevant and potentially confounding UTR sequence, respectively. While use of a bidirectional promoter in the viral vector may significantly reduce these risks, this raises the danger of cryptic splicing or polyadenylation sites in the reverse or antisense transcript (50).

In contrast, the larger relative cargo capacity of the *pBUTR* vector allows simultaneous delivery of multiple fully independent transcription units. Since transposons do not excise and integrate through an RNA intermediate, chimeric splice junctions can be included within each transcription unit, rendering associated mRNA transcripts less susceptible to nonsense mediated decay pathways (51). Because the messages are not fed into these pathways, commonly employed elements promoting mRNA stability and translation [such as the woodchuck hepatitis virus posttranscriptional regulatory element/WPRE (22)] are unnecessary. In fact, for studies of 3′ UTR-mediated post-transcriptional regulation, omission of stability elements is highly desirable since these elements might be expected to confound physiologically relevant regulation by the cellular machinery.

To reduce the possibility that our reporters might confound physiologically relevant regulation conferred by a 3′
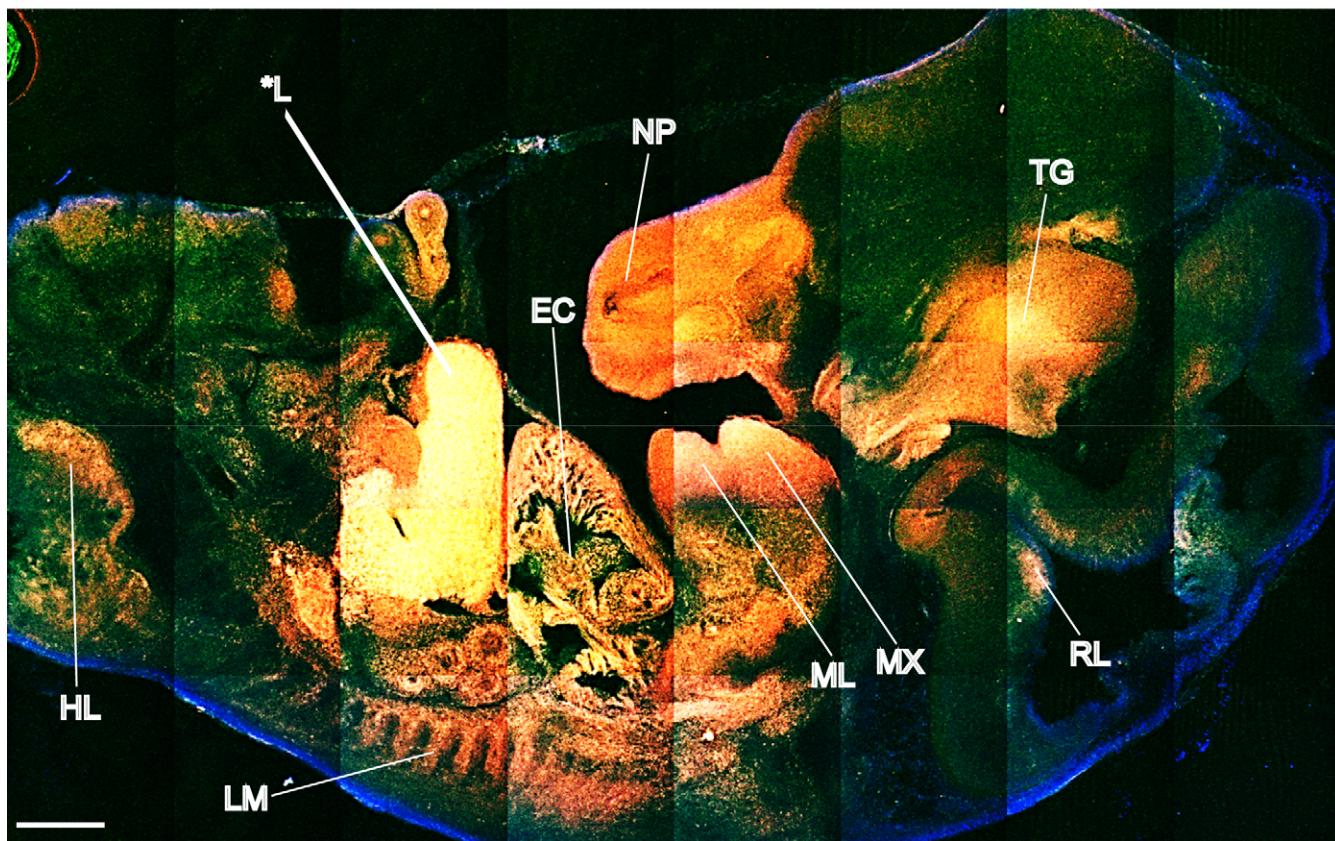
Figure 7. Patterns of expression of a ZEB1 3′ UTR reporter *in vivo*. Composite of confocal immunofluorescence images derived from sagittal section of an e11.5 murine embryo at ×10 magnification. The embryo was derived from ESCs stably transfected with the *pBUTR-ZEB1*wt vector. Scale bar = 10 μm. TG, trigeminal ganglion; RL, rhomboid lip; MX, maxillary arch; ML, mandibular arch; LM, lateral myotome; HL, hind limb bud; EC, endocardial cushion. *L, liver (autofluorescence). Red: tRFP expression. Green: tGFP expression. Blue: 4′,6-diamidino-2-phenylindole (DAPI) staining.

UTR of interest, for example by saturation of the cell's regulatory machinery, we have employed two distinct ubiquitously active transcriptional promoters (*Pgk* and *Ubc*) characterized by a consistently low relative level of expression (52). A potential drawback of this approach is that the two distinct promoters might behave somewhat differently as a function of cellular type or state. However, in an experimental context, this risk can be largely offset by inclusion of appropriate control reporters with minimal or otherwise defined 3′ UTR elements. Since both promoter activity and relative reporter stability will be a function of cellular state, any difference in expression must be due to 3′ UTR-mediated regulation in this case. It is however important to underscore that relative reporter expression within this system does not differentiate between mechanisms impacting mRNA stability or translational repression. This distinction would have to be addressed in downstream experimentation.

The functionalization of the *pBUTR* vector with Gateway® technology was implemented in anticipation of high-dimensionality screening and validation applications. Since Gateway® recombineering is scalable—meaning multiple individual 3′ UTR elements can be cloned into the vector in bulk—an inclusive, aggregate set of 3′ UTRs of interest derived from a polysomal profiling or ribosomal protection experiment can be rapidly generated to determine whether each or any of these 3′ UTRs confers any of the observed translational regulation. Alternatively, comprehensive reporter libraries might be constructed to complement RNA immunoprecipitation/sequencing-type studies via validation of potential regulatory interactions or prospective identification of regulation associated with a particular physiological context. With regard to prospective screening approaches, it should be noted that a drawback of the *pBUTR* system, relative to a retro- or lentiviral approach, is that stable transfection of cells in bulk with a pool of vectors is not straightforward. This would at first appear to argue against the use of the system in pooled screening approaches. However, provided that the initial transfection and selection are performed in an arrayed format, the inclusion of unique barcode elements with the minimal polyadenylation signal allows analysis of enrichment or depletion within pooled cell populations via flow cytometry and cell sorting.

Finally, while the proof-of-concept experiments described in this study make use of *pBUTR* as a tRFP 3′ reporter, the system is also directly adaptable to modeling the impact of single nucleotide polymorphisms or other mutations within a given 3′ UTR on the function of associated genes *in vitro* or *in vivo* model systems. This application may be useful for rapid analysis of candidate phenotypic drivers derived from large-scale genome-wide association or other

large-scale screening efforts aimed at the identification of sequence variants contributing to human disease.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **270**, 467–470.
2. Cloonan,N., Forrest,A.R., Kolle,G., Gardiner,B.B., Faulkner,G.J., Brown,M.K., Taylor,D.F., Steptoe,A.L., Wani,S., Bethel,G. *et al.* (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods*, **5**, 613–619.
3. Cloonan,N. and Grimmond,S.M. (2008) Transcriptome content and dynamics at single-nucleotide resolution. *Genome Biol.*, **9**, 234.
4. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
5. Chen,G., Gharib,T.G., Huang,C.C., Taylor,J.M., Misek,D.E., Kardia,S.L., Giordano,T.J., Iannettoni,M.D., Orringer,M.B., Hanash,S.M. *et al.* (2002) Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteomics*, **1**, 304–313.
6. de Sousa Abreu,R., Penalva,L.O., Marcotte,E.M. and Vogel,C. (2009) Global signatures of protein and mRNA expression levels. *Mol. Biosyst.*, **5**, 1512–1526.
7. Maier,T., Guell,M. and Serrano,L. (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Lett.*, **583**, 3966–3973.
8. Vogel,C., Abreu Rde,S., Ko,D., Le,S.Y., Shapiro,B.A., Burns,S.C., Sandhu,D., Boutz,D.R., Marcotte,E.M. and Penalva,L.O. (2010) Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol. Syst. Biol.*, **6**, 400.
9. Schwanhausser,B., Busse,D., Li,N., Dittmar,G., Schuchhardt,J., Wolf,J., Chen,W. and Selbach,M. (2011) Global quantification of mammalian gene expression control. *Nature*, **473**, 337–342.
10. Jansen,R.P. (2001) mRNA localization: message on the move. *Nat. Rev. Mol. Cell Biol.*, **2**, 247–256.
11. de Moor,C.H., Meijer,H. and Lissenden,S. (2005) Mechanisms of translational control by the 3′ UTR in development and differentiation. *Semin. Cell Dev. Biol.*, **16**, 49–58.
12. Garneau,N.L., Wilusz,J. and Wilusz,C.J. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.*, **8**, 113–126.
13. Keene,J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
14. Merritt,C., Rasoloson,D., Ko,D. and Seydoux,G. (2008) 3′ UTRs are the primary regulators of gene expression in the C. elegans germline. *Curr. Biol.*, **18**, 1476–1482
15. Matoulkova,E., Michalova,E., Vojtesek,B. and Hrstka,R. (2012) The role of the 3′ untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.*, **9**, 563–576.
16. Kornblihtt,A.R., Schor,I.E., Alló,M., Dujardin,G., Petrillo,E. and Muñoz,M.J. (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. *Nat. Rev. Mol. Cell Biol.*, **14**, 153–165.
17. Elkon,R., Ugalde,A.P. and Agami,R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
18. Conne,B., Stutz,A. and Vassalli,J.D. (2000) The 3′ untranslated region of messenger RNA: a molecular 'hotspot' for pathology? *Nat. Med.*, **6**, 637–641.
19. Shibayama,A., Cook,E.H. Jr, Feng,J., Glanzmann,C., Yan,J., Craddock,N., Jones,I.R., Goldman,D., Heston,L.L. and Sommer,S.S. (2004) MECP2 structural and 3′-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.*, **128B**, 50–53.
20. Chatterjee,S. and Pal,J.K. (2009) Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. *Biol. Cell*, **101**, 251–262.
21. Recillas-Targa,F. (2006) Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. *Mol. Biotechnol.*, **34**, 337–354.
22. Zufferey,R., Donello,J.E., Trono,D. and Hope,T.J. (1999) Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.*, **73**, 2886–2892.
23. Jahner,D., Stuhlmann,H., Stewart,C.L., Harbers,K., Lohler,J., Simon,I. and Jaenisch,R. (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature*, **298**, 623–628.
24. Challita,P.M. and Kohn,D.B. (1994) Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 2567–2571.
25. Yoder,J.A., Walsh,C.P. and Bestor,T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
26. Prasad Alur,R.K., Foley,B., Parente,M.K., Tobin,D.K., Heuer,G.G., Avadhani,A.N., Pongubala,J. and Wolfe,J.H. (2002) Modification of multiple transcriptional regulatory elements in a Moloney murine leukemia virus gene transfer vector circumvents silencing in fibroblast grafts and increases levels of expression of the transferred enzyme. *Gene Ther.*, **9**, 1146–1154.
27. Jones,S., Peng,P.D., Yang,S., Hsu,C., Cohen,C.J., Zhao,Y., Abad,J., Zheng,Z., Rosenberg,S.A. and Morgan,R.A. (2009) Lentiviral vector design for optimal T cell receptor gene expression in the transduction of peripheral blood lymphocytes and tumor-infiltrating lymphocytes. *Hum. Gene Ther.*, **20**, 630–640.
28. Cary,L.C., Goebel,M., Corsaro,B.G., Wang,H.G., Rosen,E. and Fraser,M.J. (1989) Transposon mutagenesis of baculoviruses: analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology*, **172**, 156–169.
29. Li,M.A., Turner,D.J., Ning,Z., Yusa,K., Liang,Q., Eckert,S., Rad,L., Fitzgerald,T.W., Craig,N.L. and Bradley,A. (2011) Mobilization of giant piggyBac transposons in the mouse genome. *Nucleic Acids Res.*, **39**, e148.
30. Ding,S., Wu,X., Li,G., Han,M., Zhuang,Y. and Xu,T. (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, **122**, 473–483.
31. Wu,S.C., Meir,Y.J., Coates,C.J., Handler,A.M., Pelczar,P., Moisyadi,S. and Kaminski,J.M. (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 15008–15013.

32. Nakanishi,H., Higuchi,Y., Kawakami,S., Yamashita,F. and Hashida,M. (2010) piggyBac transposon-mediated long-term gene expression in mice. *Mol. Ther.*, **18**, 707–714.

33. Hartley,J.L., Temple,G.F. and Brasch,M.A. (2000) DNA cloning using in vitro site-specific recombination. *Genome Res.*, **10**, 1788–1795.

34. Wilson,M.H., Coates,C.J. and George,A.L. Jr (2007) PiggyBac transposon-mediated gene transfer in human cells. *Mol. Ther.*, **15**, 139–145.

35. Xia,H., Mao,Q. and Davidson,B.L. (2001) The HIV Tat protein transduction domain improves the biodistribution of beta-glucuronidase expressed from recombinant viral vectors. *Nat. Biotechnol.*, **19**, 640–644.

36. Eggan,K., Akutsu,H., Loring,J., Jackson-Grusby,L., Klemm,M., Rideout,W.M. III, Yanagimachi,R. and Jaenisch,R. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 6209–6214.

37. Doench,J.G., Petersen,C.P. and Sharp,P.A. (2003) siRNAs can function as miRNAs. *Genes Dev.*, **17**, 438–442.

38. Yusa,K., Zhou,L., Li,M.A., Bradley,A. and Craig,N.L. (2011) A hyperactive piggyBac transposase for mammalian applications. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1531–1536.

39. Doherty,J.E., Huye,L.E., Yusa,K., Zhou,L., Craig,N.L. and Wilson,M.H. (2012) Hyperactive piggyBac gene transfer in human cells and in vivo. *Hum. Gene Ther.*, **23**, 311–320.

40. Lagos-Quintana,M., Rauhut,R., Lendeckel,W. and Tuschl,T. (2001) Identification of novel genes coding for small expressed RNAs. *Science*, **294**, 853–858.

41. Makeyev,E.V., Zhang,J., Carrasco,M.A. and Maniatis,T. (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell*, **27**, 435–448.

42. Krutzfeldt,J., Rajewsky,N., Braich,R., Rajeev,K.G., Tuschl,T., Manoharan,M. and Stoffel,M. (2005) Silencing of microRNAs in vivo with 'antagomirs'. *Nature*, **438**, 685–689.

43. Gregory,P.A., Bert,A.G., Paterson,E.L., Barry,S.C., Tsykin,A., Farshid,G., Vadas,M.A., Khew-Goodall,Y. and Goodall,G.J. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.*, **10**, 593–601.

44. Brown,K.A., Aakre,M.E., Gorska,A.E., Price,J.O., Eltom,S.E., Pietenpol,J.A. and Moses,H.L. (2004) Induction by transforming growth factor-beta1 of epithelial to mesenchymal transition is a rare event in vitro. *Breast Cancer Res.*, **6**, R215–R231.

45. Maeda,M., Johnson,K.R. and Wheelock,M.J. (2005) Cadherin switching: essential for behavioral but not morphological changes during an epithelium-to-mesenchyme transition. *J. Cell Sci.*, **118**, 873–887.

46. Blick,T., Widodo,E., Hugo,H., Waltham,M., Lenburg,M.E., Neve,R.M. and Thompson,E.W. (2008) Epithelial mesenchymal transition traits in human breast cancer cell lines. *Clin. Exp. Metastasis*, **25**, 629–642.

47. Carballo,E., Lai,W.S. and Blackshear,P.J. (1998) Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science*, **281**, 1001–1005.

48. Takagi,T., Moribe,H., Kondoh,H. and Higashi,Y. (1998) DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. *Development*, **125**, 21–31.

49. Kettlun,C., Galvan,D.L., George,A.L. Jr, Kaja,A. and Wilson,M.H. (2011) Manipulating piggyBac transposon chromosomal integration site selection in human cells. *Mol. Ther.*, **19**, 1636–1644.

50. Liu,B., Paton,J.F. and Kasparov,S. (2008) Viral vectors based on bidirectional cell-specific mammalian promoters and transcriptional amplification strategy for use in vitro and in vivo. *BMC Biotechnol.*, **8**, 49.

51. Nakano,K., Ando,T., Yamagishi,M., Yokoyama,K., Ishida,T., Ohsugi,T., Tanaka,Y., Brighty,D.W. and Watanabe,T. (2013) Viral interference with host mRNA surveillance, the nonsense-mediated mRNA decay (NMD) pathway, through a new function of HTLV-1 Rex: implications for retroviral replication. *Microbes Infect.*, **15**, 491–505.

52. Qin,J.Y., Zhang,L., Clift,K.L., Hulur,I., Xiang,A.P., Ren,B.Z. and Lahn,B.T. (2010) Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS ONE*, **5**, e10611.