A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells

A CRISPR/Cas9 platform for MS2-labelling of single mRNA in live stem cells

Jan-Hendrik Spille[a], Micca Hecht[a], Valentin Grube[a,b], Won-ki Cho[a], Choongman Lee[a], Ibrahim I. Cissé[a]*

Affiliations:

[a] Department of Physics, MIT, Cambridge, MA 02139, USA

[b] Department of Physics, LMU Munich, Geschwister Scholl Platz 1, 80539 Munich, Germany

* to whom correspondence should be addressed: icisse@mit.edu

Declarations of interest: none

Abstract

The MS2 system is a powerful tool for investigating transcription dynamics at the single molecule directly in live cells. In the past, insertion of the RNA-labelling cassette at specific gene loci has been a major hurdle. Here, we present a CRISPR/Cas9-based approach to insert an MS2 cassette with selectable marker at the start of the 3' untranslated region of any coding gene. We demonstrate applicability of our approach by tagging RNA of the stem cell transcription factor Esrrb in mouse embryonic stem cells. Using quantitative fluorescence microscopy we determine the number of nascent transcripts at the Esrrb locus and the fraction of cells expressing the gene. We find that upon differentiation towards epiblast-like cells, expression of Esrrb is down-regulated in an increasing fraction of cells in a binary manner.

Highlights

- Cloning platform for CRISPR/Cas9-based insertion of MS2 cassette in any coding gene

- Insertion of selectable marker and MS2 cassette in 3' UTR

- Instructions for verification of correct insertion

- Quantitative fluorescence microscopy of nascent RNA synthesis

- Esrrb down-regulated in binary (all-or-none) manner in mESC differentiation to epiLC

Keywords

CRISPR, stem cell, MS2, single RNA imaging, quantitative microscopy, live cell imaging

1. Introduction

Imaging of endogenous messenger RNA (mRNA) is a powerful tool to reveal dynamic variation of transcriptional activity directly in living cells [1,2]. Insertion of an array of target sites for the RNA

binding coat protein of phage MS2 (MCP) is the most widely used method to study transcription dynamics in vivo. The MS2 system has been used in model systems ranging from bacteria [3], yeast [4], slime mold [5], and eukaryotic cells [2] to live Drosophila [6–8], zebrafish [9], and mice [10]. Insertion of cassettes encoding 24 or more repeats of the MS2 sequence combined with quantitative fluorescence microscopy enables imaging of single transcripts [11,12], analysis of mRNA trafficking [13], subcellular localization of translation [14–19], as well as observation of nascent transcript synthesis at the transcription site [20,21].

Insertion of an MS2-cassette for tagging of a specific gene in stem cells was successfully demonstrated using a TALEN-based genome editing approach [22]. A CRISPR-based approach [23] could allow more versatile tagging of arbitrary gene loci. Here, we develop a versatile platform for tagging mRNA with a 24xMS2 cassette directly in live cells. We developed a universal cloning strategy employing reusable DNA fragments and highlight an example of labelling endogenous mRNA of the stem cell transcription factor Esrrb in mouse embryonic stem cells (mESCs). Using quantitative fluorescence microscopy we quantified the number of nascent transcripts produced at the *Esrrb* gene locus in living stem cells. We show that down-regulation of Esrrb expression in a population of cells upon differentiation towards epiblast-like cells (EpiLC) results from an increasing fraction of cells that completely shut down expression of the gene.

2. Cell line generation

2.1 Cell culture

R1 mouse embryonic stem cells were a gift from E. Calo (MIT). Cells were cultured in serum-free 2i media as previously described [24,25]. Cell culture flasks and dishes as well as imaging dishes were coated with 5ug/ml PLO (Sigma) in PBS buffer and subsequently 5ug/ml laminin (VWR) in PBS for at least 5h each at 37C. Cell culture media were exchanged every 24h.

2.1 Stable expression of MCP-SNAP

The bacteriophage MS2 stem loop sequence forms an RNA secondary structure that is uniquely recognized by the RNA-binding MS2-coat protein (MCP). We cloned an MCP-NLS-SNAP plasmid co-expressing the neomycin/G418 resistance gene (NeoR) to generate a cell line stably expressing the fluorescently-tagged MCP as previously described [24] (**Fig. 1A**). We transfected cells with the plasmid in a 6-well plate and grew them under G418 (Sigma Aldrich) selection for 5d, while exchanging the medium every day. Surviving cells grow in distinct colonies, and typically 20-100 colonies appear per well.

We then labelled cells by incubation with cell-permeable JF646-SNAP ligand (250nM for 25 min followed by a 10-20 min wash step in 2i medium) and used fluorescence-activated cell sorting (FACS) to select cells with stably inserted MCP-SNAP. We chose three gates to generate cell lines with low, medium, and high expression level of MCP-SNAP and kept working with cells of high expression level. We note that for the observation of single transcripts in the nucleus a lower MCP expression level may be desirable to reduce background from unbound MCP [13].

2.2 24xMS2 repair template cloning

2.2.1 DNA repair pathways used in CRISPR/Cas9-mediated sequence insertion

We chose to insert a cassette encoding 24 repeats of the MS2 stem loop sequence (24xMS2) at the start of the 3' untranslated region (3'-UTR) of the gene of interest. We reasoned that insertion at this position would interfere least with regulatory elements in the 5'- or 3'-UTR of the gene locus. For CRISPR/Cas9-based genome editing, a single guide RNA (sgRNA) is designed to direct the endonuclease Cas9 to a sequence motif determined by a programmable 20nt protospacer sequence in the sgRNA [26]. Upon recognition of an upstream protospacer-adjacent motif (PAM) - NGG in the case of *S. pyogenesis* Cas9 - and the sequence matching the protospacer encoded in the sgRNA, Cas9 induces a double-strand break ~3nt upstream from the PAM-sequence. In order to knock-in an exogenous sequence one can hijack either homologous recombination (HR) or non-homologous end-joining (NHEJ) pathways. Using NHEJ repair for sequence knock-in does not require the design of homology arms flanking the inserted sequence and therefore allows for greater flexibility in principle. [27] We find that without manipulation of further cellular processes inserts are often truncated by degradation of our linear double-stranded DNA (dsDNA) insert. Hence, we focus on an approach leveraging the less error-prone HR pathway by co-delivering a circular plasmid encoding a repair template consisting of two ~800bp homology arms flanking a selectable marker and the 24xMS2 cassette (**Fig. 1B**) [28]. CRISPR insertion frequency depends on a number of factors, including transfection efficiency of the sgRNA/Cas9 and repair template plasmids, sgRNA efficiency in mediating Cas9-cleavage, length of the insert, and length of the homology arms required for HDR.

2.2.2 Design of homology arms with insertion site for RNA-labelling cassette

We find homology arms of 2x800bp to work reliably. Homology arms can be amplified from genomic DNA using primers with appropriate overhangs for subsequent cloning into a donor plasmid or synthesized accordingly in the form of ds DNA. We opted for the latter approach and ordered a synthetic dsDNA gene fragment (Genewiz) encoding a cloning site (2xAarI) for later insertion of the RNA-labelling cassette (**Fig. 1B**), flanked by the left and right homology arms (LHA and RHA) and additional cloning sites for insertion into the universal backbone plasmid. The total length of the dsDNA fragment was ~1.6kb. A minimal mammalian expression vector synthesized by GeneArt Gene Synthesis (Life Technologies) was modified to introduce multiple cloning sites (MCS) into the universal backbone. The locus-specific LHA-2xAarI-RHA gene fragment was cloned into the universal backbone using *MfeI* (NEB) and *MluI* (NEB) to generate a locus-specific LHA-RHA backbone plasmid. The actual 24xMS2 RNA-labelling cassette was inserted in a second step described below.

2.2.3 Cloning an RNA-labelling cassette with selectable marker

We find that insertion frequency ("CRISPR efficiency") is generally low (<1%) for inserts of ~1kb and above. Therefore, a selection strategy for successfully modified cells is needed. For this purpose, we use co-expression of either a fluorescent tag (BFP) or an antibiotic resistance marker (PuroR) fused to the endogenous protein by a self-cleaving 2A peptide (T2A). Either selection marker can serve to enrich transfected cells by FACS or antibiotic selection, respectively, before genotyping monoclonal lines. Here, we cloned a T2A-PuroR-24xMS2 RNA labelling cassette flanked by restriction sites for insertion in the LHA-RHA backbone plasmid.

The ORF encoding a self-cleaving peptide (T2A) upstream of a Puromycin-resistance gene (PuroR) was created through PCR amplification of the T2A-PuroR sequence from pCRISPaint-HaloTag-PuroR (Addgene plasmid # 80960). PCR primers also encode restriction sites for insertion into pDZ415

(24MS2SL loxP-Kan-loxP) (Addgene plasmid # 45162) using *EagI* (NEB) and *BamHI* (NEB). The resulting plasmid was then digested with *EagI* and *SalI* (NEB), adding 5' overhangs to the T2A-PuroR-24xMS2 sequence for ligation into a locus-specific repair template plasmid. The final repair template plasmid (LHA-T2A-puroR-24xMS2-RHA) was generated by digestion-ligation using *AarI* (Thermo Fisher Scientific), a type IIS restriction enzyme [29]. Detailed instructions for the cloning protocol can be found in **Appendix A**.

2.3 CRISPR/Cas9-mediated insertion of the RNA labelling cassette

To mediate insertion of the RNA-labelling cassette (24xMS2 cassette) in the locus of interest cells need to be co-transfected with the repair template plasmid and a plasmid encoding both the sgRNA and the Cas9 enzyme (**Fig. 1C**).

2.3.1 sgRNA design

A multitude of online tools are freely available for the design of sgRNAs (e.g. crispr.mit.edu, benchling.com). Since cells were selected for successful insertion of the RNA-labelling cassette later, we prioritized proximity of the cut site to the desired insertion site over computationally predicted cleavage efficiency of specific sgRNAs. For murine and human cells Schmid-Burgk et al. [27] listed the sgRNA targets leading to cleavage as close as possible upstream from the stop codon of all coding transcripts. We followed the Zhang lab protocol for assembly of sgRNA expressing plasmids [30]. Briefly, we annealed oligonucleotides (**Tab. 1**) encoding the protospacer sequence and ligated them into the px459.v2 (Addgene #62988) backbone. This plasmid expresses both, the sgRNA of interest and the Cas9 protein.

2.3.2 Transfection of sgRNA and repair template and selection of monoclonal cell lines

The modified px459.v2 plasmid and the repair template plasmid were co-delivered into live R1 mESCs [31] by nucleofection using the P3 primary cell kit (Lonza). We used the manufacturer recommended protocol for mouse embryonic stem cells to deliver 1.0 µg of the px459v2 plasmid and 1.5 µg of the repair template plasmid at the same time into 5x10^5 cells. Cells were subsequently grown in a T25 flask. Medium was exchanged every 24h.

2d after transfection 1.5 µg/ml puromycin was added to the culture medium for a total of 3d. Surviving cells typically grow into small colonies (50-100 distinct colonies per T25 flask) and were grown for at least an additional 2d before sorting into a 96-well plate using FACS (1 cell per well). Surviving wells (>60%) were grown for 7-14d before further characterization. Clones with altered morphology or impaired growth were discarded prior to further analysis.

2.3.3 Genotyping

To verify insertion of the RNA-labelling cassette we extracted genomic DNA (GenElute Mammalian Genomic DNA Miniprep Kit, Sigma-Aldrich) and amplified by PCR with out-in, in-out, and out-out primer sets (**Tab. 2**). PCR products were examined by gel electrophoresis (**Fig. 1E**) [24]. The 'out-in' and 'in-out' primer sets confirmed insertion of the cassette. The 'out-out' primer set showed two bands of sizes expected for the unmodified wild-type allele and the successfully edited allele. This indicates heterozygous insertion of the RNA labelling cassette in one of the alleles. We note that heterozygous insertion of the MS2 cassette leaves the possibility for indels to occur in both the MS2-tagged and the

untagged allele. Therefore, PCR products were purified from the gel using the Monarch Gel Purification kit (NEB) and seamless insertion in the tagged allele as well as integrity of the untagged allele was verified by Sanger-sequencing (Genewiz) of all bands. See **Appendix A** for detailed instructions.

Due to the low knock-in efficiency (<1% before antibiotic selection), homozygous insertion may require a revised design of the repair template. Flanking the selection marker (PuroR) with loxP sites would allow for Cre-mediated excision and recycling of the same repair template [32]. Alternatively, an orthogonal selectable marker can be employed to target the second allele.

2.4 Differentiation into epiblast-like cells

We subjected the engineered mESCs to a differentiation protocol which leads to differentiation into epiblast-like cells (EpiLCs) within 24h. *Esrrb* is known to be down-regulated during differentiation from mESCs to EpiLCs [33].

Before EpiLC-differentiation induction, we cultured mESC in 2i-media for 1-2 days. We then trypsinized mESCs and passed the cells through a 70 µm cell strainer (Corning Falcon) to filter out cell aggregates. The single cell suspension was transferred to glass-bottom imaging dishes for microscopy. The glass-bottom dishes were coated with 5 µg/ml Fibronectin for at least 6 hours in a 37C incubator. We allowed mESCs to attache to Fibronectin-coated surfaces in regular 2i-media for 12h. The medium was then exchanged for EpiLC-differentiation buffer. EpiLC-differentiation buffer is composed of 1% knockout serum-replacement (KOSR, Invitrogen) and 12 mg/ml human basic fibroblast growth factor (hbFGF, Reprotech) in a 1:1-mixture of DMEM/F12 and Neurobasal media (Thermo Fisher Scientific).

3. Quantitative fluorescence microscopy of single mRNAs and nascent transcripts

3.1 Dual color single molecule RNA-FISH

To further verify insertion of the MS2 cassette in one allele and assess potential effects on transcription we performed dual-color single molecule fluorescence in situ hybridization (smFISH). We designed two probe sets targeting either an intronic region of the gene of interest or the MS2 cassette, respectively. The first set targeting intron 3 of the Esrrb locus were designed using the BioSearch Technologies tool (mask 5, probe length 20, 48 probes labelled with Quasar670). The latter set consisted of only 2 probes (**Tab. 3**) that each bind a site in the 24-fold repeated spacer sequence, such that up to 48 probes can bind a single transcript. These probes were ordered from IDT as DNA oligonucleotides with a Cy3-dye attached to the 3' end. FISH labelling was performed according to the manufacturer protocol (Stellaris RNA FISH protocol for adherent cells). Cells were fixed in 4% PFA, washed with PBS (3x) and incubated in 70% ethanol for 1h before labelling with both probe sets simultaneously at a final concentration of 125nM for each probe set. We stained DNA with 200ng/ml HOECHST 33342 (ThermoFisher) in Stellaris Wash Buffer A for 20 minutes before the final washing step with Stellaris Wash Buffer B and proceeded to imaging immediately.

3.2 Lattice light sheet microscopy

For dual color smFISH imaging we used a lattice light sheet microscope built in house following the design of the Betzig lab [34]. We acquired multi-color 3D stacks with an effective z-spacing of 314 nm and used LLSpy [35] to process raw data (deskewing and maximum intensity projection). Image were acquired using 405nm excitation for HOECHST 33342-stained DNA, 561nm excitation for Cy3-labelled

MS2 FISH probes, and 642nm excitation for Quasar670-labelled intronic FISH probes. We note that Quasar670 is also excited by the 561nm laser. To avoid premature bleaching and crosstalk we acquired multicolor stacks sequentially, starting with 642nm excitation followed by the 561nm and 405nm channel. The output power of the 642nm laser was set such that virtually all Quasar670 fluorescence was bleached after the first stack and crosstalk during 561nm excitation was not detectable.

3.3 Live cell labelling with cell-permeable SNAP ligand

Cells were plated on PLO/laminin-coated glass bottom dishes (CellVis) for fluorescence imaging and incubated in 2i media for at least 24h. Cell-permeable JF646-SNAP ligand was added to a final concentration of 250nM to the culture medium and incubated for 25min. Cells were briefly washed with culture medium once. Unbound ligand was washed out from cells by further incubation in culture medium for 10-20min. We imaged cells in pre-warmed Leibovitz' L15 medium (ThermoFisher) at 37C.

3.4 Setup for epi-fluorescence microscopy

Live cell imaging was performed on a custom-built epi-fluorescence microscopy setup (**Fig. 2A**) previously described [21]. Briefly, laser illumination was controlled using an AOTF (AA Opto Electronic) and coupled into the backport of a Nikon Eclipse Ti microscope equipped with a 100x oil immersion objective (NA 1.4), Perfect Focus System (Nikon) for controlling the objective z position, and a motorized stage (Prior Scientific). Cells were mounted in a stage top incubator (In Vivo Scientific) to maintain temperature at 37C during image acquisition. Fluorescence was detected with a sensitive EMCCD camera (iXon Ultra 897, Andor). All hardware was controlled using Micro Manager 1.4 [36]. All images were acquired with 100ms exposure EM gain 100 at an approximate 642nm laser power density of 1.0 kW/cm^2 (Vortran Stradus 642, 100mW). For stack acquisition, we relied on the Eclipse Ti objective focus drive and the Micro Manager multi-dimensional acquisition wizard. Stacks were acquired with the same imaging parameters at 300nm z-spacing and spanning 7.2μm, more than the typical thickness of our cells.

3.5 Flat-fielding of epi-fluorescence images

Inhomogeneous illumination is the biggest source of uncertainty in our fluorescence brightness quantification. To correct for any inhomogeneity we applied a flat-fielding strategy (**Fig. 2B**). To this end we first acquired 100 images of 100nM Cy5 solution and averaged the images to obtain a map of the illumination intensity profile. We then acquired 100 frames without laser illumination and averaged the frames to determine the constant camera offset. In order to correct inhomogeneity in the illumination profile we subtracted the average background from the average illumination profile and normalized the resulting image to a mean pixel value of "1". For subsequent analysis the background intensity was subtracted from each image, and the resulting image was divided by the normalized illumination profile.

3.6 Quantifying the brightness of fluorescent peaks

For quantitative imaging we first determined the apparent brightness of single transcripts. In principal, the total fluorescence in a peak can be calculated from the amplitude and width of a 2D Gaussian peak fitted to the signal. However, we find it more robust to calculate the total integrated intensity in a defined area around the intensity maximum. This is particularly true for transcription sites that are sometimes not well-described by a simple Gaussian function due to their extended size close to or even beyond the diffraction-limit. We calculated the integrated intensity of a peak above background

according to $I = I_1 - (I_2 - I_1) \times (\frac{A_1}{A_2 - A_1})$ (**Fig. 2C**) [37]. Here, I_1 denotes the integrated intensity in a small region of interest (ROI) with area A_1, whereas I_2 and A_2denote the same values for a slightly larger ROI. The second term in the formula calculates the integrated background signal in the area between A_2 and A_1, $(I_2 - I_1)$, and weighs it by the relative area fraction $(\frac{A_1}{A_2 - A_1})$ before subtracting from I_1. We chose the size of A_1 such that it fully contained the peak (7x7 pixels), and the size of A_2 such that background was sampled from a sufficient number of pixels around the peak (9x9 pixels). This formula calculates the local background intensity per pixel as the average value of the pixels included in the larger, but not the smaller ROI.

3.7 Counting nascent transcripts at the transcription site

In a similar manner we evaluated the brightness of transcription sites identified as distinct bright peaks of low mobility in the cell nucleus (**Fig. 3A**). The background level of unbound nuclear MCP-SNAP and mature Esrrb mRNA particles in the nucleus determines how many nascent transcripts need to be present at the transcription site to detect it above background. Typically, already a few transcripts (2-4) were sufficient to accumulate enough signal for detection. Depending on the exact insertion site of the MS2 cassette and the structure of the gene locus, transcription sites can appear elongated rather than as well-defined, diffraction-limited spots. We determined the brightness of transcription sites in the same manner as that of single transcripts. The values obtained for transcription sites were then divided by the single transcript brightness to obtain an estimate for the number of nascent transcripts present at the gene locus.

4. Results

4.1 Dual color single molecule FISH

We first performed fixed cell dual color single molecule FISH experiments on mouse embryonic stem cells to detect both, an intronic region of the Esrrb transcript highlighting the position of the gene locus, and the MS2 cassette (**Fig. 3A**). As expected we detect two bright foci in the nucleus of most cells in the intronic FISH channel, corresponding to the two alleles from which the gene is transcribed. One of the transcription sites was usually also clearly identified in the MS2 FISH channel, confirming heterozygous insertion of the MS2 cassette in one of the alleles.

To assess whether insertion of the MS2 cassette impaired transcription from the tagged allele we compared the brightness of the intronic signal for tagged and untagged alleles in cells where both alleles could clearly be identified. Since most splicing happens co-transcriptionally or at least at the transcription site [38] we did not detect single transcripts in the intronic FISH channel and could therefore not assess the number of transcripts at the transcription site in this channel. We therefore quantified the absolute brightness of MS2-tagged and untagged Esrrb alleles across 27 cells and found no significant difference between the alleles (**Fig. 3B**).

In the MS2 FISH channel, we detected not only the tagged transcription site but also single transcripts of lesser brightness throughout the nucleus and the cytoplasm (**Fig. 3A**). We found a narrow distribution of single transcript brightness well described by a normal distribution (**Fig. 3C**). Fitting the distribution allowed us to determine the typical brightness of a single transcript. This allowed us to normalize the brightness of transcription sites detected in the MS2 FISH channel (**Fig. 3D**). We found the brightness of

transcription sites to correspond to an average of 5.5 +- 0.6 (mean +- s.e.m.) nascent transcripts. We note that some of the dimmest transcription sites in the MS2 FISH channels corresponded to only 1-2 transcripts and could only be distinguished from single mature transcripts because they colocalized with transcription sites identified by the intronic FISH signal.

Taken together, these results suggest that Esrrb is constitutively expressed from both alleles in mESC, and that insertion of the MS2 cassette does not significantly impair transcription from the tagged allele.

4.2 Live cell imaging of nascent transcripts during differentiation of mESC towards epiLC

We next proceeded to live cell imaging of transcriptional activity at the Esrrb gene locus. To this end, we labelled cells with the cell-permeable JF646-SNAP ligand [39]. Since the MCP-SNAP used here also carries a nuclear localization sequence (NLS), as expected, cell nuclei showed bright fluorescent signal. Transcription sites were visible as bright foci in the cell nucleus (**Fig. 4A**). Single transcripts were readily detectable in the cytoplasm (**Fig. 4A, inset**) and sometimes above the background of unbound MCP-SNAP in the nucleus. We note that as previously found for fixed cell FISH imaging, some transcription sites were not much brighter than the signal from single transcripts. Acquisition of short image sequences allowed us to distinguish slowly moving dim transcription sites from more rapidly diffusing single transcripts. We note that at the exposure time of 100ms the signal from single transcripts is usually blurrd out due to their diffusion in the nucleus, and single transcripts become visible as distinct peaks only occasionally and for a few frames during transient phases of low mobility. Transcription sites on the other hand are continuously visible and exhibit low mobility (**Movie S1**).

We imaged 207 cells looking at a single 2D plane and identified 89 transcription sites. Normalized to the typical brightness determined from 408 single transcripts (**Fig. 4B**)., we found a broad distribution of up to 17 copies of Esrrb mRNA with a mean normalized brightness of 5.9 +- 0.3 (mean +- s.e.m.) nascent transcripts per locus (**Fig. 4C**). This value is in good agreement with the value determined from fixed cell MS2 FISH experiments. We note that due to the higher background of unbound MCP-SNAP in the cell nucleus and the missing intronic FISH reference signal, the live cell distribution may be slightly skewed towards loci with brighter signal. However, even for live cells we identified many loci with normalized brightness of only 2-4 transcripts (35% of all transcription sites).

We note further that this 2D imaging approach captures only a fraction of each cell nucleus. Therefore we also acquired z-stacks to capture the whole cellular volume and identified a transcription site in 67 +- 4% of the cells (average +- range of three biological replicates with n > 100 cells each) (**Fig. 4D**).

Esrrb is a key transcription factor highly expressed in mouse embryonic stem cells, but down-regulated during differentiation towards epiblast-like cells (epiLC) as previously revealed by qRT-PCR [24,33]. Differentiation can be induced by replacing regular 2i growth medium with appropriate differentiation buffer and occurs within 24h [33]. We applied the differentiation protocol to our cells and analyzed the brightness of Esrrb loci at different time points after inducing differentiation. We found that the number of nascent mRNA transcripts per gene locus did not change significantly after 12h (5.8 +- 0.6 transcripts, n = 30), and was only slightly reduced after 24h (4.7 +- 0.6 transcripts, n = 58) (**Fig. 4C**).

In contrast, the fraction of cells for which a transcription site could be identified in z-stacks gradually decreased to 53 +- 4% after 12h of differentiation, 30 +- 2% after 24h, and 8 +- 3% after 36h (average +- range of three biological replicates with n > 100 cells each) (**Fig. 4D**).

These result reveal that even as across the population of cells Esrrb expression is gradually turned off, a minority of individual cells may still be actively transcribing the gene locus, likely in a binary manner at an instantaneous load of ~5-6 transcripts per locus in the 'on' state and no transcription in the 'off' state.

5. Discussion

Quantitative imaging of mRNA with single molecule sensitivity directly in single living cells can yield key insights that may be hidden in population studies. The technical hurdle had historically been the insertion in the gene locus of interest of RNA labelling cassettes like MS2 that enable single molecule *in vivo* characterization. Here, we present a versatile platform for CRISPR/Cas9-based insertion of a widely used 24xMS2 cassette and use it to study the transcriptional activity of a key gene (Esrrb) in live mouse embryonic stem cells.

We use sensitive dual-color single molecule FISH to characterize transcriptional activity at the Esrrb locus and show that we can obtain the same results in live cells using MS2-labelling.

Our quantitative analysis of Esrrb transcription during differentiation from mESCs to EpiLC revealed that the gene is down-regulated in a binary manner. The data we present demonstrates that a minority of individual cells in the population may still be transcribing Esrrb with the same intensity even as the majority of cells have already shut down transcription of the gene.

Our approach includes co-expression of an antibiotic resistance marker for easy selection of successfully modified cells and is based on insertion of the RNA-labelling cassette immediately after the stop codon. This approach should be readily adaptable for insertion of other cassettes. In particular the approach would be compatible with new generations of MS2 cassettes [40,41] employing scrambled, non-repetitive spacer and stem loop sequences to maintain endogenous mRNA degradation rates.

Acknowledgements

We thank E. Calo (MIT) for the wild-type R1 cells, and differentiation protocol and L. D. Lavis (HHMI/Janelia) and J. Grimm (HHMI/Janelia) for gift of the JF646-SNAP ligand. pCRISPaint-HaloTag-PuroR was a gift from Veit Hornung (Addgene plasmid # 80960). pDZ415 (24MS2SL loxP-Kan-loxP) was a gift from Robert Singer & Daniel Zenklusen (Addgene plasmid # 45162). pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988).

Funding: This work was supported by the NIH Director's New Innovator award #DP2CA195769 (to I.I.C). J.-H.S. is supported by a postdoctoral fellowship from the German Research Foundation (DFG, SP1680/1-1).

References

[1] A. Raj, A. van Oudenaarden, Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences, Cell. 135 (2008) 216–226. doi:10.1016/j.cell.2008.09.050.

[2] Y. Shav-Tal, X. Darzacq, S.M. Shenoy, D. Fusco, S.M. Janicki, D.L. Spector, R.H. Singer, Dynamics of Single mRNPs in Nuclei of Living Cells, Science. 304 (2004) 1797–1800. doi:10.1126/science.1099754.

[3] I. Golding, J. Paulsson, S.M. Zawilski, E.C. Cox, Real-Time Kinetics of Gene Activity in Individual Bacteria, Cell. 123 (2005) 1025–1036. doi:10.1016/j.cell.2005.09.031.

[4] E. Bertrand, P. Chartrand, M. Schaefer, S.M. Shenoy, R.H. Singer, R.M. Long, Localization of ASH1 mRNA Particles in Living Yeast, Mol. Cell. 2 (1998) 437–445. doi:10.1016/S1097-2765(00)80143-4.

[5] J.R. Chubb, T. Trcek, S.M. Shenoy, R.H. Singer, Transcriptional Pulsing of a Developmental Gene, Curr. Biol. 16 (2006) 1018–1025. doi:10.1016/j.cub.2006.03.092.

[6] K.M. Forrest, E.R. Gavis, Live Imaging of Endogenous RNA Reveals a Diffusion and Entrapment Mechanism for nanos mRNA Localization in Drosophila, Curr. Biol. 13 (2003) 1159–1168. doi:10.1016/S0960-9822(03)00451-2.

[7] H.G. Garcia, M. Tikhonov, A. Lin, T. Gregor, Quantitative Imaging of Transcription in Living Drosophila Embryos Links Polymerase Activity to Patterning, Curr. Biol. 23 (2013) 2140–2145. doi:10.1016/j.cub.2013.08.054.

[8] T. Lucas, T. Ferraro, B. Roelens, J. De Las Heras Chanes, A.M. Walczak, M. Coppey, N. Dostatni, Live Imaging of Bicoid-Dependent Transcription in Drosophila Embryos, Curr. Biol. 23 (2013) 2135–2139. doi:10.1016/j.cub.2013.08.053.

[9] P.D. Campbell, J.A. Chao, R.H. Singer, F.L. Marlow, Dynamic visualization of transcription and RNA subcellular localization in zebrafish, Development. 142 (2015) 1368–1374. doi:10.1242/dev.118968.

[10] T. Lionnet, K. Czaplinski, X. Darzacq, Y. Shav-Tal, A.L. Wells, J.A. Chao, H.Y. Park, V. de Turris, M. Lopez-Jones, R.H. Singer, A transgenic mouse for *in vivo* detection of endogenous labeled mRNA, Nat. Methods. 8 (2011) 165–170. doi:10.1038/nmeth.1551.

[11] D. Fusco, E. Bertrand, R.H. Singer, Imaging of Single mRNAs in the Cytoplasm of Living Cells, in: RNA Traffick. Nucl. Struct. Dyn., Springer, Berlin, Heidelberg, 2008: pp. 135–150. doi:10.1007/978-3-540-74266-1_7.

[12] J.-H. Spille, U. Kubitscheck, Labelling and imaging of single endogenous messenger RNA particles in vivo, J Cell Sci. 128 (2015) 3695–3706. doi:10.1242/jcs.166728.

[13] D. Grünwald, R.H. Singer, In Vivo Imaging of Labelled Endogenous β-actin mRNA During Nucleocytoplasmic Transport, Nature. 467 (2010) 604–607. doi:10.1038/nature09438.

[14] J.M. Halstead, T. Lionnet, J.H. Wilbertz, F. Wippich, A. Ephrussi, R.H. Singer, J.A. Chao, An RNA biosensor for imaging the first round of translation from single cells to living animals, Science. 347 (2015) 1367–1671. doi:10.1126/science.aaa3380.

[15] B. Wu, C. Eliscovich, Y.J. Yoon, R.H. Singer, Translation dynamics of single mRNAs in live cells and neurons, Science. 352 (2016) 1430–1435. doi:10.1126/science.aaf1084.

[16] C. Wang, B. Han, R. Zhou, X. Zhuang, Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells, Cell. 165 (2016) 990–1001. doi:10.1016/j.cell.2016.04.040.

[17] X. Yan, T.A. Hoek, R.D. Vale, M.E. Tanenbaum, Dynamics of Translation of Single mRNA Molecules In Vivo, Cell. 165 (2016) 976–989. doi:10.1016/j.cell.2016.04.034.

[18] X. Pichon, A. Bastide, A. Safieddine, R. Chouaib, A. Samacoits, E. Basyuk, M. Peter, F. Mueller, E. Bertrand, Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells, J Cell Biol. 214 (2016) 769–781. doi:10.1083/jcb.201605024.

[19] T. Morisaki, K. Lyon, K.F. DeLuca, J.G. DeLuca, B.P. English, Z. Zhang, L.D. Lavis, J.B. Grimm, S. Viswanathan, L.L. Looger, T. Lionnet, T.J. Stasevich, Real-time quantification of single RNA translation dynamics in living cells, Science. 352 (2016) 1425–1429. doi:10.1126/science.aaf0899.
[20] J.P. Bothma, H.G. Garcia, S. Ng, M.W. Perry, T. Gregor, M. Levine, Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo, ELife. 4 (2015) e07956. doi:10.7554/eLife.07956.
[21] W.-K. Cho, N. Jayanth, B.P. English, T. Inoue, J.O. Andrews, W. Conway, J.B. Grimm, J.-H. Spille, L.D. Lavis, T. Lionnet, I.I. Cisse, RNA Polymerase II cluster dynamics predict mRNA output in living cells, ELife. 5 (2016). doi:10.7554/eLife.13617.
[22] H. Ochiai, T. Sugawara, T. Sakuma, T. Yamamoto, Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells, Sci. Rep. 4 (2014) 7125. doi:10.1038/srep07125.
[23] F.X. Chen, E.R. Smith, A. Shilatifard, Born to run: control of transcription elongation by RNA polymerase II, Nat. Rev. Mol. Cell Biol. 19 (2018) 464–478. doi:10.1038/s41580-018-0010-5.
[24] W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I.I. Cisse, Mediator and RNA polymerase II clusters associate in transcription-dependent condensates, Science. (2018) eaar4199. doi:10.1126/science.aar4199.
[25] J. Silva, O. Barrandon, J. Nichols, J. Kawaguchi, T.W. Theunissen, A. Smith, Promotion of Reprogramming to Ground State Pluripotency by Signal Inhibition, PLOS Biol. 6 (2008) e253. doi:10.1371/journal.pbio.0060253.
[26] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity, Science. 337 (2012) 816–821. doi:10.1126/science.1225829.
[27] J.L. Schmid-Burgk, K. Höning, T.S. Ebert, V. Hornung, CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism, Nat. Commun. 7 (2016). doi:10.1038/ncomms12338.
[28] H. Yang, H. Wang, C.S. Shivalila, A.W. Cheng, L. Shi, R. Jaenisch, One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering, Cell. 154 (2013) 1370–1379. doi:10.1016/j.cell.2013.08.022.
[29] C. Engler, R. Kandzia, S. Marillonnet, A One Pot, One Step, Precision Cloning Method with High Throughput Capability, PLOS ONE. 3 (2008) e3647. doi:10.1371/journal.pone.0003647.
[30] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, Nat. Protoc. 8 (2013) 2281–2308. doi:10.1038/nprot.2013.143.
[31] A. Nagy, J. Rossant, R. Nagy, W. Abramow-Newerly, J.C. Roder, Derivation of completely cell culture-derived mice from early-passage embryonic stem cells, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 8424–8428.
[32] T.-L. Cheng, Z. Qiu, Long non-coding RNA tagging and expression manipulation via CRISPR/Cas9-mediated targeted insertion, Protein Cell. 9 (2018) 820–825. doi:10.1007/s13238-017-0464-9.
[33] C. Buecker, R. Srinivasan, Z. Wu, E. Calo, D. Acampora, T. Faial, A. Simeone, M. Tan, T. Swigut, J. Wysocka, Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency, Cell Stem Cell. 14 (2014) 838–853. doi:10.1016/j.stem.2014.04.003.
[34] B.-C. Chen, W.R. Legant, K. Wang, L. Shao, D.E. Milkie, M.W. Davidson, C. Janetopoulos, X.S. Wu, J.A. Hammer, Z. Liu, B.P. English, Y. Mimori-Kiyosue, D.P. Romero, A.T. Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R.D. Mullins, D.M. Mitchell, J.N. Bembenek, A.-C. Reymann, R. Böhme, S.W. Grill, J.T. Wang, G. Seydoux, U.S. Tulu, D.P. Kiehart, E. Betzig, Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution, Science. 346 (2014) 1257998. doi:10.1126/science.1257998.
[35] Talley Lambert, Lin Shao, tlambert03/LLSpy: v0.3.8, Zenodo, 2018. doi:10.5281/zenodo.1204615.

[36] A.D. Edelstein, M.A. Tsuchida, N. Amodaj, H. Pinkard, R.D. Vale, N. Stuurman, Advanced methods of microscope control using μManager software, J. Biol. Methods. 1 (2014) e10. doi:10.14440/jbm.2014.36.
[37] J.S. Verdaasdonk, J. Lawrimore, K. Bloom, Determining absolute protein numbers by quantitative fluorescence microscopy, Methods Cell Biol. 123 (2014) 347–365. doi:10.1016/B978-0-12-420138-5.00019-7.
[38] Y. Brody, N. Neufeld, N. Bieberstein, S.Z. Causse, E.-M. Böhnlein, K.M. Neugebauer, X. Darzacq, Y. Shav-Tal, The In Vivo Kinetics of RNA Polymerase II Elongation during Co-Transcriptional Splicing, PLOS Biol. 9 (2011) e1000573. doi:10.1371/journal.pbio.1000573.
[39] J.B. Grimm, B.P. English, J. Chen, J.P. Slaughter, Z. Zhang, A. Revyakin, R. Patel, J.J. Macklin, D. Normanno, R.H. Singer, T. Lionnet, L.D. Lavis, A general method to improve fluorophores for live-cell and single-molecule microscopy, Nat. Methods. 12 (2015) 244–250. doi:10.1038/nmeth.3256.
[40] K. Tantale, F. Mueller, A. Kozulic-Pirher, A. Lesne, J.-M. Victor, M.-C. Robert, S. Capozi, R. Chouaib, V. Bäcker, J. Mateos-Langerak, X. Darzacq, C. Zimmer, E. Basyuk, E. Bertrand, A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting, Nat. Commun. 7 (2016) 12248. doi:10.1038/ncomms12248.
[41] E. Tutucci, M. Vera, J. Biswas, J. Garcia, R. Parker, R.H. Singer, An improved MS2 system for accurate reporting of the mRNA life cycle, Nat. Methods. 15 (2018) 81–89. doi:10.1038/nmeth.4502.

Figures

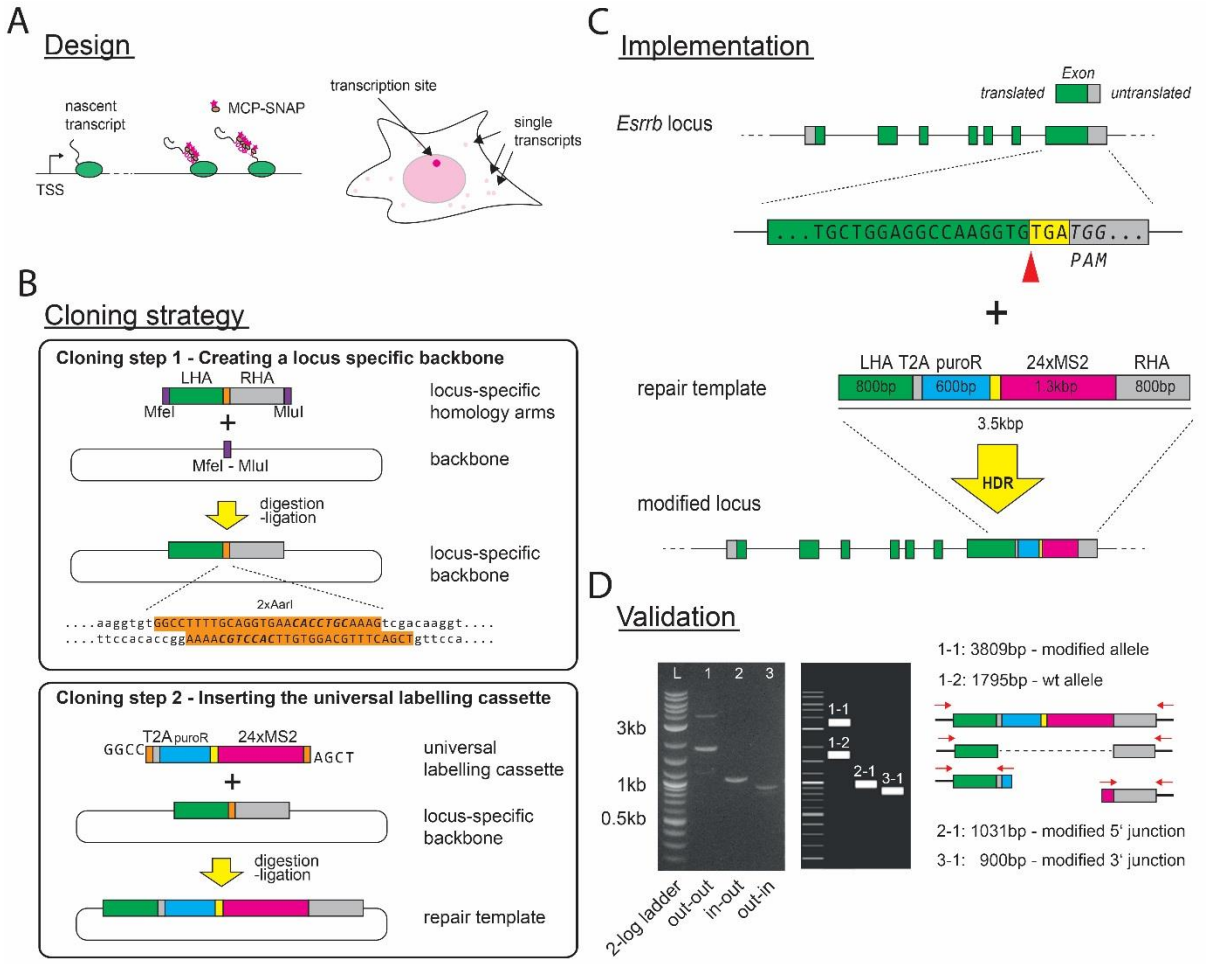


Fig. 1: Cloning and insertion of a 24xMS2 RNA labelling cassette using CRISPR/Cas9 genome engineering

A We stably expressed MCP-SNAP in R1 stem cells. MCP-SNAP decorates the MS2 stem loops in mRNA transcripts and renders single transcripts visible. Accumulation of nascent transcripts lets the transcription site appear as a bright spot in the nucleus.

B Two cloning steps are required to generate a locus-specific repair template plasmid from universal building blocks. First, locus specific dsDNA encoding left and right homology arms (LHA, RHA) and an internal cloning site is inserted into a backbone plasmid. Second, the RNA labelling cassette carrying an antibiotic selectable marker is inserted between the homology arms.

C To insert the MS2 cassette in an endogenous gene locus we co-transfected plasmids encoding the repair template as well as the Cas9 endonuclease and a sgRNA cutting (red arrowhead) immediately upstream from the stop codon (yellow) of our gene of interest.

D Successful insertion was verified by gel electrophoresis of PCR-amplified genomic DNA. Amplicons span either the entire insertion locus (lane 1, out-out primers binding outside the homology arms) or sequences across the 5’ (lane 2, in-out) and 3’ (lane 3, out-in) junction of the insertion with one primer binding outside the homology arms in the endogenous sequence and the other primer binding an exogenous sequence inside the insert.

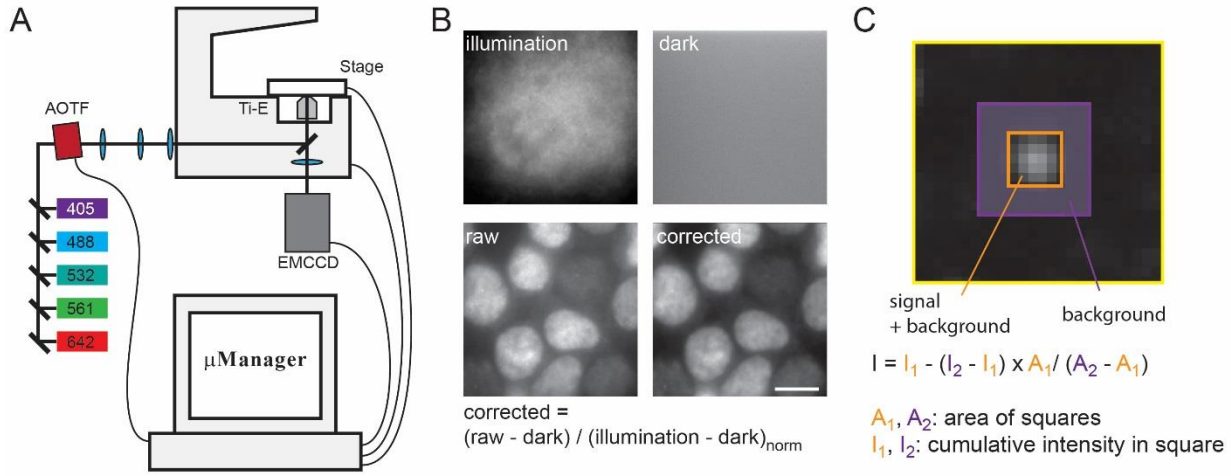


Fig. 2: Quantitative epi-fluorescence microscopy

A Sketch of the fluorescence microscopy setup used for live cell imaging. Laser lines were combined using appropriate dichroic mirrors and passed through an AOTF. The collinear beams were expanded by an achromatic telescope and coupled into the backport of a Nikon Ti-E microscope body. Cells were mounted on a motorized stage. Fluorescence was collected with a high numerical aperture objective and imaged onto an EMCCD camera. Image acquisition and all hardware were controlled using µManager.

B For quantitative image analysis, raw data was corrected for inhomogeneity in the illumination profile by flat-fielding. We acquired an image of the illumination profile using a test sample. We corrected all raw data by subtracting a dark field image and subsequently divided by a normalized form of the illumination profile. Scale bar 10 µm.

C To quantify the brightness of fluorescent peaks we determined their integrated intensity and subtracted local background. For this purpose we first calculated the cumulative intensity I1 of all pixels within a square region of interest of area A1 (orange). We then performed the same operation for a slightly larger region of interest (purple) to determine I2 and A2, accordingly. The local background intensity was calculated from the pixels between the smaller and larger region of interest and subtracted from I1 according to the formula stated in panel C.

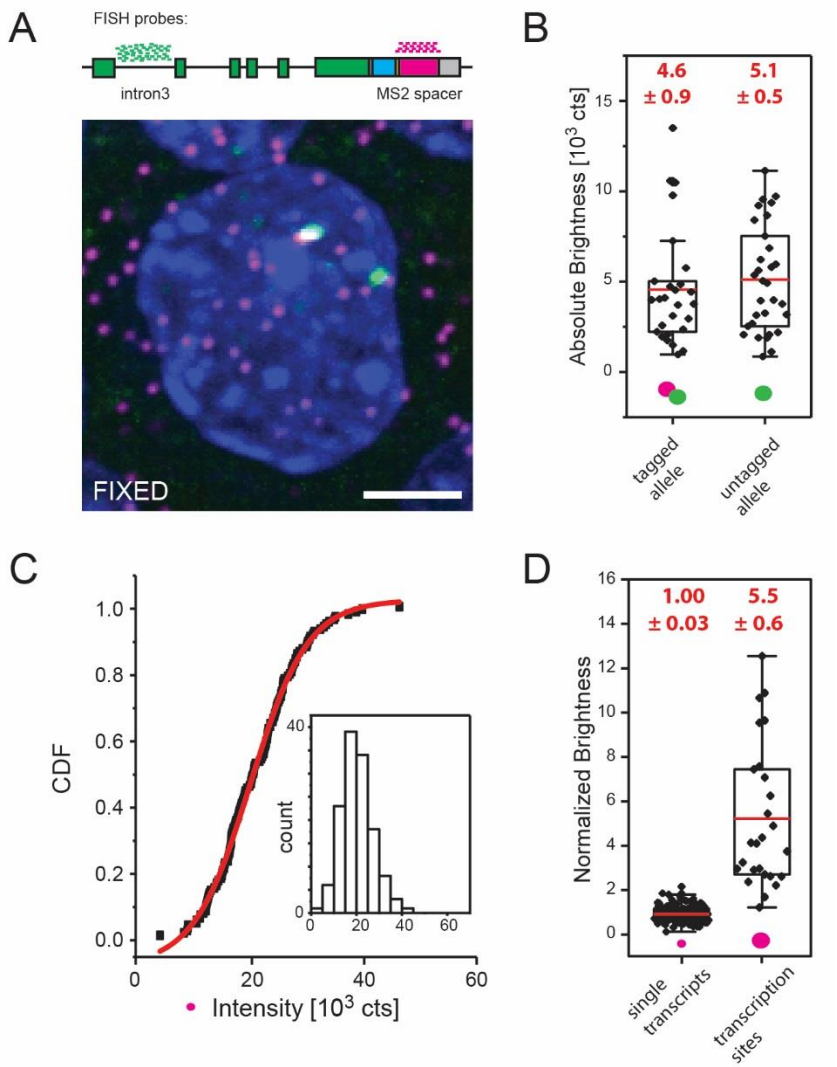


Fig. 3: Verification of MS2 insertion by dual-color smFISH and fixed cell quantification of transcripts at the Esrrb locus in fixed stem cells.

A We performed dual-color smFISH on a lattice light sheet microscope to further verify heterozygous insertion of the MS2 cassette in one allele of the Esrrb locus. We used two sets of RNA FISH probes targeting either an intron (green) or the MS2 spacer sequence (magenta) to label Esrrb transcripts. Intronic probes appear as two bright foci highlighting both alleles. Probes against the MS2 spacer sequence show only one bright locus in the nucleus that colocalizes with one of the intronic loci. Single transcripts of lesser but uniform brightness are visible in the nucleus (blue, HOECHST 33342 DNA staining) and the cytoplasm. These results confirm successful tagging of Esrrb mRNA with the MS2 cassette and productive synthesis of transcripts from the tagged allele. Maximum intensity projection of image stack. Scale bar 5 µm.

B Absolute brightness of the intronic FISH signal of tagged (colocalizing with >1transcript in MS2 FISH channel) and untagged alleles in cells with both alleles clearly identified (n = 27). There was no

significant difference between the absolute brightness of the tagged and untagged allele (red: mean +- s.e.m).

C Cumulative distribution function (CDF) of single transcript brightness in fixed cells. The integrated intensity of single transcripts (n = 133 randomly picked from 10 cells) detected in the MS2-FISH channel is well described by a unimodal normal distribution. The mean brightness of single transcripts was determined from fitting to the CDF (red). Inset: Histogrammed data.

D Normalized brightness of single transcripts and transcription sites in the MS2-FISH channel obtained by dividing the absolute brightness by the mean brightness of a single transcript. Transcription sites harbor on average 5.5 +- 0.6 transcripts (mean +- s.e.m., n = 27).

All data points shown. Boxes indicate 25%-75% quartiles. Whiskers mark 1.5x interquartile range. Red line indicates mean value.

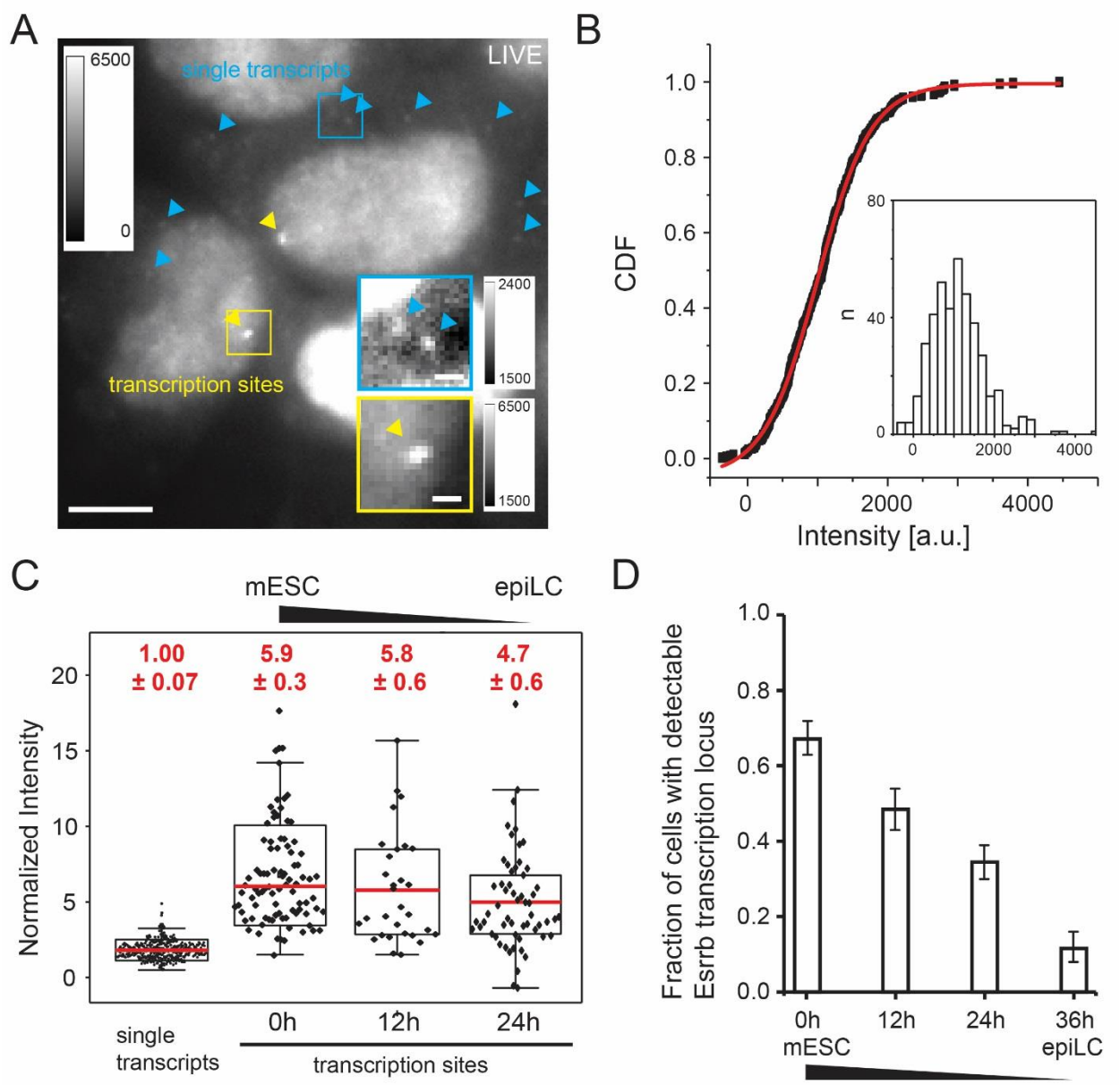


Fig. 4: Live cell epi-fluorescence imaging of nascent transcripts at the Esrrb locus during differentiation towards epiblast like cells.

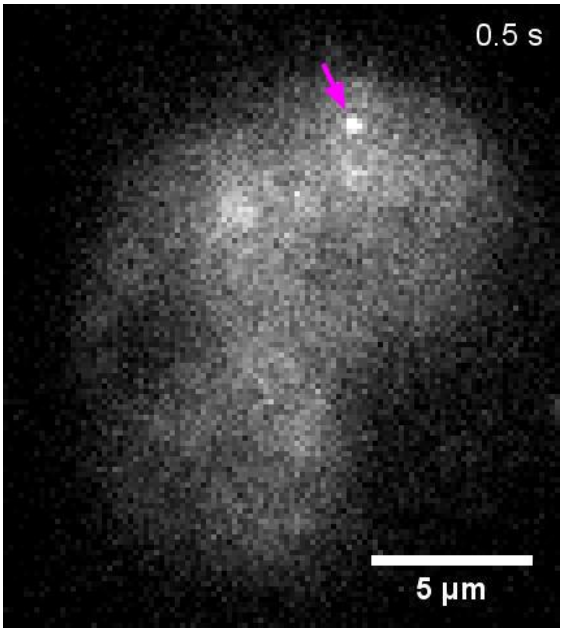
A Live imaging of Esrrb transcripts after JF646-SNAP labelling. Nuclei are clearly identified by MCP-SNAP accumulation. Transcription sites stand out as bright foci in the nucleus (yellow arrowheads). Single transcripts are readily identified in the cytoplasm (blue arrowheads). Insets: Contrast individually adjusted for better visibility of single transcripts and transcription sites. Color bars indicate contrast settings. Scale bar 10 μm (insets 1 μm).

B Cumulative distribution function of single transcript brightness in live cells. The integrated intensity of single transcripts (n = 408) is well described by a unimodal normal distribution and we determined the mean brightness by fitting to the CDF (red). Inset: Histogrammed data.

C Normalized brightness for single transcripts as well as transcription sites at different timepoints during differentiation from mESC towards epiLC. In the undifferentiated mESC state (0h) we find an average of 5.9 +- 0.3 transcripts at the Esrrb locus (n = 89). Upon differentiation towards epiLC this value does not drop significantly after 12h (5.8 +- 0.6, n = 30) or 24h (4.7 +- 0.6, n = 58). After 36h the transcription site was not detectable in >90% of cells.

All data points shown. Boxes indicate 25%-75% quartiles. Whiskers mark 1.5x interquartile range. Red line indicates mean value.

D The fraction of cells with a transcription site visible. Performing volumetric imaging on the epi-fluorescence microscope we found a transcription site in 67% +- 4% in mESC state. This value dropped continuously to 53 +- 4% after 12h in differentiation medium, 30 +- 2% after 24h, and 12% +- 3% after 36h. Mean values from three biological replicates with n > 100 cells each. Error bars indicate minimum and maximum values.



Movie S1: Time lapse epi-fluorescence imaging of a single cell with MS2-labelled transcription site (arrow). 100ms exposure time, one image acquired every 500ms.

Tab. 1: Oligo sequences for cloning sgRNAs

ESRRB_sgRNA	(forward) CACCGTGCTGGAGGCCAAGGTGTGA
	(reverse) AAACTCACACCTTGGCCTCCAGCA

Tab. 2: PCR primers for Esrrb-T2A-PuroR sequence validation

Label	Full Primer Name	5'-3' Sequence	WT	modified
'out-out'	mmEsrrb j5 F	CTGTAAGCATCCCAAGCCGA	1-2 (1795 bp)	1-1 (3809 bp)
	mmEsrrb j3 R	GCCCTTTAGAGCTCCTTCCTTT		
'out-in'	mmEsrrb j5 F	CTGTAAGCATCCCAAGCCGA		2-1 (1031 bp)
	T2A-PuroR-24xMS2 j5 R	ACGTCGTCTCTTGTAGCCAAC		
'in-out'	T2A-PuroR-24xMS2 j3 F	TGCAGGTCGACAAGGTCAC		3-1 (900 bp)
	mmEsrrb j3 R	GCCCTTTAGAGCTCCTTCCTTT		

Tab. 3: MS2 FISH probes

Name	5'-3' Sequence
MS2-FISH-1	CGTTTGAAGATTCGACCTGG-Cy3
MS2-FISH-2	AATACTGGAGCGACGCGTGA-Cy3

Appendix A - MS2-HDR Plasmid Cloning and CRISPR Methods

A. Create a Gene-Specific Entry Vector

1. Prepare the backbone and gene-specific dsDNA gene fragment (Genewiz) by double-digesting each part separately with *MfeI* and *MluI-HF* in 50 ul total volume:

   5µl 10×Cutsmart Buffer
   Xµl Plasmid or dsDNA fragment (up to 1µg)
   1µl *MfeI* (10U/µl)
   1µl *MluI*-HF (10U/µl)
   Yµl H2O up to 50 µl

   Incubate at 37 ˚C for30 minutes
   Purify digested dsDNA fragment with Qiagen PCR Purification Kit, elute 30uL H2O
   Gel-purify 2283bp backbone with Monarch Gel Purification Kit (NEB), elute 20uL H2O
   Step 2 will use 25ng of the vector and 18 ng gene-specific homology arms (1:1 molar ratio with ~17 fmol of each fragment)

2. Set up a T7 ligation reaction:

   Xµl MfeI/MluI-digested backbone (2283bp, 17 fmol~24ng)
   Yµl MfeI/MluI-digested gblock (1650bp, 17 fmol~17.3ng)
   10µl 2X T7 Ligase Buffer (NEB)
   0.5µl T7 Ligase (3000ceu, NEB)
   Z µl H_2O up to 20 µl

   Incubate 1 hour at room temperature or 23-25 ˚C, hold at 4 ˚C.

3. Plasmid-Safe DNase treatment to remove unligated fragments before transformation:

   1.5µl 10x Plasmid-Safe Buffer (Epicentre)
   1.5µl ATP (10 mM)
   11µl T7 ligation reaction
   1µl Plasmid-Safe DNase (Epicentre)

   Incubate at 37 ˚C for 30 min, cool to 4˚C or place on ice before transformation
   Transform 25µl of DH5alpha cells with 1µl of Plasmid-Safe-treated reaction (15µl total).

B. Insert T2A-PuroR-24MS2 (or another cassette) into the Entry Vector plasmid obtained in part A.

1. Obtain a donor plasmid with *EagI/SalI*-compatible restriction sites or create the T2A-PuroR-24xMS2 donor plasmid from *EagI-T2A-PuroR-BamHI*, a PCR product obtained from pCRISPaint-HaloTag-PuroR (Addgene #80960) as described in I-III below.
   I. Q5 Hot-Start High-Fidelity PCR (NEB) and cleanup
      i. Dilute pCRISPaint-HaloTag-PuroR (Addgene #80960) to 1ng/uL
      ii. Primers were ordered in Dry format from IDT at the 25nmol scale with standard desalting, re-suspended upon arrival into 100µM stock solutions with nuclease-free water and maintained at 4˚C. For each PCR reaction, a 10µM working solution of each primer is suggested.:

T2A-PuroR-EagI_F	CCTcggccgTCCGGAGGAGAGGGCAG
T2A-PuroR-BamHI_R	ACTggatccTCACGCTCCAGGCTTCCT

iii. Assemble the following PCR Reaction mixture and incubate as described:

1µl "pCrisPaint-HaloTag-PuroR" plasmid (diluted to 1 ng/µL)
2.5µl F Primer (10 µM)
2.5µl R Primer (10 µM)
10µl Q5 Reaction Buffer (5X)
1µl dNTP Mixture (10mM)
0.5µl Q5 Hot-Start HF Polymerase (NEB)
32.5µl Nuclease-Free H2O up to 50µl

Thermal cycler program:
1) 98˚C – 45 sec
2) 98˚ C - 10 sec
3) 71˚ C – 30 sec
4) 72˚ C – 1 min
5) Repeat steps 2-4 for 35 cycles
6) 72˚ C – 2 min (final elongation)
7) 4˚ C - hold

iv. Purify 50µl of PCR product with Qiagen Qiaquick PCR purification Kit. Elute up to 50µl H2O. Scale up the PCR reaction as needed to obtain 1µg of product for *EagI/BamHI* double-digestion.

II. Digest *EagI-T2A-PuroR-BamHI* PCR product or dsDNA fragment and pDZ415 (24MS2SL loxP-Kan-loxP) with *EagI-HF* and *BamHI-HF* :

5µl 10×Cutsmart Buffer
Xµl Purified PCR product or plasmid DNA (up to 1µg)
1µl *EagI-HF* (10U/µl)
1µl *BamHI*-HF (10U/µl)
Yµl H2O up to 50 µl

Incubate at 37 ˚C for 30 minutes
Gel-purify 669bp digested cassette with Monarch Gel Purification Kit, elute 20uL H2O
Gel-purify 5795bp backbone fragment with Monarch Gel Purification Kit, elute 20uL H2O

III. Insert digested T2A-Puro into pDZ415-MS2 vector to create T2A-Puro-24xMS2 donor plasmid. Adjust concentrations to obtain a desired 1:1-n:1 molar ratio of insert to vector. Higher n factors (n>3) may be helpful if the vector is many times longer than the insert.

Xµl *EagI/BamHI*-digested 24xMS2 backbone (5795bp, 11.17fmol~40ng)
Yµl *EagI/BamHI*-digested T2A-Puro (669bp,11.17-33.51fmol~4.62-13.85ng)
10µl 2X T7 Ligase Buffer (NEB)
0.5µl T7 Ligase (3000ceu, NEB)
Zµl H_2O up to 20µl

Incubate 1 hour at room temperature or 23-25˚C, hold at 4 ˚C.

1.5µl 10x Plasmid-Safe Buffer (Epicentre)
1.5µl ATP (10 mM)
11µl T7 ligation reaction
1.0µl Plasmid-Safe DNase (Epicentre)

Incubate at 37 ˚C for 30 min, cool to 4˚C or place on ice before transformation
Transform 25µl of DH5alpha cells with 1µl of Plasmid-Safe-treated reaction (15µl total).

2. Digest the T2A-PuroR-24xMS2 donor plasmid with *EagI-HF* and *SalI*, then Gel-purify the cassette:

   5µl 10×NEBuffer 3.1
   Xµl Plasmid DNA (1µg)
   1µl *EagI-HF* (10U/µl)
   1µl *SalI* (10U/µl)
   Yµl H2O up to 50 µl

   incubate at 37 ˚C for 30 minutes
   Gel-purify the 1999bp cassette with Monarch Gel Purification Kit and elute 20uL H2O

3. Set up an *AarI* digestion-ligation reaction to insert the cassette into the Entry Vector plasmid:

   1µl Entry Vector plasmid (3933bp,41.15fmol/µl ~100 ng/µl)
   Xµl Digested and purified casette (1999bp,41.15 fmol~50.83ng)
   2µl 10X AarI Buffer (Thermo Fisher)
   0.4µl 50x AarI oligos (Thermo Fisher)
   1µl DTT (10 mM)
   1µl ATP (10 mM)
   1µl AarI (Thermo Fisher)
   0.5µl T7 Ligase (NEB)
   Yµl H_2O up to 20 µl

   Thermal cycler program (ca. 2 hours)
   1) 37˚C - 10 min
   2) 23˚ C - 10 min
   3) Repeat steps 1 & 2 for 6 cycles
   4) 4˚ C - hold

   Plasmid-Safe treatment to remove un-ligated fragments:

   1.5µl 10x Plasmid-Safe Buffer (Epicentre)
   1.5µl ATP (10 mM)
   11µl *AarI* digestion-ligation reaction
   1.0µl Plasmid-Safe DNase (Epicentre)

   Incubate at 37 ˚C for 30 min, cool to 4 ˚C or place on ice before transformation. Transform 25µl of competent cells with 1µl of Plasmid-Safe-treated reaction (15µl total).

   .
4. Grow bacteria in LB or LB+ agar plates at 37 ˚C with 50ug/uL Ampicillin or under the appropriate conditions for the chosen backbone. Select 3 bacterial colonies from each plate and verify sequences according to motifs expected in the final Entry Vector construct.

C. Transfection and growth of monoclonal cell lines

See main text for transfection instructions and growth of monoclonal cell lines.

D. gDNA isolation and PCR to confirm seamless insertion.

1. Harvest ~2x10^6 cells and extract genomic DNA using GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma), following the manufacturer's instructions for Cultured Cell Preparation and DNA Isolation. Elute 100-200µl with Elution Buffer provided in the kit. The gDNA Q5 PCR reactions will use 3x150ng of genomic DNA from each culture.

2. Using the Primer-Blast tool from NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), check primer specificity against the rest of the mouse genome (taxid:10090). We designed the following allele-specific primers to amplify the Esrrb-T2A-PuroR-24xMS2 locus. These were ordered in dry format from IDT at the 25nmol scale with standard desalting, resuspended upon arrival into 100µM stock solutions with nuclease-free water, then maintained at 4˚C.

For the PCR reaction, a 10uM working solution of each primer was assembled. A 50uL stock of each primer should be sufficient to screen up to 10 colonies with three primer pairs each (**Tab. 2**):

Out/out:	
mmEsrrbj5F	CTGTAAGCATCCCAAGCCGA
mmEsrrbj3R	GCCCTTTAGAGCTCCTTCCTTT
Out/in:	
mmEsrrbj5F	CTGTAAGCATCCCAAGCCGA
T2A-PuroR-24xMS2j5R	ACGTCGTCTCTTGTAGCCAAC
In/out:	
T2A-PuroR-24xMS2j3F	tgcaggtcgacaaggtcac
mmEsrrbj3R	GCCCTTTAGAGCTCCTTCCTTT

3. Q5 High-Fidelity Hot-Start PCR (NEB) and gel analysis (gDNA PCR version):

i. Dilute gDNA to 120-150 ng/uL in H2O.
ii. Assemble PCR Reaction mixtures and incubate as described:

1 µl	gDNA (diluted to 150 ng/µL)
2.5 µl	F Primer (10 µM)
2.5 µl	R Primer (10 µM)
10 µl	Q5 Reaction Buffer (5X)
1 µl	dNTP Mixture (10mM)
0.5 µl	Q5 Hot Start High-Fidelity Polymerase (NEB)
32.5 µl	Nuclease-Free H2O up to 50µl

Thermal cycler program:
1) 98˚C – 1 min
2) 98˚C - 15 sec
3) 68˚C – 30 sec
4) 72˚C – 4 min
5) Repeat steps 2-4 for 30 cycles
6) 72˚C – 5 min (final elongation)
7) 4˚C - hold

iii. Resolve all three 50uL PCR products (out/out, out/in, in/out) on a 1% Agarose gel in separate lanes, as shown in **Fig.1**. For a single Esrrb-T2A-PuroR-24xMS2 allele, in the case of seamless insertion at the 5' and 3' homology-arm junctions respectively, the out/in and in/out lanes show a single fragment each of 1031bp and 900bp. These were Gel-purified and directly sequenced with the generative primers using Sanger-sequencing (Genewiz).

In the presence of heterozygous alleles, the out/out lane shows a mixture of bands with 1795bp (WT) and 3809bp (seamless insertion) lengths. Different alleles may also be observed, especially when generated by non-HDR-mediated pathways, which makes out/out PCR products difficult to predict and often impossible to sequence directly. Further genotyping and verification should be performed as described in the main text.

Suggested Materials:

Material	Vendor	Cat. #
FragmentGENE 1,501-1,750 bp	Genewiz	
Custom Backbone with MfeI/MluI cloning sites	Geneart (Life Technologies)	
MfeI (10U/µL)	NEB	R0589
MluI-HF® (10U/µL)	NEB	R3198
T7 Ligase (3,000,000 units/mL)	NEB	M0318
QIAquick PCR Purification Kit(50)	QIAGEN	28104
Monarch® DNA Gel Extraction Kit	NEB	T1020
Agarose	Sigma	A9539
Plasmid-Safe™ ATP-Dependent Dnase (10U/µL)	Lucigen	E3101K
Subcloning Efficiency™ DH5α™ Competent Cells	Invitrogen™	18265017
Nuclease-Free Water (not DEPC-treated)	Invitrogen™	AM9937
EagI-T2A-PuroR-BamHI cassette can be obtained in 2 ways: *Follow step B) O.I.i.-iii.* ***Or*** *Order as FragmentGENE 1,501-1,750 bp*		
FragmentGENE 1,501-1,750 bp: The 5'-3' sequence shown here is for EagI-T2A-PuroR-BamHI. This can be modified to replate PuroR with HygroR or NeoR. LoxP or FRT sites may also be added around the resistance fragment for recominase-based exision.	Genewiz	
pCRISPaint-HaloTag-PuroR (Addgene #80960)	Addgene	
T2A-PuroR-EagI, T2A-PuroR-BamHI primers	IDT	
EagI-HF® (10U/µL)	NEB	R3505
BamHI-HF® (10U/µL)	NEB	R3198
pDZ415 (24MS2SL loxP-Kan-loxP) (Addgene #45162) **or** pDZ416 (24PP7SL loxP-Kan-loxP) (Addgene#45163)	Addgene	
SalI (10U/µL)	NEB	R0138
AarI (2 U/µL)	Thermo Scientific™	ER1581
USB Dithiothreitol (DTT), 0.1M Solution	Thermo Scientific™	707265 ML

GenElute™ Mammalian Genomic DNA Miniprep Kit	Sigma	G1N70
Q5® Hot Start High-Fidelity DNA Polymerase	NEB	M0493
Deoxynucleotide (dNTP) Solution Mix (10mM)	NEB	N0447
Primers for gDNA amplification and sequencing	IDT	