CHAPTER 3

3 PROOF-OF-CONCEPT STUDIES FOR THE USE OF CRISPR/CAS9 GENETIC SCREENS IN THE DIFFERENTIATION OF ESCS TO TSCS

3.1 Introduction

Stem cell lines can be established from the three layers of cells that compose a mouse blastocyst. These are embryonic stem cells (ESCs), derived from the epiblast (Evans and Kaufman 1981, Martin 1981), trophoblast stem cells (TSCs), derived from the trophectoderm layer (Tanaka, Kunath et al. 1998) and extraembryonic endoderm cells (XEN), derived from the primitive endoderm (Kunath, Arnaud et al. 2005). They represent important *in vitro* models to study embryo development considering their ability to indefinitely self-renew, maintain an undifferentiated state when cultured in appropriate conditions, and their capacity to retain both the developmental potential and lineage restrictions of their embryonic counterparts. I focus particularly on stem cells representative of the two layers generated in the embryo upon the first cell lineage decision, ESCs and TSCs. ESCs are classically regarded as pluripotent, based on their capacity to generate chimeras and differentiate into all embryonic lineages when engrafted into morulae or blastocysts, and their lack of contribution to extraembryonic tissues (Evans and Kaufman 1981, Martin 1981, Smith, Heath et al. 1988, Beddington and Robertson 1989, Ying, Wray et al. 2008). On the other hand, TSCs are multipotent cells restricted to the extraembryonic lineage in chimera assays, unable to contribute the embryo proper (Tanaka, Kunath et al. 1998). The study of ESCs and TSCs provides important tools to address the regulation of their embryonic counterparts, both for lineage establishment and fate restriction.

3.1.1 Bypassing the first cell lineage decision: conversion of ESCs to TSC fate

Under normal conditions, ESCs retain epiblast lineage restrictions and have limited ability to spontaneously differentiate to extraembryonic tissues (Beddington and Robertson 1989). However, several genetic approaches were reported that could bypass the lineage barrier and convert ESCs to a TSC fate shedding light on important regulators of lineage decision and maintenance, as well as key factors for TSC and TE transcriptional networks. These have different extents of lineage conversion as shown by *in vivo* potency and are summarised on Table 3.1. They can be subdivided in knockout approaches, which likely result in abolishing the mechanisms of fate restriction, and overexpression approaches, which enforce the activation of TSC transcription circuitry.

Conditional knockout of *Oct4* was the first study demonstrating that expression levels of this transcription factor in ESCs are tightly connected with cell fate, whereby forced repression of *Oct4* allows differentiation to TSC (Niwa, Miyazaki et al. 2000). The role of *Oct4* repression was later mimicked by enforced expression of *Cdx2* or *Eomes*, which redirected the transcriptional network to acquire a TSC fate (Niwa, Toyooka et al. 2005), demonstrating the key role of these two transcription factors in TSC. Furthermore, this study established the direct interaction between *Oct4* and *Cdx2*, resulting in reciprocal repression, as a key determinant for lineage segregation in the mouse embryo. Activation of *Cdx2* was also achieved through conditional expression of activated *Ras*, enabling the derivation of TSCs from ESCs and implicating Ras-MAPK signalling in promoting TE formation in embryos (Lu, Yabuuchi et al. 2008).

A key barrier in lineage restriction is achieved through DNA methylation. Indeed, this was demonstrated in the ESC/TSC context by the fact that *Dnmt1* deficient ESCs bypass their lineage commitment and differentiate to TSC (Ng, Dean et al. 2008). Analysis of differentially methylated promoters in ESCs, TSCs and Dnmt1 deficient ESCs, revealed that *Elf5* was the key hypermethylated and repressed gene in ESC state, and demethylated and expressed in the remaining two cases. This established *Elf5* promoter methylation as a key player on the maintenance of the first cell lineage decision. Furthermore, overexpression of *Elf5* in ESCs induced trophoblast transition that was not stable in the stem cell state, progressing to terminal differentiation (Ng, Dean et al. 2008). Highlighting the importance of a correct balance between DNA methylation and demethylation for proper lineage restriction, later experiments showed that

downregulation of the 5-mC hydroxylase *Tet1* in ESCs also allows differentiation towards TSC (Koh, Yabuuchi et al. 2011).

The demonstration that *Tead4* was involved in initiation of TE formation upstream of *Cdx2* (Nishioka, Yamamoto et al. 2008), prompted the follow up study showing that inducible expression of active *Tead4* (*Tead4VP16ER*) also resulted in activation of *Cdx2* in ESCs and their conversion to TSC (Nishioka, Inoue et al. 2009). Ralston and colleagues analysed a previous microarray expression dataset for genes specifically enriched in TSCs compared to ESC and XEN cells (Kunath, Arnaud et al. 2005), finding *Gata3* among these, in addition to the already established transcription factors *Cdx2*, *Eomes* and *Tead4*. They proved that overexpression of *Gata3* in ESCs can promote TSC differentiation independently of *Cdx2*, demonstrating a parallel pathway to Cdx2 regulation downstream of Tead4 that also coordinates trophoblast fate (Ralston, Cox et al. 2010). Moreover, in this case, Gata3 overexpression induces differentiation of ESCs to differentiated trophoblast rather than maintaining a stable stem cell state. Similarly to *Gata3*, overexpression of *Tfap2c* in ESCs resulted in induction of TSC fate independently of *Cdx2* (Kuckenberg, Buhl et al. 2010). The authors also showed that both *Cdx2* and *Tfap2c* are required for the upregulation of *Elf5*, which further promotes TSC maintenance (Kuckenberg, Buhl et al. 2010).

More recently, ESC to TSC differentiation was explored to elucidate the role of microRNAs in this lineage transition. It demonstrated that upregulation of TE-specific microRNAs *miR-15b*, *miR-322* or *miR-467g* in ESCs was sufficient to generate stable TSC lines, expanding the TSC regulatory network beyond transcription factors and epigenetic modifiers (Nosi, Lanner et al. 2017).

All these studies illustrate that differentiation of ESC to TSC is a faithful *in vitro* model that allows meaningful dissection of regulators of the first cell lineage decision in the mouse embryo. This is clearly illustrated by the fact that most of the studies mentioned either confirmed that: i) *in vivo* findings were also relevant for *in vitro* cell lines; or ii) conversely, *in vitro* discoveries were often confirmed *in vivo*.

Table 3.1 **I Summary of genetic approaches used to circumvent lineage restriction in ESCs and allow differentiation to TSC-like states.** N.A., not assessed.

Approach	Chimera Assay	*Elf5* expression / DMR analysis	Reference
Oct4 knockout	N.A.	Upregulation *Elf5* / partial demethylation DMR region	(Niwa, Miyazaki et al. 2000, Cambuli, Murray et al. 2014)
Cdx2 overexpression	Placenta contribution	Upregulation *Elf5* / slight demethylation DMR region	(Niwa, Toyooka et al. 2005, Tolkunova, Cavaleri et al. 2006, Cambuli, Murray et al. 2014)
Eomes overexpression	N.A.	N.A.	(Niwa, Toyooka et al. 2005)
Dnmt1 knockout	*Ex vivo* contribution to TE in blastocysts	Upregulation of *Elf5* / partial demethylation DMR region	(Ng, Dean et al. 2008)
Elf5 overexpression	N.A.	Upregulation of *Elf5*	(Ng, Dean et al. 2008)
Ras activation	Placenta contribution	No *Elf5* upregulation / hypermethylated DMR	(Lu, Yabuuchi et al. 2008, Cambuli, Murray et al. 2014)
Tead4-VP16 activation	N.A.	N.A.	(Nishioka, Inoue et al. 2009)
Gata3 overexpression	N.A.	N.A.	(Ralston, Cox et al. 2010)
Tfap2c overexpression	*Ex vivo* contribution to TE in blastocysts	Low *Elf5* upregulation / N.A.	(Kuckenberg, Buhl et al. 2010)
Tet1 knockdown	Placenta contribution	Moderate *Elf5* upregulation / hypermethylated DMR	(Koh, Yabuuchi et al. 2011)
Arid3a upregulation	*Ex vivo* contribution to TE in blastocysts	N.A.	(Rhee, Lee et al. 2014)
CRISPRa-mediated *Cdx2* overexpression	Placenta contribution	*Elf5* upregulation / DMR hypomethylation	(Wei, Zou et al. 2016)
Overexpression *miR-15b* (or *miR-322, miR-467b*)	Extraembryonic contribution in E6.5 embryos	*Elf5* upregulation	(Nosi, Lanner et al. 2017)

3.1.2 Epigenetic memory of the first cell lineage

ESCs and TSCs are characterised by distinct distribution of active and repressive histone modifications (Alder, Lavial et al. 2010, Rugg-Gunn, Cox et al. 2010), in addition to unique DNA methylation profiles (Senner, Krueger et al. 2012) which contribute to both define their identity and expression patterns, and restrain them to their respective lineage. Despite the cumulative reports of ESC conversion to TSC, the extent to which different strategies fully reprogramme their transcriptomic and epigenomic landscapes was only recently analysed in a study by (Cambuli, Murray et al. 2014). The authors performed an in-depth characterisation of several methods that were shown to achieve TSC differentiation from ESC lines - conditional *Oct4* deletion, overexpression of *Cdx2* and conditional activation of *Ras-Erk1/2*.

Gene expression and epigenetic analysis, found that none of the abovementioned approaches achieves an expression profile similar to bona-fide TSCs. Indeed, the two methods with best performance were *Oct4* deletion (achieving only about 20% of *Cdx2*, *Elf5* and *Eomes* expression relative to TSCs) and *Cdx2* upregulation (which achieved 40% of TSC expression levels). Activation of Ras-Erk1/2 produced low upregulation of these transcription factors, with *Elf5* expression not detected. Importantly, analysis of DNA methylation patterns in these models revealed that the *Elf5* promoter remains heavily methylated in most cases, demonstrating these differentiated cells retain an epigenetic memory of the first cell lineage decision. In this case, the concept of "epigenetic memory" is used to describe the fact that converted cells maintain a methylation status at the *Elf5* promoter that is characteristic of their ESC-state rather than fully erasing this mark and acquiring a methylation level equivalent to bona-fine TSCs.

Consistent with their expression profiles, methylation levels in *Cdx2* overexpression-derived TSCs were partially reprogrammed to 64% whereas *Oct4* knockout-derived TSCs showed a persistence of 26%. Exploring genome-wide methylation patterns led to the identification of nine other genes with potential "gatekeeper" function, similarly to *Elf5* – that is, highly methylated and repressed in ESCs and demethylated and expressed in TSCs, which further confirmed the partial reprogramming to TSC fate.

Nevertheless, many of the published models have reported contribution of induced TSCs to placenta in chimera assays (see Table 3.1), demonstrating either the presence of small populations that fully reprogramme to functional TSCs or completion of reprogramming upon integration in the embryo environment. Most of these studies however rely on positioning of injected cells in the blastocyst/placenta structures and do not assess their true functional identity. In fact, only cells resulting from *Cdx2*

overexpression were analysed by co-immunostainings to demonstrate that their presence in extraembryonic regions indeed is complemented with *in vivo* expression of trophoblast markers (Cambuli, Murray et al. 2014). Further evaluation of the remaining models will help clarify their *in vivo* potency.

In a process equivalent to the generation of induced pluripotent stem cells (iPSCs), the conditions to reprogramme mouse embryonic fibroblasts (MEFs) to induced trophoblast stem cells (iTSCs) were recently reported by two independent groups. These relied on the overexpression of a transcription factor cocktail with either *Gata3*, *Eomes*, *Tfap2c* and *Ets2* (Kubaczka, Senner et al. 2015), or *Gata3*, *Eomes*, *Tfap2c* and the optional use of *Ets2* or *Myc* (Benchetrit, Herman et al. 2015). The resulting iTSCs had both transcriptome and *Elf5* promoter methylation status similar to bona-fide TSCs. They also achieved *in vivo* contribution to placenta when injected into chimeras. Strikingly, when the same transcription factor cocktail was overexpressed in ESCs, these proved resistant to iTSC reprogramming, further demonstrating the strength of the first cell lineage restriction (Kubaczka, Senner et al. 2015).

A recent study that converted ESCs to TSCs through the activation of *Cdx2* using CRISPRa, also generated stable TSCs with *Elf5* upregulation and hypomethylated promoter (Wei, Zou et al. 2016).

Nevertheless, all of the reports that induce TSC fate achieving equivalent expression and epigenetic landscape to bona-fide TSCs (Benchetrit, Herman et al. 2015, Kubaczka, Senner et al. 2015, Wei, Zou et al. 2016) had subcloning steps in their protocols, in which TSC-like colonies were picked and further expanded. This might be selective for cells that had the capacity to bypass the first cell lineage memory and fully reprogramme to TSC fate, in a similar process to establishing iPSC lines from primary colonies.

A key need in the field is to define the end point of each experimental model. Currently, many studies rely on upregulation of key marker genes and evaluate heterogeneous populations by the end of a fixed differentiation time-course. It will be valuable to distinguish which genetic modifications allow establishment of self-renewing TSC-like states in the future. Additionally, systematic analysis of the extent of epigenetic reprogramming of ESCs to TSCs will continue to elucidate mechanisms involved in the maintenance of ICM and TE segregation.

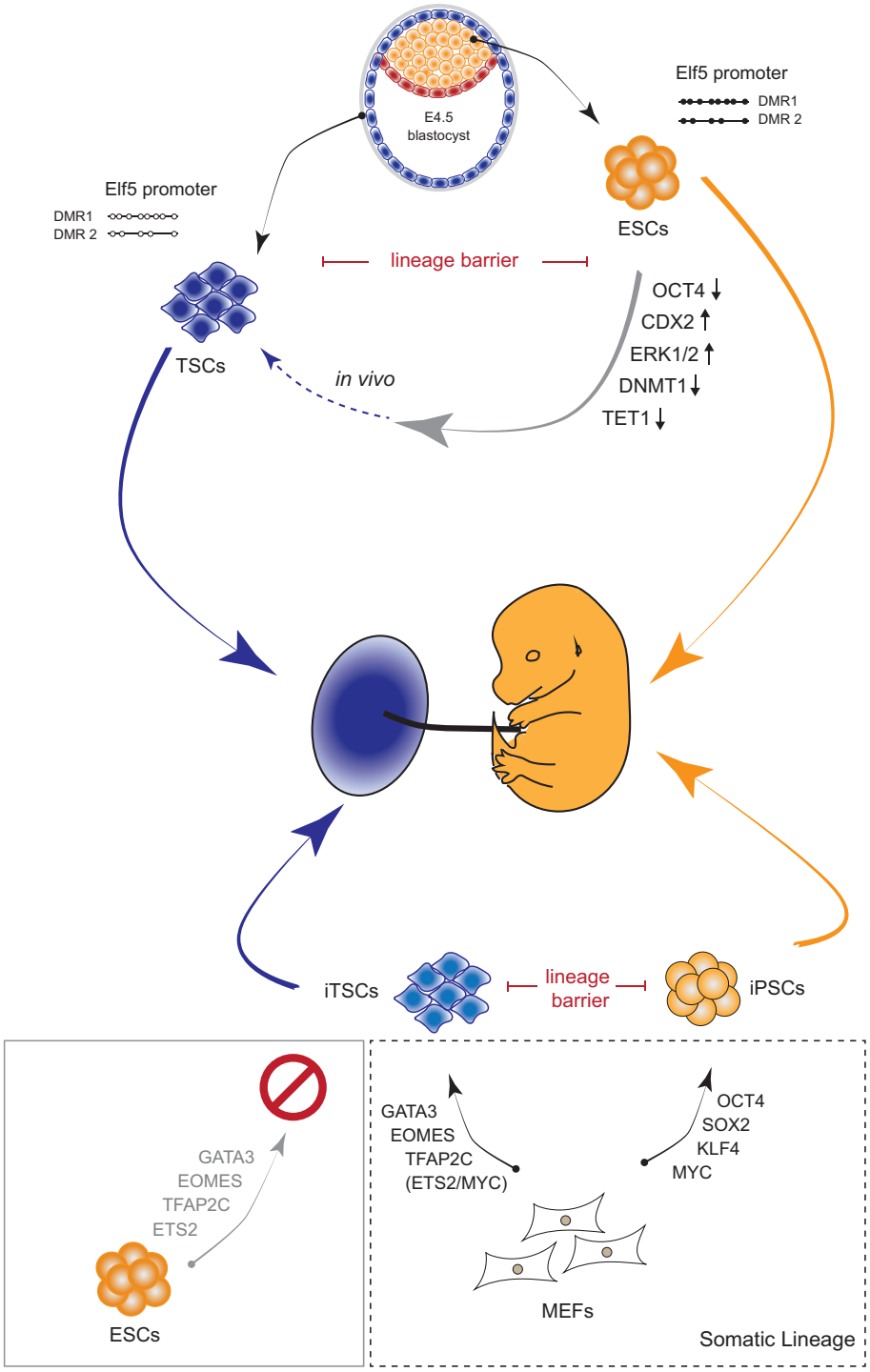


Figure 3.1 | Epigenetic restriction of the first cell lineage decision. Model for the epigenetic memory of ICM/TE segregation in ESCs, that prevents their complete reprogramming to TSCs. This lineage barrier is specially preserved at the *Elf5* promoter methylation status, which is hypermethylated in ESCs and hypomethylated in TSCs, but remains mostly methylated in many genetic approaches for ES-to-TS cell conversion. Conversion can be completed upon incorporation into *in vivo* embryonic environment. This lineage restriction is strong enough to prevent ESC reprogramming to iTSCs, following the overexpression of a cocktail capable of successfully reprogramme MEFs. Image based on Cambuli *et al* (Cambuli, Murray et al. 2014).

3.1.3 Role of *Elf5* in the trophoblast compartment

Elf5-null embryos implant and form the ectoplacental cone, but fail to develop extraembryonic ectoderm by E6.5 and arrest between E8.5 and E10.5 (Donnison, Beaton et al. 2005). The inability to derive TSC lines from these mutant embryos shows it has an essential role for the establishment and maintenance of these *in vitro* stem cells (Donnison, Beaton et al. 2005).

Using ESC differentiation to TSC as a model system, *Elf5* was later established as a gatekeeper of the first cell lineage decision due to the differential methylation profile of its promoter in ESCs and TSCs (Ng, Dean et al. 2008). *Elf5* acts downstream of *Cdx2* and *Eomes* transcription factors and creates a positive feedback loop that reinforces trophoblast stem cell circuitry and is necessary to expand the trophoblast compartment. In the absence of *Elf5*, low levels of these two transcription factors do not result in establishment of trophoblast fate upon ESC differentiation. This feedback loop acts specifically in the trophoblast lineage due to hypomethylation of *Elf5* promoter. On the other hand, hypermethylation blocks *Elf5* expression in ESCs thereby restricting the activation of a stable TSC fate in the pluripotent compartment.

Elf5 reinforcement of TSC network is limited in time, as continued expression results in terminal differentiation of ESC-converted TSCs, which positions *Elf5* both at the centre of self-renewal and onset of differentiation (Ng, Dean et al. 2008, Latos, Sienerth et al. 2015). In fact, a recent study (Latos, Sienerth et al. 2015) showed that *Elf5* regulates TSC self-renewal and differentiation in a tight stoichiometry balance with *Eomes* and *Tfap2c*. In a stem cell state where *Eomes* is expressed in at least equal stoichiometry to *Elf5* and *Tfap2c*, the three associate in a tripartite complex and bind preferably to self-renewal genes, promoting their expression and reinforcing the network. On the other hand, at the onset of TSC differentiation, the expression of *Eomes* is reduced and the association of *Elf5* and *Tfap2c* as a bipartite complex alters their binding profile to a differentiation promoting program. This way, tight regulation of *Elf5* levels, in a parallel to the role of *Oct4* in pluripotency (Niwa, Miyazaki et al. 2000, Radzisheuskaya, Chia Gle et al. 2013), dictate self-renewal and differentiation programmes in TSCs.

3.1.4 Scope of this chapter

This chapter describes the framework to allow genome-wide CRISPR/Cas9 knockout screening for the unbiased identification of genes preventing ESCs from adopting a trophoblast fate. I first describe the development of a *Elf5::Venus* reporter ESC line that constitutively expresses Cas9 from the *Rosa26* locus. This allows both efficient genetic manipulation using CRISPR/Cas9 and monitoring differentiation towards trophoblast stem cell lineage. I next describe the validation of this line and confirmation of the transcription profile of Venus positive cells. I then proceed with exploring the feasibility of using CRISPR/Cas9 to knockout two genes known to prevent ESC to TSC differentiation, *Oct4* and *Dnmt1*, and optimise the experimental setup for differentiation. I finally summarise the lessons learnt from the different proof-of-concept experiments and discuss considerations for the experimental setup to be used in the genome-wide screening to follow.

3.2 Results

3.2.1 Generation of an *Elf5::Venus* ESC reporter line for trackable detection of trophoblast differentiation

3.2.1.1 Development of *Elf5::Venus* knock-in ESCs with stable expression of Cas9

Homogeneous expression of Cas9 is an important parameter when considering genome-wide screening to avoid bias arising from gRNAs transduced into subpopulations with either higher Cas9 expression or less functional protein (Zhou, Zhu et al. 2014). Indeed, it has been shown that in a mixed population of Cas9-integrated cell lines, often mutations can arise in Cas9 that render it inactive (Tzelepis, Koike-Yusa et al. 2016). I therefore chose to engineer the JM8-*R26-Cas9* ESC line to generate the *Elf5::Venus* reporter, as it has stable integration of *EF1α-Cas9::IRES::Neo*^{R} in the *Rosa26* locus (Figure 3.2A) and was clonally selected for high editing efficiency (Tzelepis, Koike-Yusa et al. 2016).

To ensure proper regulation of the Venus reporter, I decided to target the endogenous *Elf5* locus. I used the T2A self-cleavage peptide so that Elf5 was not affected by fusion to Venus, and an H2B sequence was included to induce nuclear localisation of Venus and allow a more defined fluorescent signal for microscopy.

A schematic representation of the CRISPR/Cas9-mediated homology-directed repair strategy used for knock-in is shown in Figure 3.2. I constructed a targeting vector that contains 700-bp homology arms, a *T2A::H2B::Venus* cassette to express nuclear-localised Venus protein upon *Elf5* expression, and an excisable selection cassette composed of *PGK-Bsd*^{R} flanked by *LoxP* sites for blasticidin selection (Figure 3.2C). The excisable cassette was included to allow selection of integrated clones with blasticidin, followed by removal of this cassette to promote minimal disruption of the endogenous *Elf5* locus in the final cell line. I then co-transfected ESCs with the donor construct and a gRNA-expressing plasmid (Figure 3.2D) used to induce a double-strand break downstream of the stop codon and facilitate homology-directed repair (Figure 3.2B). Colonies were selected for stable integration in Blasticidin (Figure 3.2E). Targeted clones were identified by genotyping PCR, using the primer strategy illustrated in Figure 3.2. Briefly, these primers bind outside the homology arms in the endogenous *Elf5* locus, amplifying a 2 kb band in *wild-type* locus (Figure 3.2B) and a 5 kb band in successfully targeted clones (Figure 3.2E). Finally, to guarantee minimal perturbation of the

endogenous *Elf5* locus, successfully cloned cells were transiently transfected with *Cre* to mediate excision of the selection cassette, leaving the targeted locus with *Elf5::T2A::H2B::Venus* (Figure 3.2F). Following transfection with Cre, colonies were picked and genotyped with the same pair of primers used before, which generate a 3.5 kb band when the selection cassette is efficiently removed (Figure 3.2F).

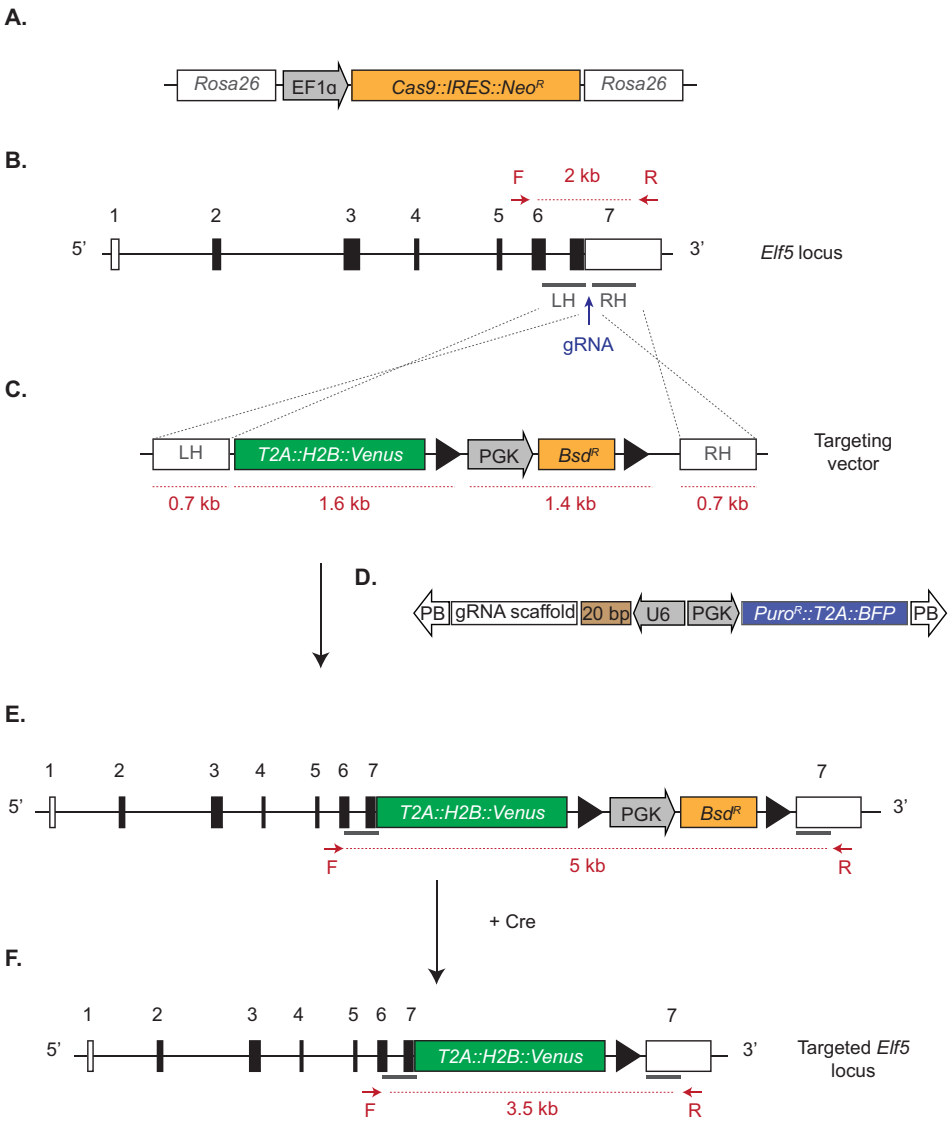


Figure 3.2 | Schematic representation of the strategy used to generate the *Elf5::Venus* reporter ESC line. A | The JM8-*R26-Cas9* line used in this work has stable integration of Cas9 in the Rosa26 locus, expressed from the *EF1α* promoter followed a neomycin resistant gene (*Neo*^{R}). *EF1α*, human elongation factor 1a. *IRES*, internal ribosome entry site. **B |** Representation of the *Elf5* locus. Underlined in grey are the regions used for left (LH) or right (RH) homology arms and outlined in red is the location of the primers used for genotyping. gRNAs designed target the end of *Elf5* coding sequence. **C |** Targeting vector, composed of LH and RH sequences, a cassette for integration *of*

T2A::H2B::Venus, and a excisable selection cassette that expresses blasticidin resistant gene (*Bsd*^{R}) from the PGK promoter, flanked by LoxP sites (triangles). T2A, T2A self-cleaving peptide. H2B, Histone 2B. PGK, mouse *Pgk1* promoter. **D |** *piggyBac* vector carrying a gRNA expression cassette driven by the human U6 promoter (U6) and a Puromycin-resistance gene (*Puro*^{R}) fused with a T2A to blue fluorescent protein (BFP), under the regulation PGK promoter. PB, *piggyBac* repeats. 20 bp, gRNA specificity sequence. **E |** Schematic representation of the *Elf5* locus upon the first round of targeting. **F |** Final edited *Elf5* locus, obtained after removal of the blasticidin selection cassette upon transfection with *Cre*.

In order to obtain efficient homology-directed repair mediated by CRISPR/Cas9, one should design gRNAs up to 100-bp from the desired insertion site with the distance between this and the PAM sequence being one of the key determinants for targeting efficiency (Paquet, Kwart et al. 2016, Liang, Potter et al. 2017). This limited sequence does not always allow the design of a gRNA with high cutting efficiency and minimal off-targets. With this in mind, I tested the use of four different gRNAs to induce double-stand break up to 30-bp downstream of the *Elf5* stop codon (Figure 3.3A). Note the exclusion of the stop codon in the left homology arm to allow for *T2A::H2B::Venus* expression following Elf5 expression (Figure 3.3A).

Four independent transfections were carried-out using each gRNA and the targeting vector, followed by selection with blasticidin. It is important to highlight that even though the expression plasmid for gRNA delivery contains *piggyBac* repeats, the transfections to generate the reporter line were all transient to both minimise gRNA off-targets, and avoid conflicts with downstream applications of the final line. These will rely on both *BFP* and *Puro*^{R} expression. Twenty-four blasticidin resistant colonies were picked per transfection, expanded on feeders and genotyped to detect the targeted insertion of the *T2A::Venus-LoxPBsd* cassette. Figure 3.3B shows the genotyping results from colonies transfected with gRNA #1. The gRNAs had similar efficiencies independently of the gRNA used, ranging from 21% to 29% (Figure 3.3C). At this point, I chose to proceed with clones generated with gRNA #1 given it was the one with the lowest number of predicted off-targets (3 exonic off-targets, compared to 7, 6 and 9 for gRNAs #2, #3 and #4, respectively). ESC clone 1.18 was then transiently transfected with *Cre* to mediate excision of the blasticidin selection cassette followed by colony picking and genotyping (Figure 3.3D). Note that before this point, cells were grown on a feeder layer and hence show a wild-type band which is no longer visible in the clones picked after Cre excision, routinely grown on gelatin plates (Figure 3.3B and D). The final line carries homozygous knock-in and is referred from now on as *Elf5::Venus* ESC.

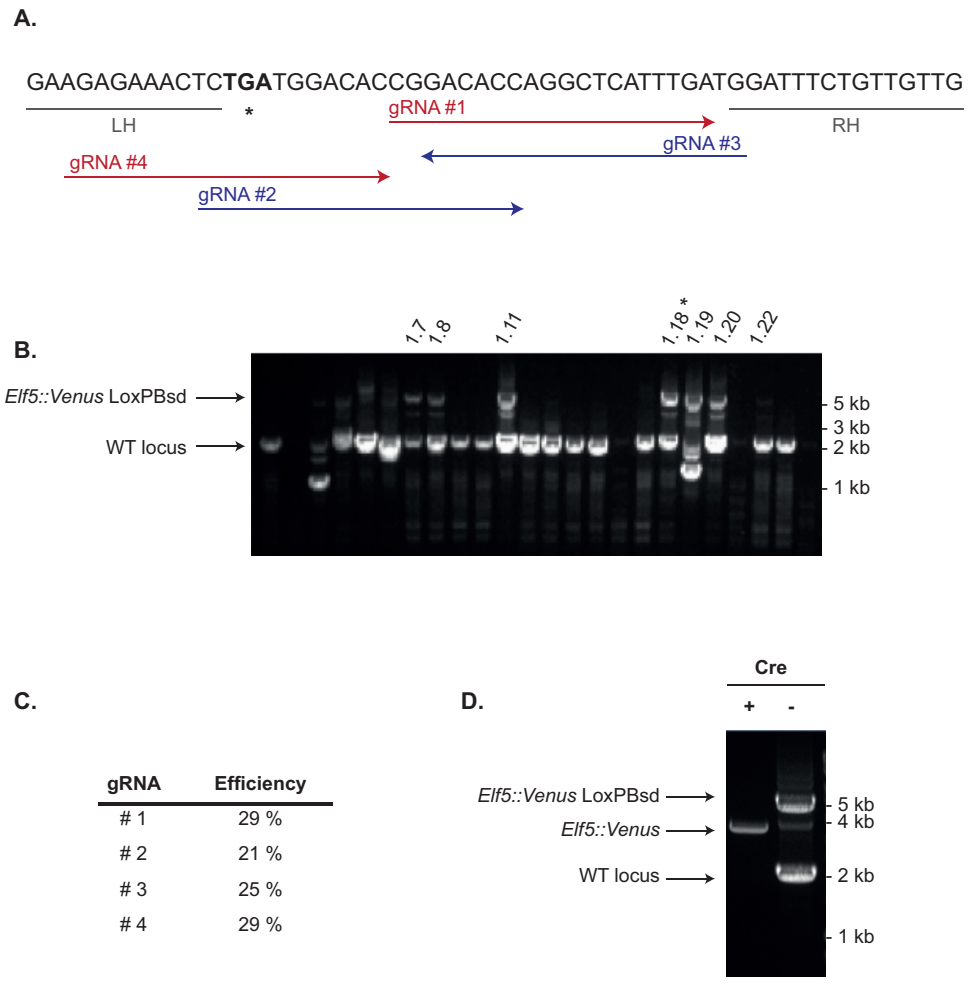


Figure 3.3 | Generation of the *Elf5::Venus* knock-in ESC line. A | Diagram showing the design of the four gRNAs used to induce double-strand break upstream of mElf5 stop codon (highlighted in bold and asterisk). The boundaries of both left homology (LH) and right homology (RH) arms are underlined in grey. **B |** Genotyping results for 24 colonies resistant to blasticidin selection, picked from targeting reaction using gRNA #1. Positive clones are identified by numbers, with an asterisk marking the clone chosen for *downstream work*. **C |** Table depicting the efficiency of homologous recombination by gRNA used. **D |** Confirmation of excision of Blasticidin selection cassette, mediated by Cre.

As a quality control to guarantee that the line generated was not affected by gRNA #1 outside its predicted target, I PCR amplified and sequenced the three predicted exonic off-target regions. I sequenced both the original JM8-*R26-Cas9* ESC line, and four different clones generated through this gRNA. All of them displayed the *wild-type* sequence at these sites, with absence of CRISPR/Cas9-mediated indels (Figure 3.4).

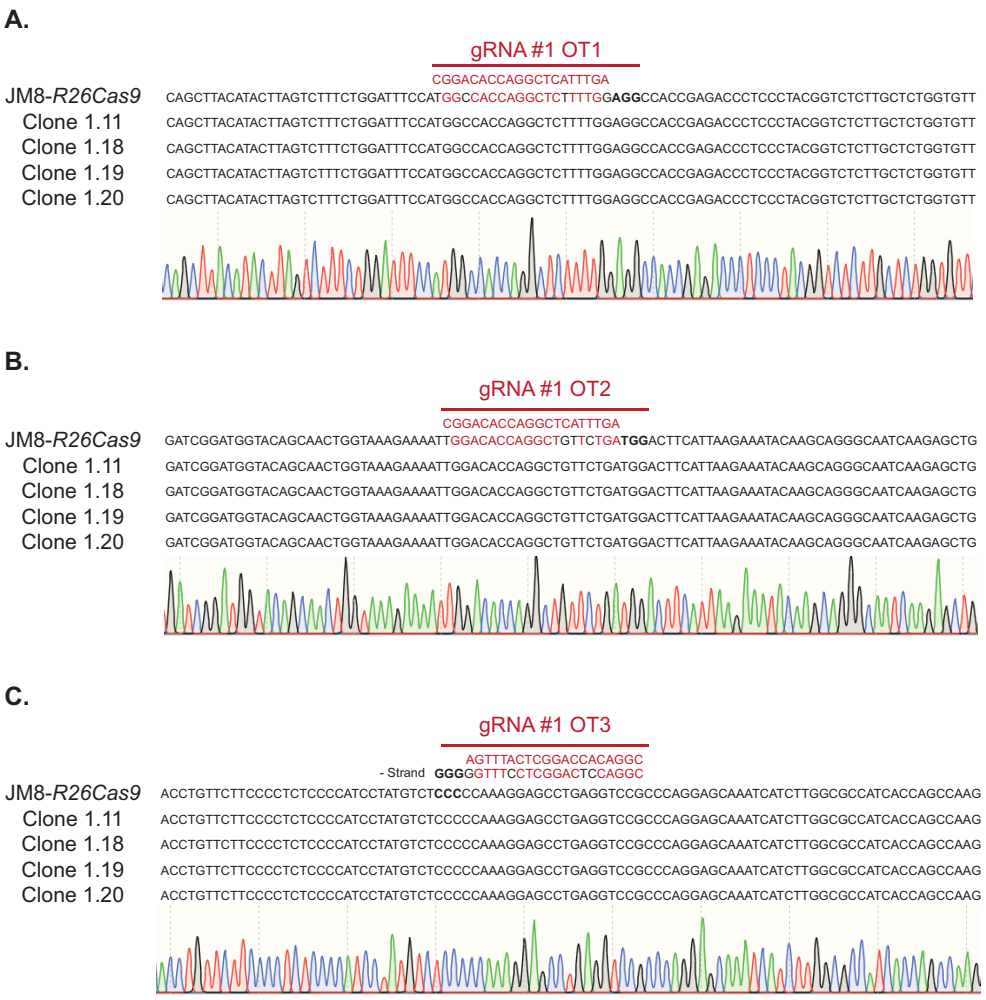


Figure 3.4 | Absence of Cas9-mediated indels in the three predicted exonic off-targets of gRNA #1. The genomic regions for these were amplified by PCR and sequenced for the non-targeted JM8-*R26Cas9* line, and for four different clones generated with gRNA #1. Shown are the alignments of the results from the five samples with a representative Sanger sequencing chromatogram depicted for each off-target (OT). The OT region is shown aligned with gRNA #1 in red, with mismatches in black. PAM sequence in each OT is highlighted in bold. **A |** Off-target 1 (OT1), located on gene Nt5m (NM134029). **B |** Off-target 2 (OT2), located on gene Slc12a8 (NM001013766). **C |** Off-target 3 (OT3), located on gene Samd3 (NM001083902). Note that in this case, the OT sequence is present in the minus strand (-Strand).

3.2.1.2 Validation of *Elf5::Venus* ESC for its use in TSC differentiation

One of the key distinctions between ESCs and TSCs is the methylation status of *Elf5* promoter region, which appears hypermethylated in ESCs and unmethylated in TSCs (Ng, Dean et al. 2008). It is therefore important to verify the integrity of this regulation layer upon *Elf5* targeting as it can affect the lineage barrier in a direct manner.

To assess the effect of targeting *Elf5* locus on the methylation status of *Elf5* promoter DMRs, the genomic DNA from JM8-*R26-Cas9* and the knock-in *Elf5::Venus* ESC lines was bisulphite converted followed by PCR amplification of the DMR regions (Frommer, McDonald et al. 1992, Herman, Graff et al. 1996). 10 colonies obtained from cloning of each PCR product were analysed by sanger sequencing to determine their methylation levels. The results demonstrate that *Elf5* promoter remains highly methylated upon gene editing in *Elf5::Venus* ESCs (Figure 3.5). This dataset was obtained in collaboration with Dr. Sandeep Rajan.

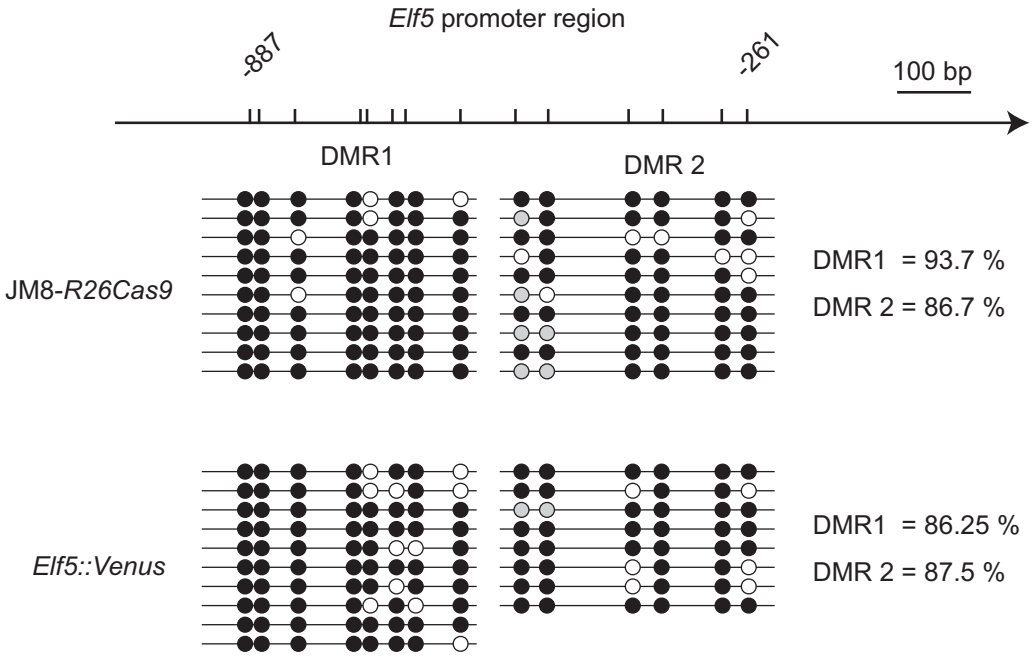


Figure 3.5 | Analysis of *Elf5* promoter DMR methylation status upon gene targeting demonstrates it remains highly methylated. Circles represent CpGs, with closed and open circles indicating methylated and unmethylated CpG, respectively. Grey circles indicate CpG that could not be assessed due to sequencing quality.

In order for *Elf5::Venus* ESC to be used in this study, one needs to prove that expression of Venus faithfully reflected the dynamics of Elf5 activation. To test this, and given that *Elf5* is not expressed in ESCs, I had to induce TSC differentiation. *Oct4* knockout is one of the most robust methods for differentiation of ESCs to TSCs, being reproduced independently in several labs (Niwa, Miyazaki et al. 2000, Niwa, Toyooka et al. 2005, Alder, Lavial et al. 2010, Cambuli, Murray et al. 2014), in addition to its well-known role as a direct regulator of the first cell lineage decision in the mouse embryo (Nichols, Zevnik et al. 1998). Importantly, it has been shown that *Oct4* knockout is one

of the few published methods achieving *Elf5* upregulation during TSC differentiation, as well as *Elf5* promoter demethylation (Cambuli, Murray et al. 2014).

For this validation experiment (Figure 3.6A and B), I transfected *Elf5::Venus* ESCs with a pool of three gRNAs targeting *Oct4*. In this experiment I used PBase to mediate gRNA integration and stable expression. Given the role of *Oct4* in ESC self-renewal, BFP positive cells containing gRNA expression plasmid were sorted 48h post transfection and directly replated for differentiation on MEF feeders and TSC media (Tanaka, Kunath et al. 1998). The cell density at day zero was chosen based on published work and set to 20.000 cells per well of a 12-well plate. TSC differentiation was carried-out for 10 days with the percentage of Venus positive cells assessed by flow cytometry. In this experimental setting, cells transfected with *Oct4* gRNAs, generated about 4% Venus expressing cells, peaking at day 6 of differentiation (Figure 3.6C). In contrast, untransfected cells peaked at only about 1% Venus expressing cells, also by day 6 (Figure 3.6C). Importantly, bulk RT-qPCR analysis at the peak of differentiation demonstrated modest upregulation of *Elf5* at the population level, detectable upon *Oct4* knockout but not in the untransfected control (Figure 3.6D). On the other hand, similar levels of *Cdx2* upregulation were observed for both control and *Oct4* knockout, pointing to an easier activation of *Cdx2* expression upon induction of TSC differentiation, in comparison with *Elf5*. *Eomes and Tpbpa*, other two markers of trophoblast lineage were not detected. Analysis at the population level showed a modest tendency of activation of trophoblast stem cell programme, in agreement with the low percentage of Venus positive cells present in culture.

In order to assess if Venus positive cells specifically reflect the activation of *Elf5* and other trophoblast markers, these were sorted at day 10 of differentiation and profiled by RT-qPCR. About 0.3% Venus positive cells could be detected in the control experiment, compared to 3% in the *Oct4* knockout case (Figure 3.6E). Both of these positive populations were sorted and analysed for their expression profile using a low input cDNA preparation method adapted from single-cell Smart-seq2 protocol (Picelli, Faridani et al. 2014), as described in the methods section (2.2.4.2). Importantly, Venus negative and ESC wild-type controls were prepared with the same cell numbers and overall protocol as Venus positive cells for suitable comparison of the results. RT-qPCR analysis (Figure 3.6F) demonstrated that Venus positive cells clearly upregulated *Cdx2*, *Elf5* and *Tpbpa* trophoblast markers, in comparison with Venus negative or wild-type ESC control. *Eomes* did not seem to be specifically upregulated as its expression levels are similar between negative and positive populations, and between untransfected control and *Oct4* knockout cells. Critically, the small percentage of Venus positive cells observed in

untransfected control, displayed an expression profile consistent with the same cells resulting from *Oct4* knockout, demonstrating that these are more likely to reflect background spontaneous differentiation rather than Venus leaky expression.

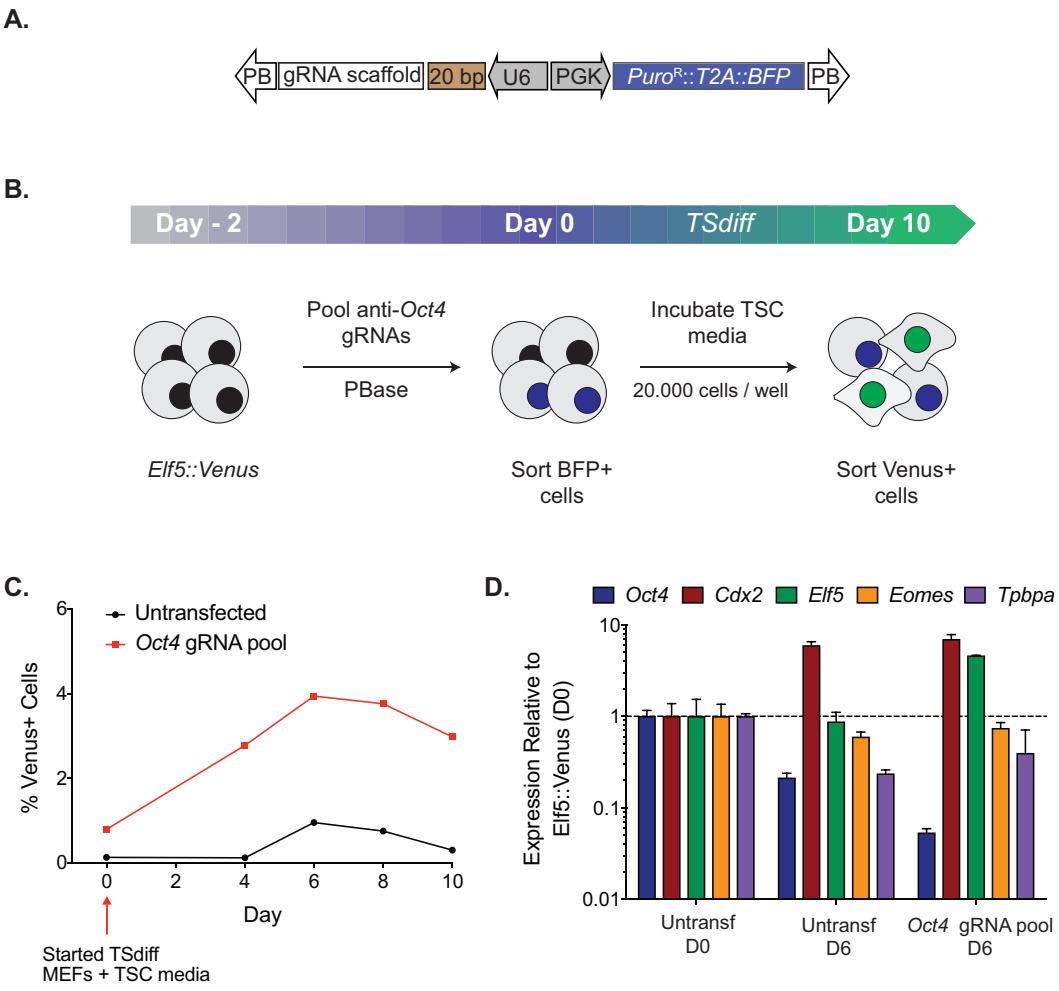
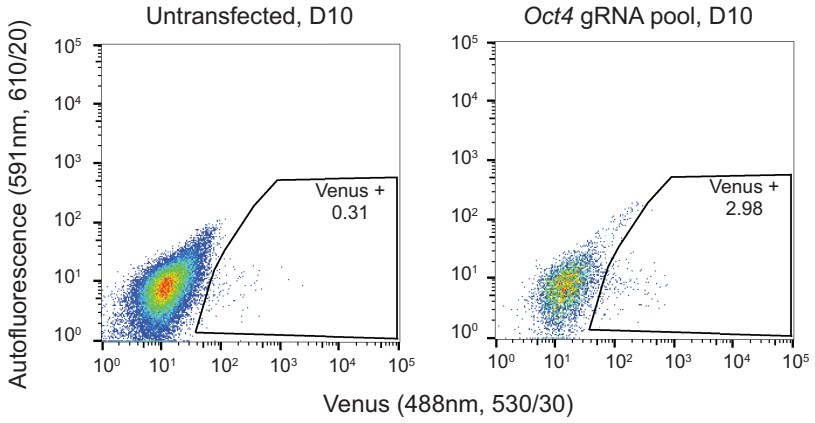


Figure 3.6 | Validation of the *Elf5::Venus* reporter ESC line, through CRISPR/Cas9-mediated *Oct4* knockout followed by induction of trophoblast differentiation. A | Schematic of the *piggyBac* vector carrying a gRNA expression cassette driven by the human *U6* promoter (*U6*) and a Puromycin-resistance gene (*Puro*^{R}) fused with a *T2A* self-cleaving peptide (*T2A*) to blue fluorescent protein (*BFP*), under the regulation of the mouse *Pgk1* promoter (*PGK*). *PB*, *piggyBac* repeats. 20 bp, gRNA specificity sequence. *bpA*, bovine polyadenylation signal sequence. **B |** Diagram depicting the experimental setup. *Elf5::Venus* ESCs were transfected at day -2 with a pool of three gRNAs targeting Oct4, and *piggyBac* transposase (PBase). 48h later, BFP expressing cells were sorted and plated for trophoblast stem cell differentiation (TSdiff) on MEFs and TSC media. At day 10, cells were sorted for Venus expression and their profile analysed by RT-qPCR. **C |** Flow cytometry analysis showing the percentage of Venus positive cells over the differentiation time-course. **D |** Bulk RT-qPCR depicting the expression profile of untransfected or *Oct4* gRNA transfected cells at the peak of Venus expression, day 6 (D6). Results are shown for *Oct4*, and four markers of trophoblast lineage (*Cdx2, Elf5, Eomes and Tpbpa*). Expression is normalized to *Gapdh*, and relative to *Elf5::Venus* at day 0 (D0).

E.



F.

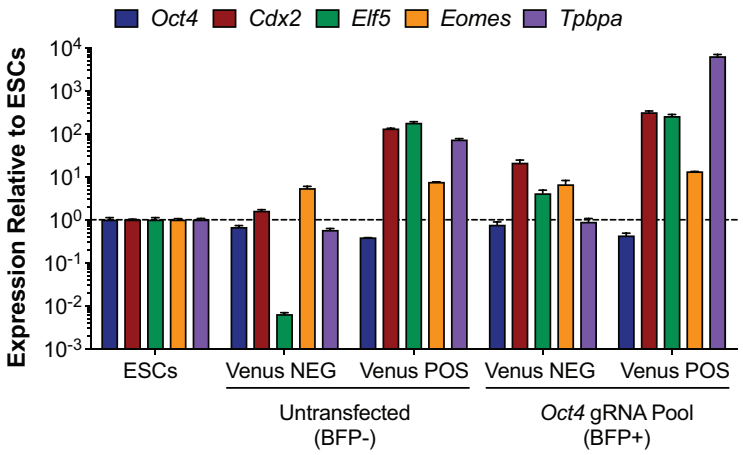


Figure 3.6 (continuation) | Validation of the *Elf5::Venus* reporter ESC line, through CRISPR/Cas9-mediated *Oct4* knockout followed by induction of trophoblast differentiation. E | Flow cytometry analysis of differentiated cells at day 10 (D10). **F |** At D10, cells were sorted for Venus and its expression profile analysed by RT-qPCR. Expression is normalized to *Gapdh*, and relative to ESC line.

3.2.2 Proof-of-concept for the use of CRISPR/Cas9 in genome-wide knockout screening for TSC differentiation: *Oct4* knockout and optimisation experiments

In this chapter I discuss the development of tools and experimental conditions required to perform a genome wide CRISPR/Cas9 knockout screen. This type of pooled screening relies on the use of lentiviral gRNA libraries that are transduced at a low multiplicity of infection (MOI) for the delivery of ideally one single gRNA per cell. The cell

knockout libraries thus generated are then submitted to a phenotypic pressure and the gRNAs present in selected cells are sequenced in order to identify enriched or depleted gRNAs and this way infer genes involved in the phenotype.

Proof-of-concept experiments should mimic the experimental setup to be taken in a screening context, that is, in a positive control, delivery of a single gRNA through lentiviral transduction needs to be sufficient to allow differentiation towards TSC lineage. Also, given that in a screen only a few genes are expected to allow differentiation, the relevant clone/cell line must be capable of eliciting a phenotype in a pool of cells that are not doing so. In other words, it must behave cell autonomously and not be subject to signals from adjacent cells in the culture.

Considering the reasons stated in the previous section for the choice of *Oct4* knockout as a model for induction of TSC differentiation, I first set to design six different gRNAs targeting exon 1 to select for efficient gRNAs capable of mediating high indel frequencies in the *Oct4* locus (Figure 3.7A). These were cloned in the gRNA expression plasmid described in Figure 3.6A and transfected into *Elf5::Venus* ESCs together with PBase for stable integration. Transfected cells were selected with Puromycin and six days post-transfection, genomic DNA was extracted for indel analysis. To determine gRNA cutting efficiency, the region surrounding the cutting site was PCR amplified (schematic representation on Figure 3.7A) and first analysed by surveyor assay. This assay relies on the use of Surveyor Endonuclease, a mismatch-specific DNA endonuclease, which cleaves at the 3' side of any mismatch with high specificity and in both DNA strands (Qiu, Shandilya et al. 2004). This makes it a useful tool to quickly assess indel frequency mediated by CRISPR/Cas9 editing as simple PCR amplification of the gRNA target region followed by denaturing and re-hybridization of the PCR product before incubation with Surveyor endonuclease should allow the detection of mismatches produced according to indel frequency in the population visualised by cleavage of the initial PCR product. The results of this assay (Figure 3.7B) demonstrate that all gRNAs can induce indels in the *Oct4* locus, detected by the presence of smaller bands upon incubation with Surveyor enzyme. To gain a quantitative insight on cutting efficiency per gRNA, the PCR products were also sequenced and analysed by TIDE (Tracking of Indels by Decomposition) (Brinkman, Chen et al. 2014). A typical Sanger sequencing chromatogram obtained from cells treated with CRISPR/Cas9 is marked by mixed traces surrounding the gRNA cutting site, as a consequence of NHEJ and the presence of indels in the edited population (Figure 3.7C). For TIDE analysis, wild-type control and sample sequence traces are submitted to an online analysis software (https://tide.deskgen.com/) which deconvolutes the mixed traces present on edited

population, reporting discrete indel frequencies in the overall sample and an estimation of cutting efficiency. TIDE analysis of the different *Oct4* gRNAs revealed that gRNAs #1 and #7 display the highest cutting efficiency (Figure 3.7D).

As the kinetics of PBase-mediated integration is different from lentiviral integration, I next performed a time-course experiment in order to understand cutting dynamics using lentivirus. I produced lentiviruses for the independent delivery of gRNAs #1 and #7 and transduced *Elf5::Venus* ESCs with these, approximately at MOI of 0.3. Cells were selected with Puromycin from day 2 post transduction. I then extracted genomic DNA at different days and analysed the indel complexity by TIDE. Results demonstrate that cutting is detected from day 3, and increases over time, with the highest cutting efficiency visible at day 8 (Figure 3.7E).

Standard conditions for TSC maintenance require a MEF feeder layer, or MEF-conditioned TSC media. TSC media is widely established in the scientific community as it was used to derive the first TSC lines from mouse blastocysts (Tanaka, Kunath et al. 1998). It contains serum and provides the cytokine Fgf4 and cofactor Heparin to promote TSC proliferation. TX media is a serum-free defined formulation developed more recently for derivation and maintenance of TSCs (Kubaczka, Senner et al. 2014) and in addition to Fgf4 and Heparin, it relies on the cytokine Tgf-β1 which replaces the need for MEF-conditioned media (Erlebacher, Price et al. 2004).

Having determined efficient gRNAs to induce *Oct4* knockout, and the kinetics of lentiviral-mediated editing, I next started proof-of-concept experiments to assess the capacity for the mutated population to differentiate towards TSCs. Additionally, I designed the experiments to test how different plating densities and culturing media might affect differentiation efficiency. I chose to test both TSC media and TX media. Moreover, in this first round, I tested plating the pool of mutated cells for differentiation at low density, according to most published protocols, or increasing 10-fold this initial cell number. This is an important parameter given that in a genome-wide screening context, millions of cells should be used for differentiation to guarantee library representation, and therefore higher plating density becomes a valuable technical advantage provided it is not detrimental to differentiation efficiency.

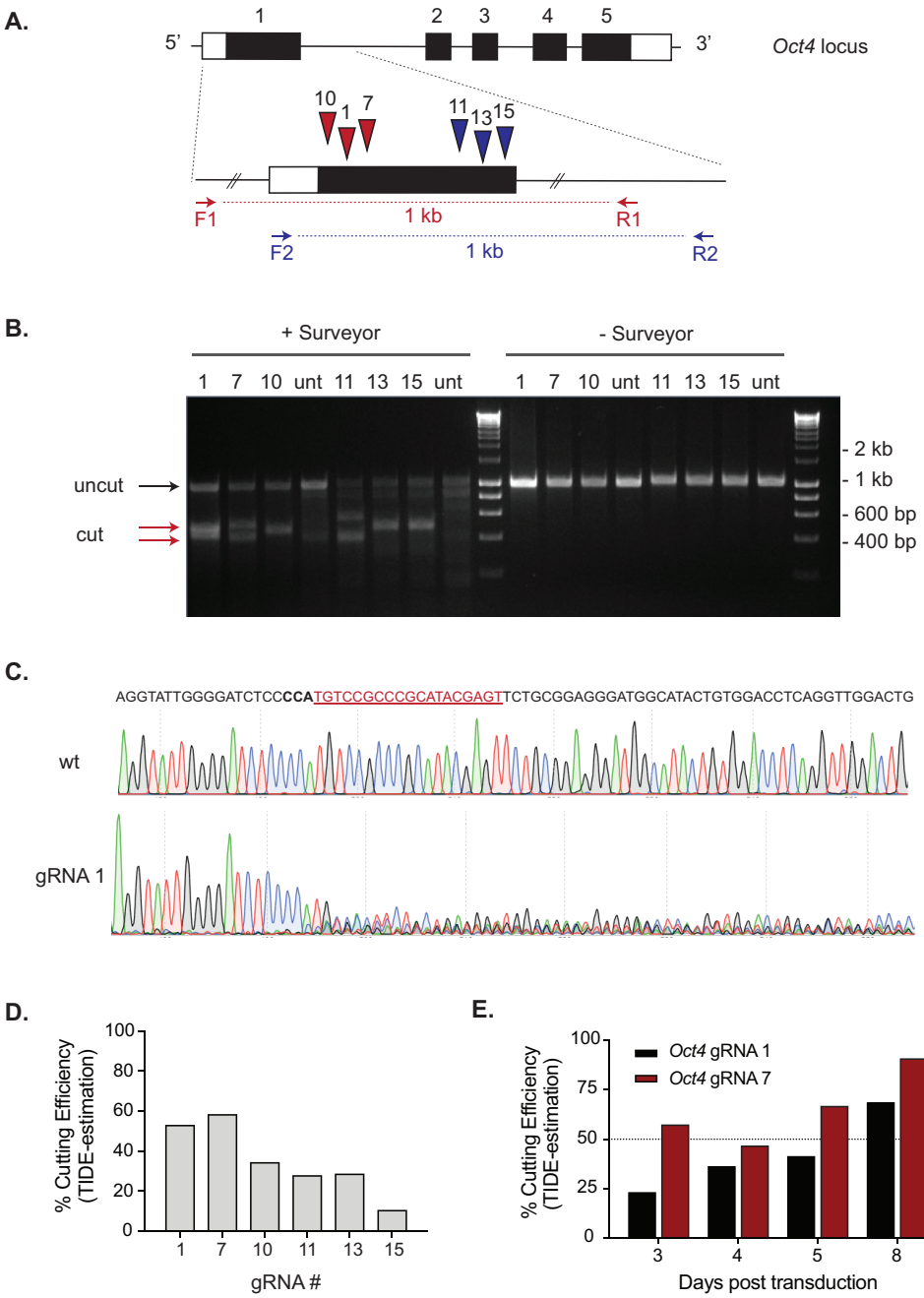


Figure 3.7 | Selection of gRNAs to efficiently target the *Oct4* locus in ESCs. A | Schematic representation of the mouse *Oct4* locus and gRNA design. Six gRNAs (denoted by arrow heads) targeting different sequences in exon 1 were characterized for their cutting efficiency 6 days post lipofection. For indel quantification, a 1kb region surrounding the target sequence was amplified by PCR (primers F1 and R1 for gRNAs number 1, 7 and 10; and primers F2 and R2 for gRNAs number 11, 13 and 15). **B |** Surveyor assay depicting gRNA efficiency. Unt, untransfected control. **C |** Representative Sanger sequencing chromatogram demonstrating the presence of mixed traces upon gRNA 1 transfection. Highlighted in red is the gRNA 1 targeting sequence, with the PAM region in bold. The top chromatogram represents the wild-type (wt) sequence, whereas the bottom one shows gRNA 1-edited cells. **D |** TIDE-estimation of cutting efficiency 6 days post transfection. **E |** Cutting efficiency of gRNAs 1 and 7 over time, estimated by TIDE analysis.

Elf5::Venus ESCs were transduced with *Oct4* gRNA #1 or a *U6* empty vector control. A mock control was also included in this first lentiviral experiment to measure transduction effects on background spontaneous differentiation. Transduced cells were selected with puromycin. Seven days post-transduction (differentiation day zero), cells were plated at 20.000 or 200.000 cells per well of a 12-well plate, on MEF feeders, and cultured on TSC media or TX media for 14 days (Figure 3.8A). To test the effect of editing on Oct4 protein levels at the start of differentiation, part of the culture was fixed and stained with anti-Oct4 antibody followed by flow cytometry analysis. The results demonstrate that in comparison with mock or *U6* controls, there was a shift towards lower levels of protein expression upon editing with *Oct4* gRNA #1 (Figure 3.8B). I next analysed indel complexity in the edited ESC culture at day zero of differentiation (corresponding to P2 post-transduction) and one passage later. TIDE estimation determined a cutting efficiency of 52.6% by P2 and an increase to 63.1% by P3 (Figure 3.8C). Further analysis of indel profile demonstrated that despite high cutting efficiency, most of the edited population by P2 displayed deletions of 6, 9 or 12 bp generating *in-frame* mutations, with only about 7.6% *frame-shifting* mutations present. Additionally, by P3 the culture was taken over by cells with *in-frame* deletions.

TSC differentiation was monitored using flow cytometry to measure percentage of Venus positive cells over time. Both mock and *U6* controls showed equivalent dynamics, with very low percentage of Venus cells by day 14 in any of the conditions tested (Figure 3.8D and E), demonstrating that transduction does not affect background differentiation significantly. On the other hand, despite the low percentage of *frame-shifting* mutations by day zero, differentiating *Oct4* knockout cells in TSC media yielded an increasing Venus expression over time, culminating in about 20% Venus positive cells by day 14 when plating 20.000 cells per well, and 30% when starting with 200.000 cells (Figure 3.8D). The same differentiations in TX media, resulted in lower efficiencies (Figure 3.8E), with 20.000 cells per well providing better results. Moreover, TX media induced high levels of cell death in culture.

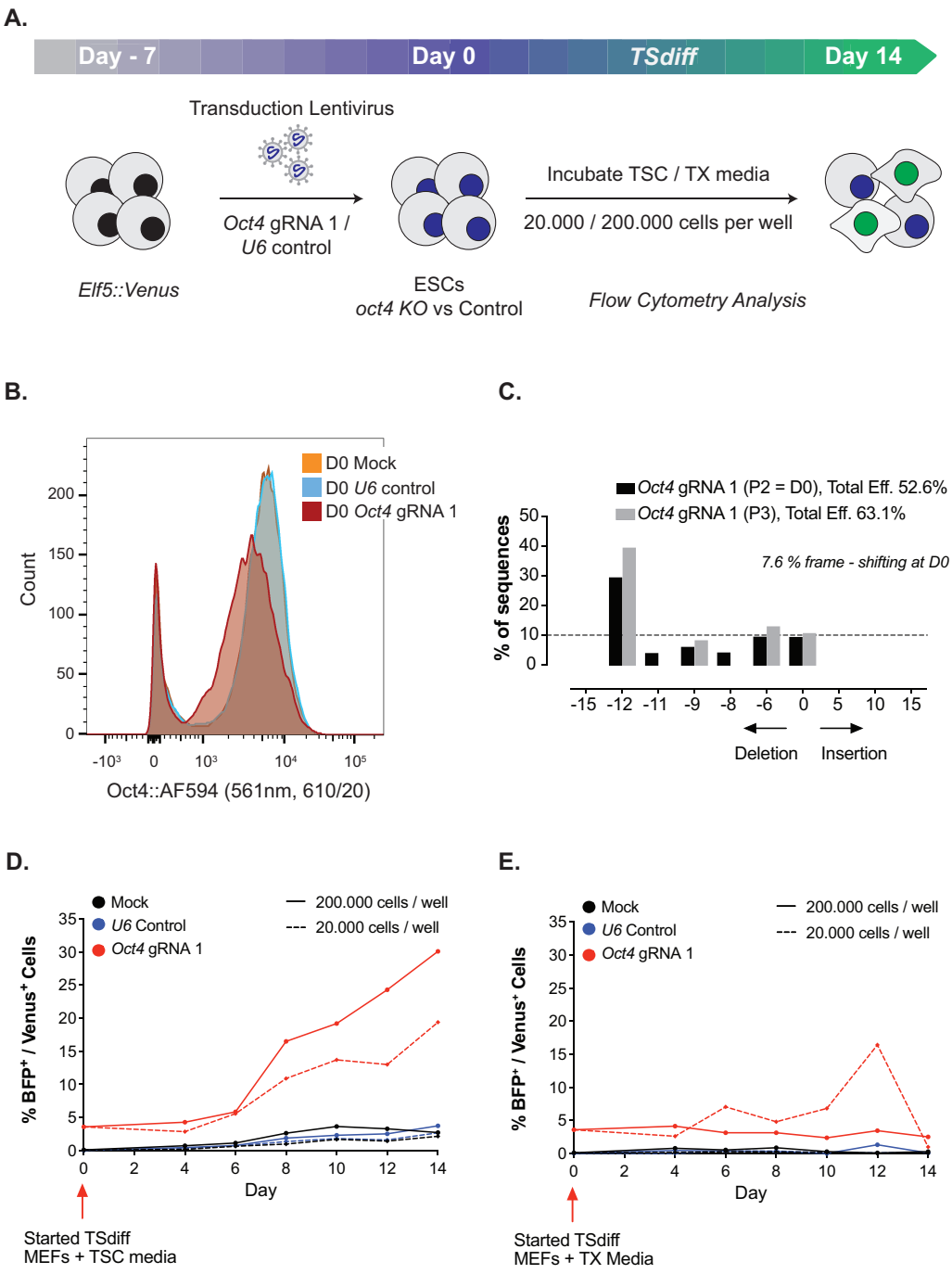


Figure 3.8 | Using lentiviral delivery of gRNA against *Oct4* allows for differentiation towards trophoblast lineage. **A |** Experimental setup. Cells were transduced with lentivirus containing control gRNA or *Oct4* gRNA 1. Seven days later, cells were plated in the indicated media and density for differentiation. The percentage of Venus positive cells was assessed by flow cytometry for 14 days. **B |** Flow cytometry analysis of cells at D0, stained for Oct4 protein levels. **C |** TIDE estimation of indel profile of cells transduced with *Oct4* gRNA 1 at D0 of differentiation (Passage number 2 post-transduction, P2) and at one passage later (P3). **D and E |** Percentage of Venus positive cells over the differentiation time-course in TSC media (D) and TX media (E), analysed by flow cytometry. The gating strategies for analysis of Venus positive populations in D | and E | can be found on **Supplementary Figure 3.1** and **Supplementary Figure 3.2**, respectively, at the end of this chapter.

Considering that by day seven post-transduction, I could only detect 7.6% *frame-shifting* mutations in the *Oct4* knockout pool, and that one passage seemed to have a drastic effect on indel profiles, I next examined whether initiating differentiation at an earlier time post-transduction could capture a higher degree of *frame-shifting* mutations and improve differentiation yields.

Elf5::Venus ESCs were transduced with *U6* control and *Oct4* gRNA#1, as previously, and selected with puromycin. Three days post-transduction, cells were plated for differentiation again in either TSC media or TX media on MEF feeders. Plating cell density was 200.000 or 400.000 cells per well of a 12 well plate and differentiation carried on for 14 days (Figure 3.9A). TIDE analysis of indel profile by day zero of differentiation confirmed the expected result, and one could detect 43% cutting efficiency, with 30.4% *frame-shifting* indels (Figure 3.9B). However, analysis of Oct4 protein levels by immunostaining and flow cytometry did not show lower expression compared to control untransfected cells (Figure 3.9C). Differentiation in TSC media showed again an increase in Venus percentage over time, with little difference between the two plating densities tested (Figure 3.9D). The overall efficiency was nevertheless about half the previous experiment. On the other hand, efficiency of differentiation in TX media was greatly improved in this setting, displaying an increasing Venus positive population over time and reaching about 40% by day 14 when plating the cells at 200.000 cells per well (Figure 3.9E). In this case, increasing plating density clearly had a detrimental effect on differentiation as seen by the curve for 400.000 cells per well (Figure 3.9E). Nevertheless, TX media seems to either be highly toxic or highly selective as most cells in culture die during differentiation. High levels of cell death could also be detected in TSC media due to the confluency each well reaches over time, but by the end of differentiation I could count about 240.000 cells per well in TX media, as opposed to about 2.3 million cells in TSC media.

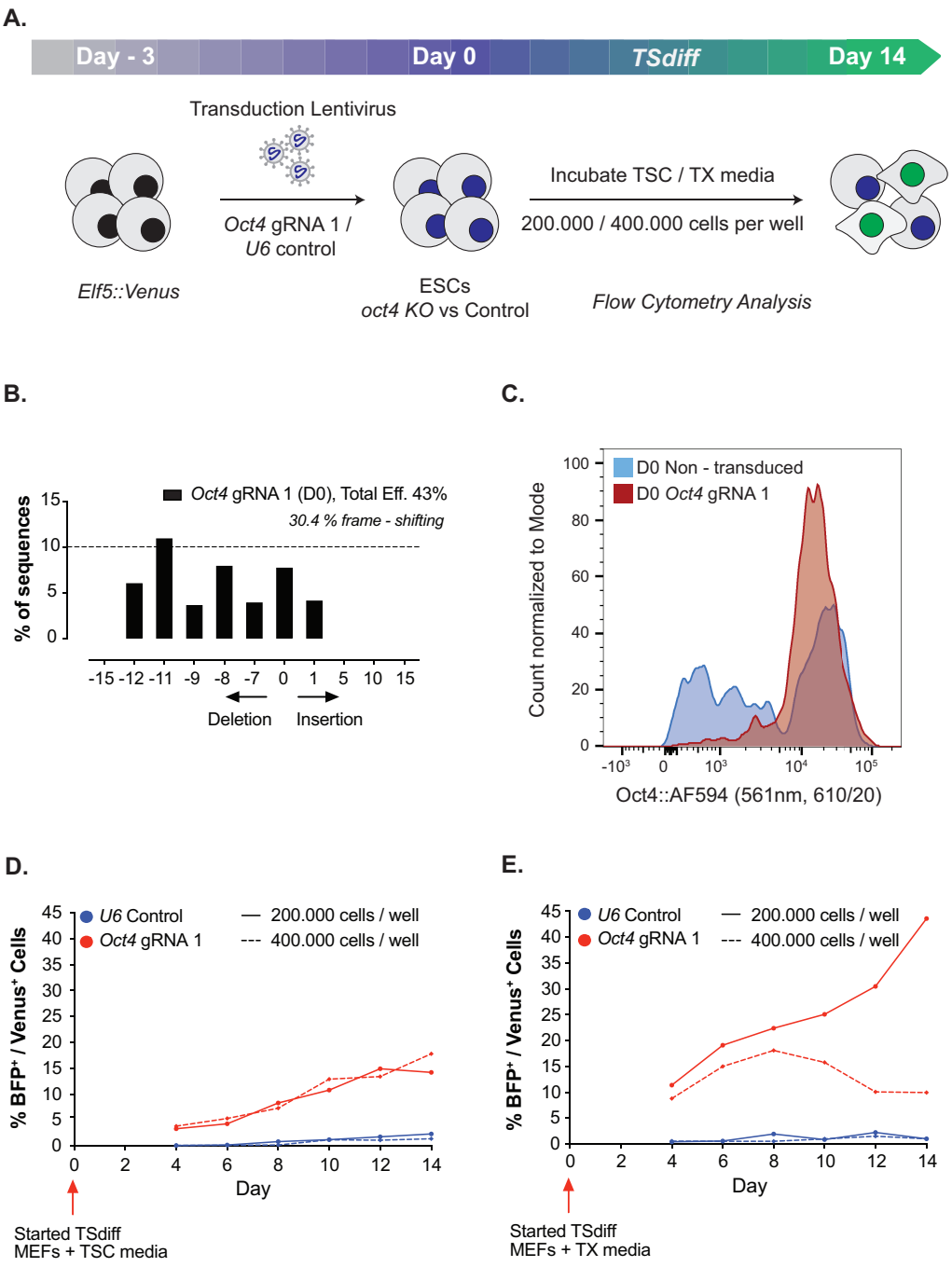


Figure 3.9 | Initiating differentiation 3 days post-transduction captures a higher percentage of frame-shifting mutations of *Oct4* but does not increase trophoblast differentiation efficiency. A | Experimental setup. Cells were transduced with lentivirus containing control gRNA or *Oct4* gRNA 1. Three days later, cells were plated in the indicated media and density for differentiation. The percentage of Venus positive cells was assessed by flow cytometry for 14 days. **B |** TIDE estimation of indel profile of cells transduced with *Oct4* gRNA 1 at D0 of differentiation. **C |** Flow cytometry analysis of cells at D0, stained for Oct4 protein levels. **D and E |** Percentage of Venus positive cells over the differentiation time-course in TSC media (D) and TX media (E), analysed by flow cytometry. The gating strategies for analysis of Venus positive populations in D | and E | can be found on **Supplementary Figure 3.1** and **Supplementary Figure 3.2**, respectively, at the end of this chapter.

To evaluate whether Elf5 protein can be detected in cells expressing Venus, I sorted these cells at day 14 of differentiation and cytospun them to poly-lysine coated slides followed by immunostaining for Elf5. This confirmed that Elf5 is co-localised with Venus in individual cells (Figure 3.10A).

I have also examined Cdx2 protein expression through immunostaining by day 10 of differentiation is TS media, in cells transduced with either *U6* control or *Oct4* gRNA #1. These results demonstrate that Cdx2 can be readily detected in the control experiment whereas clear Venus positive cells are difficult to find (Figure 3.10B). On the other hand, knocking out *Oct4* clearly increased the number of cells expressing Cdx2 (Figure 3.10C), keeping Venus positive cells at low levels. Importantly, Venus positive cells co-express Cdx2.

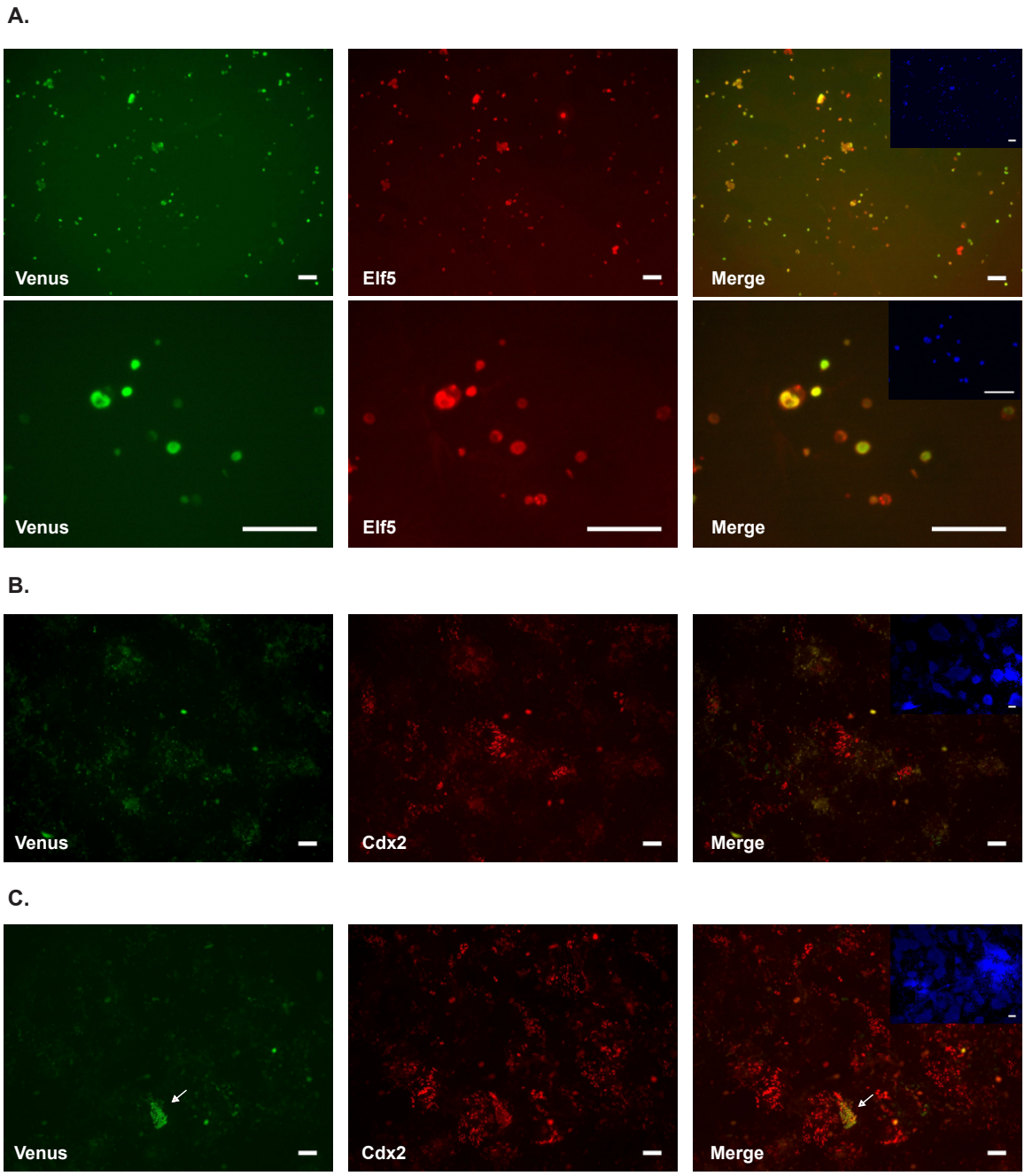


Figure 3.10 | Immunostaining of differentiated cells confirms co-expression of Venus and TSC markers Elf5 and Cdx2. A | Venus positive cells resulting from differentiation upon *Oct4* knockout and culture in TSC media were sorted at day 14, cytospun to poly-lysine coated slides, and stained for Elf5 expression. **B and C |** Differentiating cells were fixed at day 10 and stained for Cdx2 expression for both *U6* control (B) and *Oct4* knockout (C). Note the high level of autofluorescence in the green channel due to extensive cell death in the cultures. Arrow denotes a Venus positive colony co-staining with Cdx2. Scale bar, 100 μm.

3.2.3 Using *Dnmt1* knockout as a second case-study for TSC differentiation

Oct4 is a gene essential for ESC self-renewal (Niwa, Miyazaki et al. 2000). As a consequence, it has strong effects on the indel profiles that could be detected on a mixed population resulting from CRISPR/Cas9 editing, as shown in subchapter 3.2.2. In order to try to maximise the type of genes I might pick from the genome-wide screening setup I chose to perform, I next decided to test knockout of *Dnmt1* to get a different profile for editing and differentiation. *Dnmt1* null mutations do not affect self-renewal in ESCs (Lei, Oh et al. 1996). Furthermore, it has been implicated in the regulation of the epigenetic memory of the first cell lineage decision, as ESCs homozygous for *Dnmt1* mutation can differentiate to TSC (Ng, Dean et al. 2008). Importantly, this was the seminal work revealing the role of *Elf5* in the epigenetic restriction of cell fate.

As for the *Oct4* case, I designed and evaluated gRNAs to mediate efficient *Dnmt1* knockout. *Dnmt1* is a large locus and published work relies on replacing or disrupting its catalytic domain to affect protein function (Lei, Oh et al. 1996, Tsumura, Hayakawa et al. 2006). As *Dnmt1* is a gene composed of 40 exons, I designed ten gRNAs targeting different regions of the gene, with a special focus towards the catalytic domain (Figure 3.11A). I have also included gRNAs #9 and #10, which target exon 5 and were validated previously (Sanjana, Shalem et al. 2014). gRNA cutting efficiency was assessed by transfection of *Elf5::Venus* ESCs with each gRNA and PBase followed by puromycin selection for 6 days. At this point, genomic DNA was extracted and the regions surrounding each gRNA target site amplified by PCR (schematic representation Figure 3.11A). The surveyor assay demonstrated that all gRNAs can produce indels at the desired locus (Figure 3.11B). Cutting efficiency was then evaluated by TIDE estimation revealing a range of efficiencies from 40 – 80% for most gRNAs apart from #6 and #7 (Figure 3.11C).

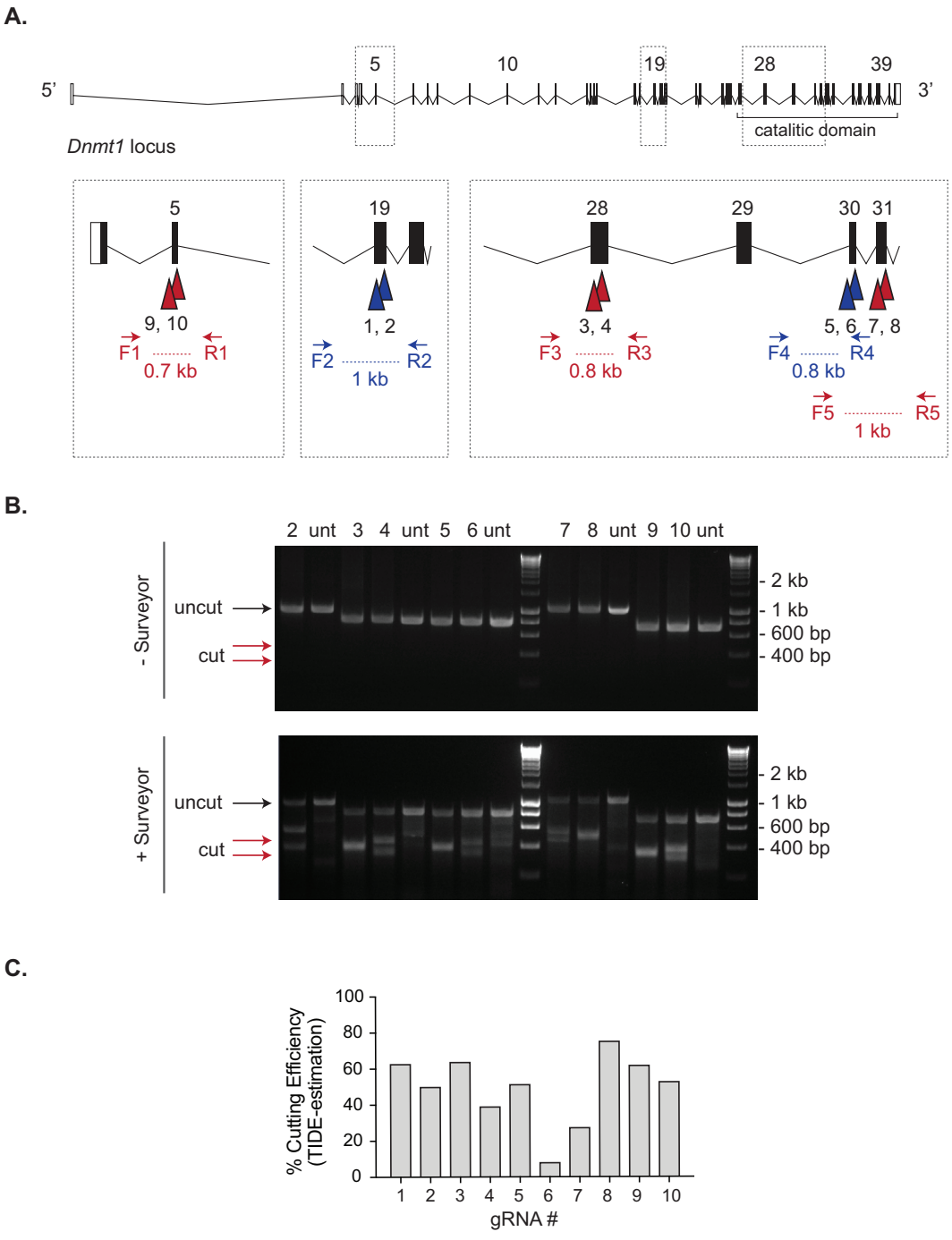


Figure 3.11 | Selection of gRNAs to efficiently target the *Dnmt1* locus in ESCs. A | Schematic representation of the mouse *Dnmt1* locus and gRNA design. Ten gRNAs (denoted by arrow heads) targeting different sequences throughout the gene were characterized for their cutting efficiency 6 days post lipofection. For indel quantification, a 0.7 – 1 kb region surrounding the target sequence was amplified by PCR (primers F1 and R1 for gRNAs 9 and 10; F2 and R2 for gRNAs 1 and 2; F3 and R3 for gRNAs 3 and 4; F4 and R4 for gRNAs 5 and 6; and F5 and R5 for gRNAs 7 and 8). **B |** Surveyor assay depicting gRNA efficiency. Unt, untransfected control. **C |** TIDE-estimation of cutting efficiency 6 days post transfection.

I next proceeded with lentiviral experiments in a similar setting as used for *Oct4*. Cells were transduced with either *U6* control or *Dnmt1* gRNA #3 followed by selection with puromycin. Seven days post-transduction they were seeded for differentiation in 12 well plates with MEF feeders at a plating density of 200.000 or 400.000 cells per well, and cultured for 14 days in either TSC or TX media (Figure 3.12A). The effect of *Dnmt1* knockout on Dnmt1 protein expression by day zero of differentiation was assessed by immunostaining followed by flow cytometry analysis. A clear reduction on Dnmt1 expression levels could be detected in cells transduced with *Dnmt1* gRNA#3 compared with *U6* control (Figure 3.12B). TIDE estimation of cutting efficiency at different passage number post-transduction (TSC differentiation was started at P3), showed stable percentage of edited ESCs with 76.5% cutting efficiency by P2 and P3 and a lower level by P5, 62% (Figure 3.12C). Importantly, indel diversity could be maintained over time, with a consistent percentage of *frame-shifting* mutations. Nevertheless, the slow increase of wild-type sequences, reaching about 10% by P5, might indicate a growth advantage of cells with *wild-type Dnmt1* sequences compared with those with mutated *Dnmt1*.

Analysis of the presence of Venus positive cells over the differentiation time-course demonstrated that despite achieving high mutation rates and protein knockdown, this was not sufficient to allow differentiation in any of the conditions tested (Figure 3.12D and E). I have additionally tested the use of lentiviral delivery of gRNA #8, another highly efficient gRNA for *Dnmt1* editing, followed by initiation of differentiation at an earlier timepoint post-transduction, with the same negative outcome (data not shown). Together, these experiments indicate that *Dnmt1* knockout might not allow TSC differentiation in an equivalent setting to *Oct4* knockout.

Dnmt1 phenotype in ESC to TSC differentiation was shown for a homozygous mutant (Ng, Dean et al. 2008). I therefore decided to run a final test using co-transfection of gRNAs and PBase to aim for multiple gRNA integrations and this way try to achieve higher mutation rates. I transfected ESCs with all the gRNAs shown to be effective for *Dnmt1* knockout (individually) or two pools of five gRNAs each. I then selected with puromycin and six days later seeded mutated ESCs for differentiation using an experimental design equivalent to published work (Ng, Dean et al. 2008): low plating density (25.000 cells per well of a 12 well plate) and culture in TSC media for 12 days (Figure 3.13A). The levels of *Dnmt1* knockdown were measured by RT-qPCR demonstrating a significant downregulation in all gRNA transfections, compared with untransfected control (Figure 3.13B). Once again, analysis of the percentage of Venus positive cells over the differentiation time-course maintained a level similar to

spontaneous differentiation observed in untransfected cells, even when pools of gRNAs were used to induce mutation.

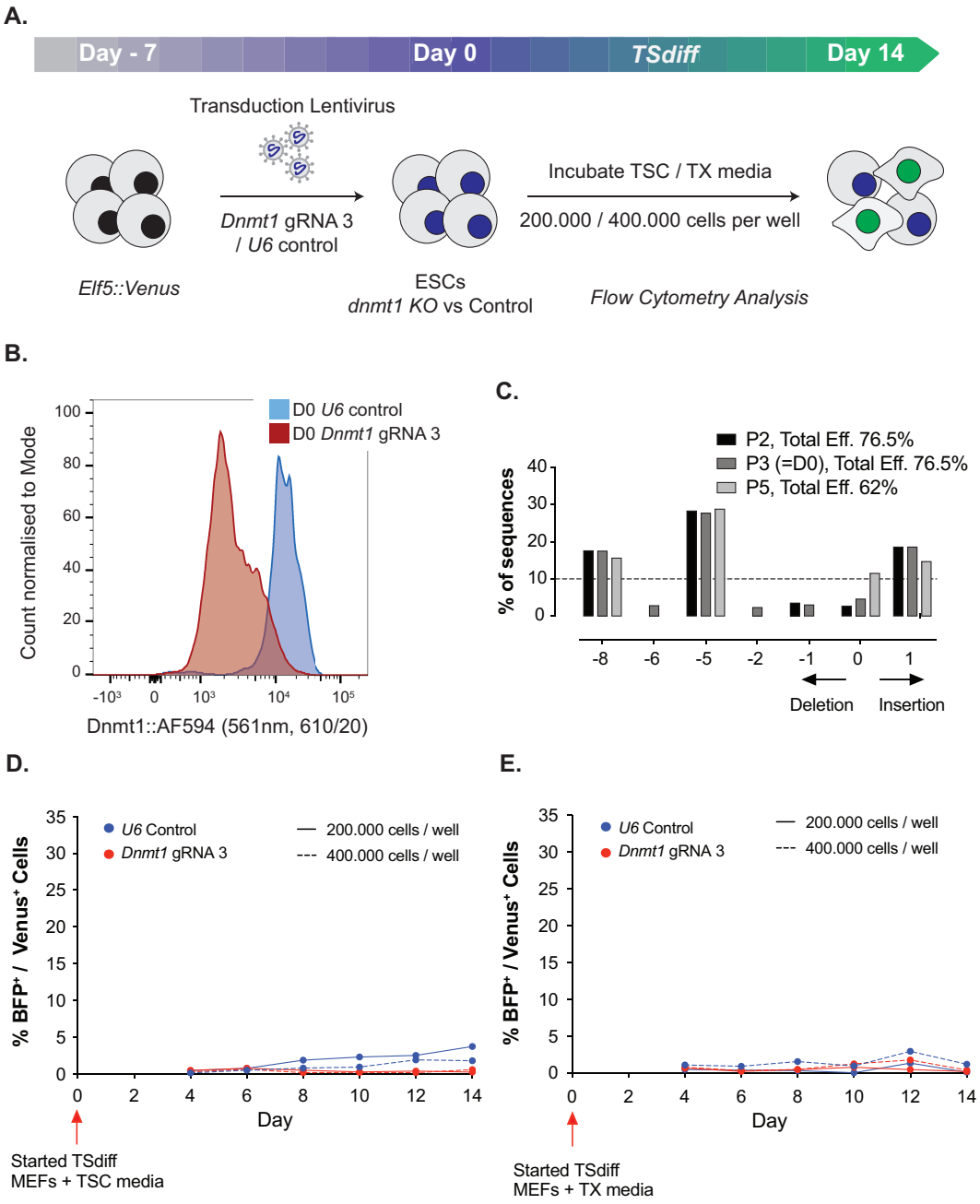


Figure 3.12 | Knockout of *Dnmt1* through CRISPR/Cas9 does not allow for differentiation towards trophoblast lineage in the same setting as *Oct4* knockout. A | Experimental setup. Cells were transduced with lentivirus containing control gRNA or *Dnmt1* gRNA 3. Seven days later, cells were plated in the indicated media and density for differentiation. The percentage of Venus positive cells was assessed by flow cytometry for 14 days. B | Flow cytometry analysis of cells at D0, stained for Dnmt1 protein levels. C | TIDE estimation of indel profile of cells transduced with *Dnmt1* gRNA 3 at different passage number post-transduction. Differentiation was started at P3. D and E | Percentage

of Venus positive cells over the differentiation time-course in TSC media (D) and TX media (E), analysed by flow cytometry.

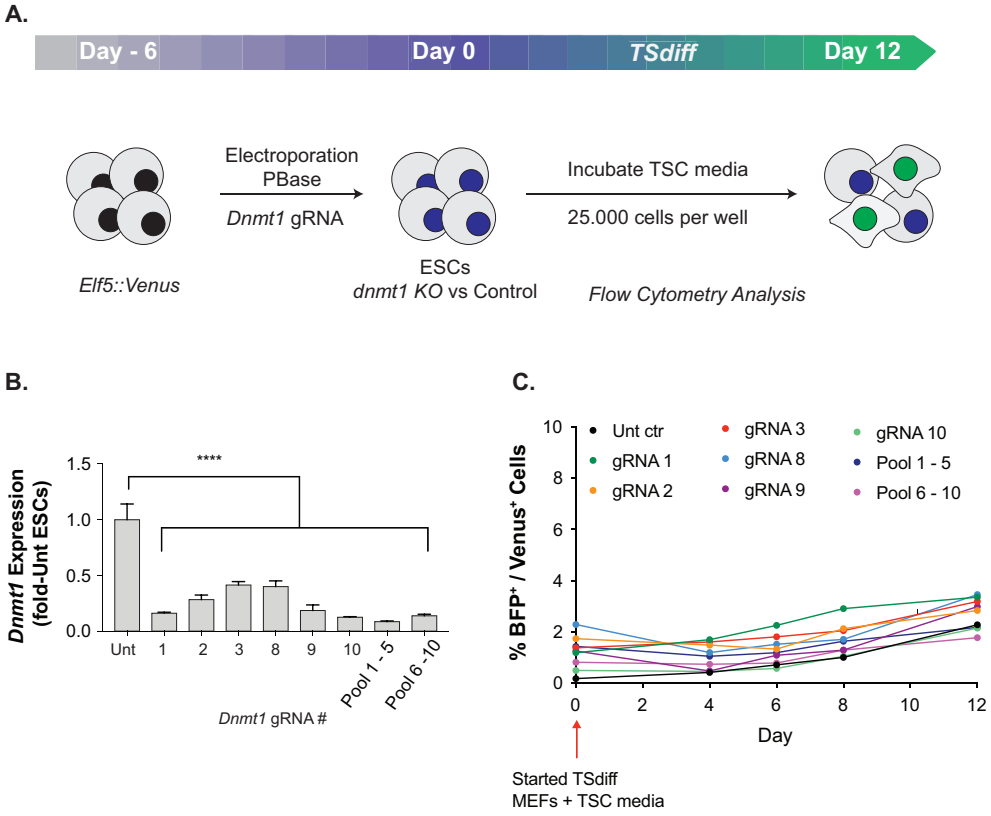


Figure 3.13 | Knocking out *Dnmt1* with a different range of gRNAs does not result in improved differentiation capacity. A | Experimental setup. Cells were transfected with PBase and different gRNAs targeting *Dnmt1*. Six days later, cells were plated in TSC media at low density for differentiation. The percentage of Venus positive cells was assessed by flow cytometry for 12 days. **B |** RT-qPCR analysis of *Dnmt1* expression levels at D0 of differentiation. Results were compared to untransfected control (Unt) and analysed using one-way ANOVA with Dunnett's multiple comparison test. **** = adjusted p-value < 0.0001. **C |** Percentage of Venus positive cells over the differentiation time-course, analysed by flow cytometry.

3.3 Discussion

Several groups have reported differentiating ESCs to TSC fate, as discussed in section 3.1.1. However, comparison of multiple strategies in a systematic manner has shown that most of these methods result in partial differentiation due to the epigenetic memory of the first cell lineage decision and mainly fail to efficiently activate *Elf5* transcription as a consequence of its promoter methylation status (Cambuli, Murray et al. 2014). With this current bottleneck in mind, I set out to build an *Elf5* reporter cell line, and follow the activation of this stringent marker to unveil new molecular players in the maintenance of the first cell lineage barrier. In this chapter, I describe tools and conditions to perform a genome-wide CRISPR/Cas9 knockout screen to address this question.

3.3.1 Development of *Elf5::Venus* ESCs that faithfully report TSC differentiation

I started by engineering JM8-*R26-Cas9* ESC line, which has stable integration of *Cas9* in the *Rosa26* locus and was clonally selected for high Cas9 activity (Tzelepis, Koike-Yusa et al. 2016). I successfully knocked-in *T2A::H2B::Venus* into the *Elf5* locus, using a CRISPR/Cas9-mediated homology-directed repair strategy. In this approach, the use of CRISPR/Cas9 induces a double-strand break at the 3' end of the *Elf5* coding sequence. It then exploits the endogenous homology-directed repair machinery to repair this break by recombination with homology arms on the exogenous repair template provided, thereby knocking-in the *T2A::H2B::Venus* sequence. As a quality control, I then showed edited clones had no indels present in any of the predicted exonic off-target sequences of the gRNA employed.

I next verified the integrity of *Elf5* promoter DMR regions upon editing as this is a key regulatory element preventing ESCs from adopting a trophoblast fate. I confirmed that *Elf5::Venus* ESCs maintain the levels of methylation seen in wild-type cells, reflecting an intact lineage barrier after editing and are therefore a valid system to address genetic suppressors of TSC differentiation in ESCs.

In order to completely validate the use of *Elf5::Venus* ESC as a reporter system, I next employed CRISPR/Cas9-mediated *Oct4* knockout to allow differentiation towards TSC lineage and test whether i) Venus positive cells could be detected over time, and ii) cells

expressing Venus truly reflect *Elf5* expression and the activation of a TSC transcriptional profile.

I performed this experiment with a pool of gRNAs co-transfected with PBase and followed the published protocol for differentiation (Ng, Dean et al. 2008). This setting was not the most efficient, generating low percentage of Venus positive cells, likely due to starting differentiation 48h post-transfection which is not enough time to induce high indel frequency, as shown by others (Kosicki, Rajan et al. 2017). It is also a short period to allow mRNA and protein downregulation. Nevertheless, this preliminary experiment allowed sorting and RT-qPCR profiling of Venus positive cells from both untransfected and *Oct4* knockout samples. This demonstrated Venus expressing cells have a distinct profile from Venus negative ones, and upregulate key TSC markers *Elf5*, *Cdx2* and *Tpbpa*, reflecting the activation of trophoblast transcriptional programme. I later further demonstrate faithful expression of Elf5 by sorting Venus positive cells followed by immunostaining, corroborating that individual Venus cells co-stain with Elf5 protein.

Population level RT-qPCR data from this experiment together with immunostainings of differentiating cells reinforced the choice of *Elf5* as a marker for TSC differentiation over *Cdx2*, another core transcription factor in TSC (Ralston and Rossant 2008). Indeed, analysing bulk RT-qPCR data, I could find that *Cdx2* was readily expressed in control experiments, which was later confirmed by immunostaining. This upregulation was more prominent upon *Oct4* knockout. On the other hand, *Elf5* is more stringent, and was not detected at the population level in RT-qPCR of control experiments, but could specifically be induced upon *Oct4* KO. This was consistently demonstrated in the immunostaining of differentiating cells resulting from *Oct4* KO, where the proportion of Cdx2 positive cells was higher than Venus and, importantly, Venus cells co-expressed Cdx2.

Together, data generated corroborate the suitability of *Elf5::Venus* ESCs as a useful tool to address differentiation of ESCs to TSCs.

3.3.2 Optimisation of experimental setup for CRISPR/Cas9 genome-wide screening using *Oct4* knockout model

For a genome-wide knockout screen to be possible, delivery of a single gRNA per cell needs to produce efficient gene-editing to overcome the lineage restriction that prevents ESCs from adopting a trophoblast fate, and result in detectable activation of Venus / Elf5 expression. At the other end of the screen, one should be able to retrieve

enough Venus positive cells so that genomic DNA can be extracted and analysed for gRNA enrichment. It was therefore key to have a good positive control to both prove the feasibility of the screening and optimise experimental conditions to maximise differentiation efficiency.

I chose to use *Oct4* knockout as a model for these experiments, due to its robust role on ESC to TSC differentiation. Upon defining efficient gRNAs to target this locus, I could demonstrate that lentiviral transduction at MOI 0.3 (presumed to provide an average of 1 gRNA per selected cell), is sufficient to allow differentiation of *Elf5::Venus* ESCs towards TSC lineage. This validated the feasibility of a CRISPR/Cas9 genome-wide screening.

I then explored this system for optimisation experiments in order to define the best conditions for TSC differentiation. Several variables were considered and are discussed below.

1. Plating density

In a genome-wide screening context it is important to guarantee high library coverage, with a recommended guideline of about 500 cells per gRNA (Joung, Konermann et al. 2017). Considering a library of about 90.000 gRNAs (Tzelepis, Koike-Yusa et al. 2016), achieving a coverage of 500x means having to differentiate 45 million cells per screen. Most of the published protocols for ESC to TSC differentiation rely on low plating density (Niwa, Toyooka et al. 2005, Ng, Dean et al. 2008, Abad, Mosteiro et al. 2013, Cambuli, Murray et al. 2014). My results demonstrate that increasing plating density to 200.000 cells per well had no adverse effects on the percentage of Venus positive cells detected. When testing an increased cell density to 400.000 cells per well, differentiation in TX media was unreliable. I decided to carry on with 200.000 cells per well for the genome-wide screening setup.

2. Medium used to induce differentiation

There are different culture conditions for the maintenance of TSCs (Tanaka, Kunath et al. 1998, Erlebacher, Price et al. 2004, Kubaczka, Senner et al. 2014, Ohinata and Tsukiyama 2014). Of these, I focused on media formulations according to three criteria: i) they allow derivation of TSC lines from mouse embryos; ii) TSCs maintained on these have demonstrated *in vivo* contribution to placenta; iii) they contain small number of cytokines and small molecules. This last point was included given that complex media

formulations could result on the screening picking up genes involved in specific response to media components rather than their role on lineage restriction.

This way, I have tested the use of TSC medium (Tanaka, Kunath et al. 1998) and TX medium (Kubaczka, Senner et al. 2014). TSC medium revealed relatively consistent results between the different experimental settings tested, demonstrating a consistent increase in Venus population overtime. It allowed efficiencies of 15 - 30 % Venus positive cells by day 14 independently of the cell density or timing post-transfection used for induction of differentiation.

On the other hand, TX media had dramatic differences in efficiencies depending on cell density and the starting point for differentiation. Initiating differentiation 7 days post-transduction resulted in poor differentiation, with a peak of Venus expression by day 12, which was lost by day 14. Starting differentiation 3 days post-transduction could result in 40% Venus positive cells by day 14 when plating 200.000 cells per well, compared to a decrease to 10% if plating density was doubled.

Even though TX media seemed to allow high differentiation efficiency in one of the conditions tested, this formulation was always accompanied by massive cell death in culture. In fact, counting cell numbers by day 14 of differentiation demonstrated that TX medium cultures had about 240.000 cells per well, compared to 2.3 million cells per well in TSC medium, which would prove challenging to yield enough Venus positive cells for analysis in a genome-wide screening setting.

Considering the consistency between experiments with TSC medium, and high cell survival through differentiation, it has the best formulation for differentiation towards TSC.

3. Timing for induction of TSC differentiation

Adjusting the timing post-transduction for initiation of differentiation was the final parameter considered. *Oct4* knockout turned out to be an interesting case-study for this due to its effect on ESC self-renewal. On one hand, before differentiation, one should allow enough time for lentiviral integration of the gRNA cassette, gRNA expression and editing of the *Oct4* locus. On the other hand, given the role of *Oct4* in self-renewal, each ESC passage reduced the complexity of indels present in the population and quickly cultures would be overgrown by cells with *in-frame* mutations. Indeed, starting differentiation seven days post-transduction identified just 7.6% *frame-shifting* mutations that resulted in 30% Venus positive cells by day 14 of differentiation in TSC media. Instead, starting differentiation three days post-transduction captured higher indel diversity in culture, with 30.4% *frame-shifting* mutations, but yielded only about 15% of

Venus positive cells in the same condition. This was likely correlated with protein turn-over as Oct4 immunostaining did not detect downregulation of Oct4 levels in this case.

Given the *Oct4* example, and in order to be more likely to detect higher indel diversity on a genome-wide screening experiment, I concluded that the best approach would be to start replicates covering different timepoints post-transduction.

3.3.3 Limitations in CRISPR/Cas9 screening: the *Dnmt1* knockout scenario

To maximise the type of genes a genome-wide knockout screen could pick up, I decided to also test *Dnmt1*, which is not necessary for ESC self-renewal (Lei, Oh et al. 1996, Tsumura, Hayakawa et al. 2006) and allows differentiation to TSC (Ng, Dean et al. 2008).

I identified efficient gRNAs which target *Dnmt1*. TIDE estimations reached 60 – 80% cutting efficiency for most of the gRNAs tested. Dnmt1 immunostaining followed by cytometry analysis demonstrated a clear reduction in knockout experiments, compared to control ESCs and the data was further corroborated through RT-qPCR. Nevertheless, none of the experimental setups tested for differentiation resulted in significant expression of Venus positive cells demonstrating a failure in TSC differentiation, compared with the *Oct4* knockout case.

This absence of differentiation towards TSC might be related with the level of *Dnmt1* knockout achieved through CRISPR/Cas9, as the original work was developed in an ESC line with a *Dnmt1* null mutation through removal of its catalytic domain (Lei, Oh et al. 1996, Ng, Dean et al. 2008). Despite clearly showing a reduction in Dnmt1 expression, the current data cannot guarantee I had fully depleted this protein in any cell.

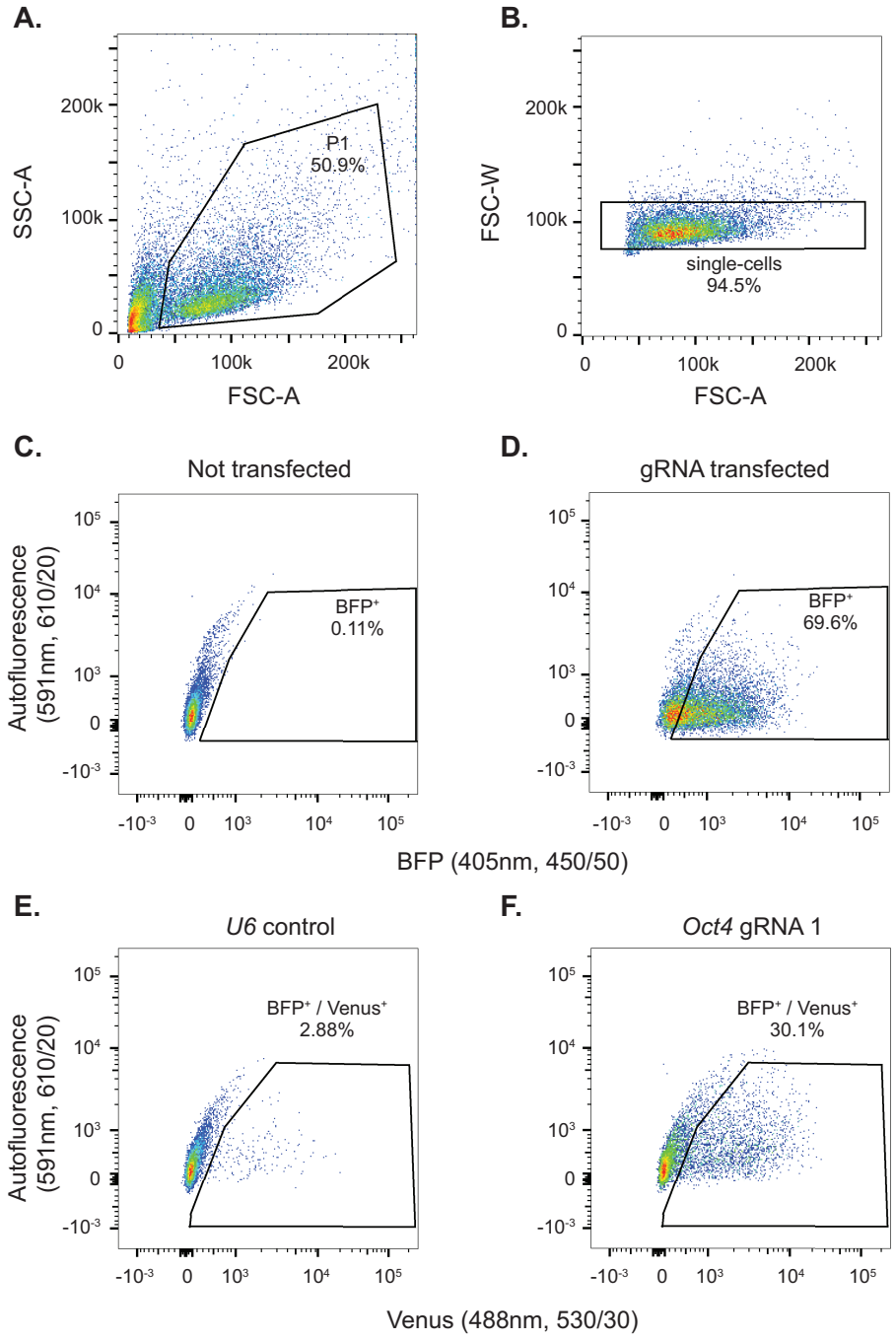
On the other hand, the phenotype associated with *Dnmt1* knockout cells is dependent on CpG island methylation. In a wild-type scenario, *Dnmt1* enzyme ensures that methylation patterns are maintained during cell division. However, it's the demethylation of *Elf5* promoter region that allows *Dnmt1* null cells to differentiate towards TSCs (Ng, Dean et al. 2008). In the absence of *Dnmt1*, passive demethylation through cell division is likely the mechanism for loss of methylation marks. Therefore, CRISPR/Cas9-mediated knockout cells might require longer culture periods before achieving a TSC-permissive state.

Further experiments would be needed to address lack of TSC differentiation. For instance, ensuring the generation of a homozygous knockout by using two gRNAs to

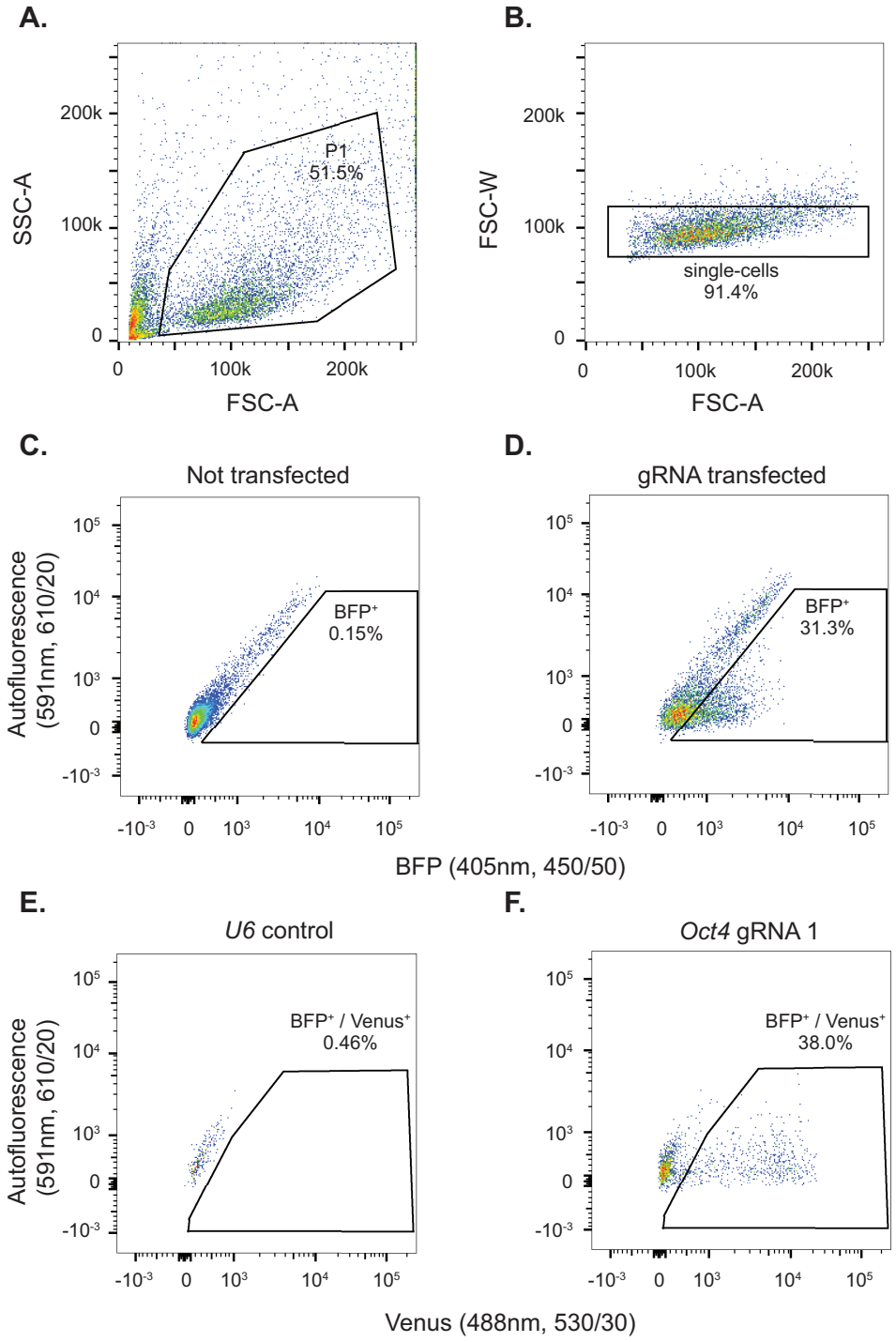
induce deletion of the catalytic domain, followed by colony picking and genotyping. This type of experiments would allow to understand the effect of *Dnmt1* levels in differentiation to TSC, but would not be helpful in the context of a pooled genome-wide screening as current methodology relies on indel induction through a single gRNA generating a pool of mutant cells. Additionally, culturing knockout ESCs for longer periods might help to reveal demethylation-related phenotypes, but would introduce higher bias and misrepresentation of genes that have an effect on ESC survival/growth.

Nevertheless, it is important to highlight that a small percentage of frame-shifting mutations in the case of *Oct4* gene were enough to allow differentiation towards TSC at the population level, which is clearly not the case for *Dnmt1*. Perhaps this is related with the mechanistic nature of *Dnmt1* phenotype in comparison with strong regulatory genes such as *Oct4* which have an acute effect on this differentiation. Identifying genes with a profile similar to *Dnmt1* will be a limitation for a genome-wide screening, as highlighted with this proof-of-principle experiment. The current setting is more likely to detect genes with a strong phenotype, which do not affect ESC self-renewal.

3.4 Supplementary Figures



Supplementary Figure 3.1 | Gating strategy for flow cytometry analysis of trophoblast differentiation in TSC media. The plots represent data acquired by day 14 of a typical differentiation experiment. Upon gating of P1 (A) and single cells (B), cells transfected with a gRNA construct were selected based on BFP expression: C | non-transfected control; and D | gRNA transfected line. BFP-positive cells were then evaluated for Venus expression: E | Representative result for U6 control; F | Representative result for *Oct4* gRNA 1 transfection.



Supplementary Figure 3.2 | Gating strategy for flow cytometry analysis of trophoblast differentiation in TX media. The plots represent data acquired by day 14 of a typical differentiation experiment. Upon gating of P1 **(A)** and single cells **(B)**, cells transfected with a gRNA construct were selected based on BFP expression: **C |** non-transfected control; and **D |** gRNA transfected line. BFP-positive cells were then evaluated for Venus expression: **E |** Representative result for U6 control; **F |** Representative result for *Oct4* gRNA 1 transfection.