© 2016. Published by The Company of Biologists Ltd | Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

Optimized inducible shRNA and CRISPR/Cas9 platforms for *in vitro* studies of human development using hPSCs

Alessandro Bertero[1,2,‡,§,¶], Matthias Pawlowski[1,3,§], Daniel Ortmann[1,2,§], Kirsten Snijders[1,2], Loukia Yiangou[1,4], Miguel Cardoso de Brito[1,2], Stephanie Brown[1,2], William G. Bernard[1,4], James D. Cooper[1,4], Elisa Giacomelli[1,2], Laure Gambardella[1,4], Nicholas R. F. Hannan[1,2,*], Dharini Iyer[1,4], Fotios Sampaziotis[1,2], Felipe Serrano[1,4], Mariëlle C. F. Zonneveld[1,2], Sanjay Sinha[1,4], Mark Kotter[1,3] and Ludovic Vallier[1,2,5,¶]

ABSTRACT

Inducible loss of gene function experiments are necessary to uncover mechanisms underlying development, physiology and disease. However, current methods are complex, lack robustness and do not work in multiple cell types. Here we address these limitations by developing single-step optimized inducible gene knockdown or knockout (sOPTiKD or sOPTiKO) platforms. These are based on genetic engineering of human genomic safe harbors combined with an improved tetracycline-inducible system and CRISPR/Cas9 technology. We exemplify the efficacy of these methods in human pluripotent stem cells (hPSCs), and show that generation of sOPTiKD/KO hPSCs is simple, rapid and allows tightly controlled individual or multiplexed gene knockdown or knockout in hPSCs and in a wide variety of differentiated cells. Finally, we illustrate the general applicability of this approach by investigating the function of transcription factors (*OCT4* and *T*), cell cycle regulators (cyclin D family members) and epigenetic modifiers (*DPY30*). Overall, sOPTiKD and sOPTiKO provide a unique opportunity for functional analyses in multiple cell types relevant for the study of human development.

KEY WORDS: Human pluripotent stem cells, shRNA, CRISPR/Cas9, OCT4, POU5F1, T, brachyury, DPY30

INTRODUCTION

Loss-of-function experiments in human pluripotent stem cells [hPSCs; comprising human embryonic stem cells (hESCs) or human induced pluripotent stem cells (hiPSCs)] provide a unique opportunity to study the mechanisms that regulate human development, physiology and disease (Avior et al., 2016; Pourquié et al., 2015; Zhu and Huangfu, 2013). However, functional genomic applications of hPSCs are currently limited by the lack of an easy and efficient method to conditionally manipulate gene expression in both hPSCs and hPSC-derived cells. Indeed, such a system is necessary both for the study of genes essential for hPSC self-renewal and for functional analyses at specific stages of differentiation.

Historically, the expression of inducible short hairpin RNAs (shRNAs) has been the most popular method to trigger gene knockdown in human cells. This has been achieved using a TET-ON system, which relies on a modified RNA polymerase (Pol) III promoter that is responsive to a tetracycline-sensitive repressor protein (tetR) to induce shRNA expression by simple tetracycline (TET) treatment (Lambeth and Smith, 2013). Nevertheless, application of this TET-ON system in hPSCs has proved challenging for two main reasons: (1) tight control of shRNA expression is difficult to achieve, thereby resulting in uncontrolled knockdown; (2) induction of shRNA rarely works in differentiated derivatives. Indeed, very high and homogenous expression of both the tetR and the inducible shRNA is required to obtain potent yet controlled knockdown. However, transgene silencing is a recurring problem in hPSCs (Ellis, 2005; Herbst et al., 2012; Yao et al., 2004), and randomly integrated promoters are often subject to positional effects that can strongly limit their activity (Zafarana et al., 2009). Differentiation further increases the chances of silencing, as transgenes can be located in regions where heterochromatin forms following cell fate choices (Herbst et al., 2012; Raya et al., 2009). As a consequence, inducible shRNA expression in both hPSCs and a wide variety of their differentiated progenies has never been reported.

More recently, CRISPR/Cas9-mediated gene knockout has emerged as a powerful method to interrogate gene function (Wright et al., 2016), and inducible manipulation of gene expression in hPSCs using this approach has been reported (Chen et al., 2015; González et al., 2014; Mandegar et al., 2016). However, these methods are either very complex and time consuming, as they involve multiple genome editing steps that need to be individually tailored for each gene to be examined (Chen et al., 2015), or are not widely applicable in multiple differentiated cell types as they rely on inducible promoters that are not stably and homogeneously expressed following hPSC differentiation (González et al., 2014; Haenebalcke et al., 2013; Mandegar et al., 2016; Ordovas et al., 2015). Overall, there are currently no methods for inducible gene knockout in hPSCs that fulfill all the criteria described above.

Here we describe novel platforms for single-step optimized inducible gene knockdown or knockout (sOPTiKD or sOPTiKO) that address all the limitations of current inducible shRNA or CRISPR/Cas9 systems, thus providing powerful and scalable

[1]Wellcome Trust-MRC Stem Cell Institute, Anne McLaren Laboratory, University of Cambridge, Cambridge, CB2 0SZ, UK. [2]Department of Surgery, University of Cambridge, Cambridge, CB2 0QQ, UK. [3]Department of Clinical Neuroscience, University of Cambridge, Cambridge, CB2 0QQ, UK. [4]Division of Cardiovascular Medicine, University of Cambridge, Cambridge, UK. [5]Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK.
*Present address: Centre for Biomolecular Sciences, University of Nottingham, Nottingham, NG7 2RD, UK. ‡Present address: Department of Pathology, University of Washington, Seattle, WA 98109, USA.
§These authors contributed equally to this work

¶Authors for correspondence (abertero@uw.edu; lv225@cam.ac.uk)

L.V., 0000-0002-3848-2602

Received 31 March 2016; Accepted 7 October 2016

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

platforms that have the potential to greatly simplify the study of human gene function.

RESULTS

Validation of the *ROSA26* and *AAVS1* loci as genomic safe harbors in hPSCs and their differentiated derivatives

We aimed to develop optimal conditional loss-of-function platforms using inducible shRNAs or guide RNAs (gRNAs) for CRISPR/Cas9. We reasoned that inserting each element of the TET-ON system into a different genomic safe harbor (GSH; Sadelain et al., 2012) would maximize expression in hPSCs and their differentiated progenies while avoiding potential promoter interference (Shearwin et al., 2005). The *AAVS1* and *ROSA26* loci appeared particularly suitable for this purpose as these GSHs have been suggested to allow strong expression of various transgenes in hPSCs, including constitutively expressed shRNAs (DeKelver et al., 2010; Hockemeyer et al., 2009; Irion et al., 2007). We first improved the targeting efficiency for both GSHs by developing a CRISPR/Cas9n-based gene-trap strategy to target the human *ROSA26* locus (Fig. 1A,B, Fig. S1A) and by refining an existing zinc-finger nuclease (ZFN)-based targeting strategy for the *AAVS1* locus (Hockemeyer et al., 2009) (Fig. 1A,B). In both cases, hPSC targeting occurred with very high efficiency (59-100%; Table S1), while neither *ROSA26* nor *AAVS1* modifications resulted in chromosomal abnormalities (data not shown).

We then sought to identify the most efficient promoter to drive constitutive transgene expression from GSHs. We tested the ability of different promoter configurations to express an enhanced green

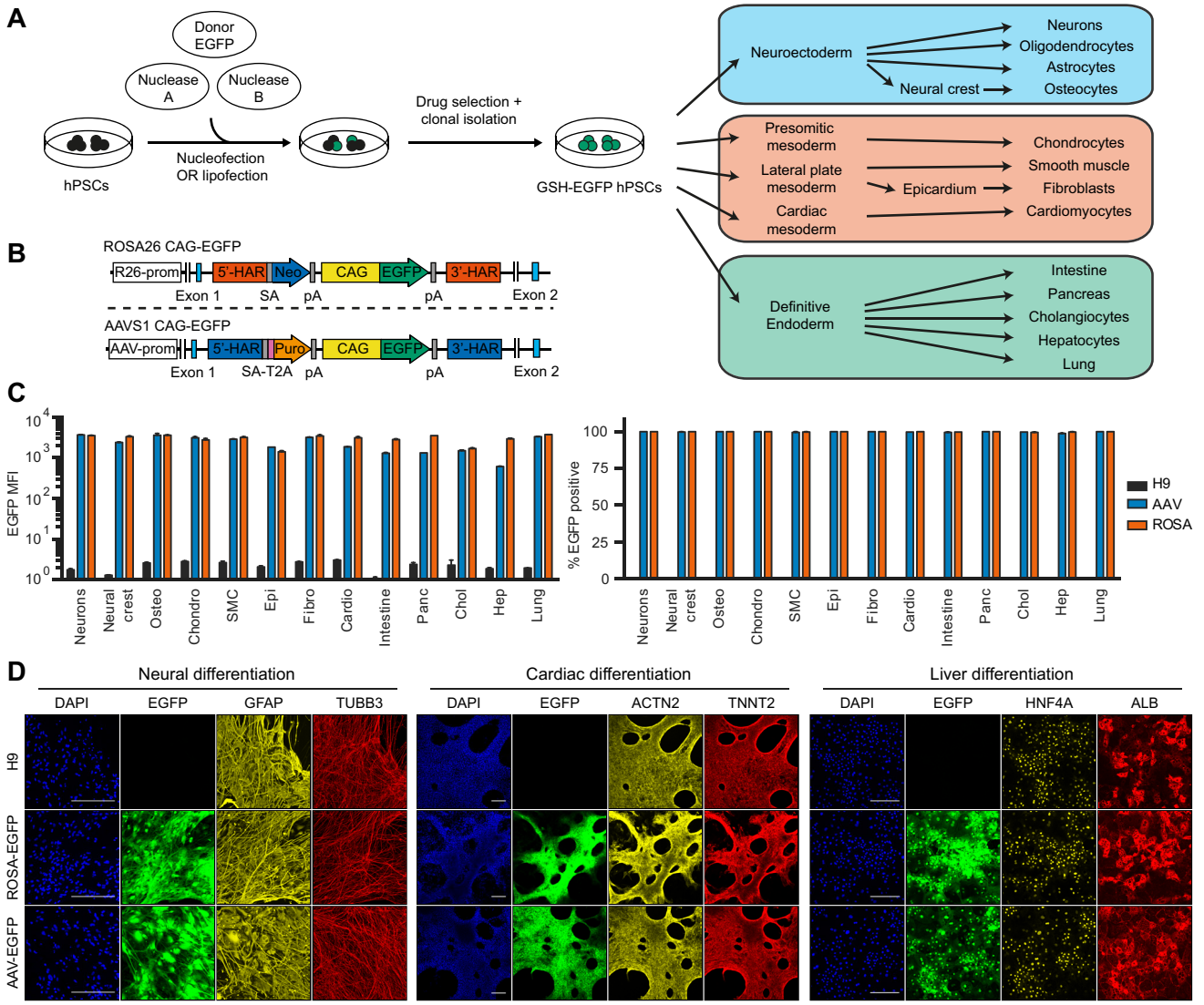


Fig. 1. Validation of the *ROSA26* and *AAVS1* loci as bona fide genomic safe harbors. (A) Experimental approach behind the generation of genomic safe harbor (GSH) EGFP reporter hPSCs to test GSH expression during differentiation. Neurons, oligodendrocytes and astrocytes were obtained in bulk cultures containing a mixture of these cell lineages, whereas all other cell types were individually generated. (B) *ROSA26* and *AAVS1* EGFP reporter transgenic alleles. R26-prom, *ROSA26* locus promoter; AAV-prom, *AAVS1* locus promoter; 5′-HAR/3′-HAR, upstream/downstream homology arm; SA, splice acceptor; T2A, self-cleaving T2A peptide; Neo, neomycin resistance; Puro, puromycin resistance; pA, polyadenylation signal; CAG, CAG promoter. (C) Summary of EGFP flow cytometry quantification experiments in the indicated cell types generated from GSH EGFP reporter hPSCs (abbreviations indicate the lineages described in A). The percentage of EGFP-positive cells and the EGFP median fluorescence intensity (MFI) are reported. Wild-type hESCs (H9) were used as negative controls, and results are from two independent cultures per lineage. (D) Representative immunofluorescent stainings for lineage-specific markers in three of the mature cell types analyzed. EGFP fluorescence from the reporter lines is in green, and DAPI (blue) shows nuclear staining. Scale bars: 200 µm.

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

fluorescent protein (EGFP) transgene from the *ROSA26* locus in hESCs (Fig. S1A,B). The highest and most homogenous EGFP expression (100%) was achieved with the artificial CAG promoter (Fig. S1C-E), which was stronger by an order of magnitude than the endogenous *ROSA26* promoter (Irion et al., 2007). Interestingly, and in contrast to previous reports (Ramachandra et al., 2011), we observed that the *EF1α* (*EEF1A1*) promoter was strongly silenced, as shown by mosaic EGFP expression (Fig. S1C-E). Similar results were obtained after EGFP targeting into the *AAVS1* locus (data not shown), thereby preventing the use of this promoter in subsequent experiments.

To further evaluate the robustness of the CAG promoter activity, we analyzed in detail hESCs with heterozygous or homozygous targeting of a CAG-EGFP transgene in the *ROSA26* or *AAVS1* loci (Fig. 1A,B). For both GSHs EGFP was homogeneously expressed at high and comparable levels for more than 30 passages (Fig. S1F), and similar results were obtained after differentiation of hESCs into the three primary germ layers (Fig. S1G-M). Importantly, targeting did not interfere with pluripotency or differentiation, as shown by appropriate expression of lineage markers (Fig. S1N,O). We further differentiated these EGFP-hESC lines into fifteen different cell types (Fig. 1A), and both GSHs allowed homogeneous and strong EGFP expression in all cell types analyzed (Fig. 1C,D, Fig. S2). Overall, these results validate the *ROSA26* and *AAVS1* loci as suitable for robust transgene expression in both hPSCs and their derivatives.

Development of an optimized inducible knockdown platform in hPSCs

Having demonstrated the suitability of the *ROSA26* and *AAVS1* loci for transgene expression, we developed a TET-ON inducible knockdown system based on dual GSH targeting (Fig. 2A, Fig. S3A). To simplify knockdown evaluation and method optimization we generated hESC lines in which an EGFP transgene could be silenced in an inducible fashion (Fig. 2B). To achieve this we targeted: (1) a CAG-tetR expression cassette into the *ROSA26* locus; and (2) a CAG-EGFP transgene plus an inducible EGFP shRNA cassette into the *AAVS1* locus (Fig. 2A,B). Interestingly, we observed a strong and homogeneous decrease in EGFP fluorescence following tetracycline treatment for 5 days (>95%; Fig. 2C), thereby confirming efficient knockdown. However, a decrease in EGFP expression was also noticed in the absence of tetracycline (Fig. 2C), suggesting a significant leakiness in the expression of the shRNA and thus confirming previous reports (Henriksen et al., 2007).

We then hypothesized that this limitation could be bypassed by expressing higher levels of the tetR protein to more strongly repress shRNA expression in the absence of tetracycline. We performed a multi-parameter RNA and codon optimization of the bacterial tetR cDNA (Fath et al., 2011) and used the resulting codon-optimized tetR (OPTtetR) to generate new EGFP inducible knockdown hESC lines (Fig. 2B). This modification achieved a tenfold increase in tetR expression compared with the standard sequence (STDtetR; Fig. 2D). Furthermore, homozygous expression of OPTtetR was sufficient to completely prevent shRNA leakiness while fully preserving efficient knockdown induction (Fig. 2C, Fig. S3B). Of note, the inducible knockdown was rapid, reversible and dose responsive (Fig. 2E,F, Fig. S3C-E). Finally, inducible hESCs displayed a normal karyotype (data not shown), demonstrating that the genome engineering necessary to create these lines did not alter their genetic stability.

Based on these encouraging results, we further validated this method in the context of endogenous genes by generating hESCs carrying inducible shRNAs against *OCT4* (*POU5F1*) or *B2M* (Fig. S3F). Remarkably, all the sublines analyzed (six for each gene) showed robust inducible knockdown with no significant shRNA leakiness (Fig. S3G,H). Tetracycline titration identified optimal concentrations to partially or fully knockdown *OCT4* (Fig. 2G, Fig. S3I,J). As expected, a strong decrease in *OCT4* specifically resulted in loss of pluripotency and induction of neuroectoderm and definitive endoderm markers (Fig. 2H, Fig. S3I,J) (Thomson et al., 2011; Wang et al., 2012). Similar results were obtained with 20 additional *OCT4* inducible knockdown hESC sublines, confirming the robustness and reproducibility of this method (Fig. S3K). Importantly, the generation of hESCs with strong and tightly regulated knockdown was so efficient that phenotypic analyses could be performed immediately after antibiotic selection on a mixed population of cells, thereby entirely bypassing the need to pick individual colonies for clonal isolation (Fig. S3K).

Overall, these results establish that dual targeting of GSHs with an optimized inducible knockdown system is a powerful method to control gene expression in hPSCs. This approach is hereafter named optimized inducible knockdown, or OPTiKD (Fig. 2A, Fig. S3F).

Single-step generation of optimized inducible knockdown hPSCs

We then sought to further improve the OPTiKD system by developing an all-in-one targeting approach that would facilitate the rapid and scalable generation of inducible knockdown hPSCs. We constructed a single *AAVS1* targeting vector to carry both the inducible shRNA and the CAG-tetR expression cassette (Fig. 3A), and validated this approach by knocking down the expression of an EGFP transgene targeted in the *ROSA26* locus (Fig. S3L). Remarkably, this method shared key properties with OPTiKD, such as both the absence of shRNA leakiness (Fig. S3L,M) and rapid, reversible and dose-responsive inducible knockdown (Fig. S3N,O). Thus, this all-in-one strategy, which we named single-step optimized inducible knockdown, or sOPTiKD (Fig. 3A), is as efficient as our original dual targeting approach.

To further demonstrate the versatility of sOPTiKD, we generated both hESC and hiPSC lines carrying an inducible shRNA against *OCT4* or *B2M*. Generation of sOPTiKD hPSCs following lipofection was rapid (2 weeks) and extremely efficient, as all the sublines generated showed robust inducible knockdown (Fig. 3B,C). qPCR analyses confirmed that knockdown of *OCT4* using sOPTiKD induced differentiation of both hESCs and hiPSCs, whereas knockdown of *B2M* had no effect (Fig. 3D). Overall, these experiments show that sOPTiKD provides an efficient system to knock down gene expression that can be easily applied to a large number of hPSC lines.

Finally, we explored whether sOPTiKD could enable simultaneous knockdown of multiple genes (Fig. 3E). We focused on the cyclin D family (*CCND1*, *CCND2* and *CCND3*). These cell cycle regulators are functionally redundant, and thus their study in hESCs has previously required laborious multiple rounds of stable shRNA transfection in order to achieve double or triple knockdown (Pauklin and Vallier, 2013). We developed a method to easily combine multiple shRNAs into the same targeting vector using a one-step Gibson assembly, and generated sOPTiKD plasmids carrying one, two or three shRNAs against cyclin D genes or scrambled control shRNAs (Fig. 3E). These vectors were tested in hESCs without isolation of clonal sublines, and inducible knockdown proved highly efficient and comparable with single, double and triple shRNA constructs (Fig. 3F). Interestingly, prolonged knockdown of one or two cyclin Ds was compatible with hESC self-renewal (Fig. 3G), whereas triple knockdown

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

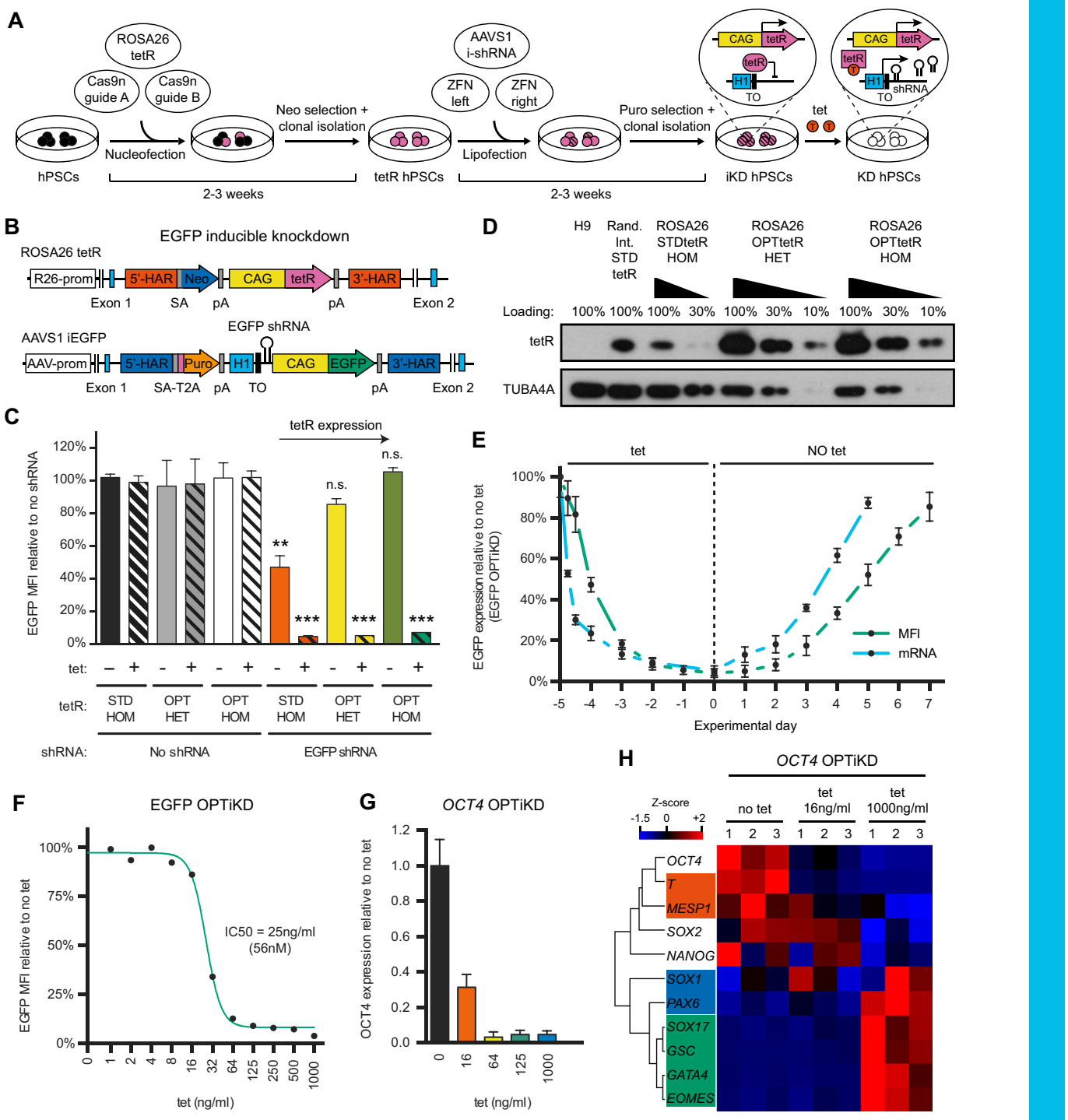


Fig. 2. Development of an optimized inducible knockdown system (OPTiKD) based on dual GSH targeting of hPSCs. (A) Experimental approach for the generation of inducible knockdown (iKD) hPSCs. H1, H1 promoter; TO, tet operon; tetR, tetracycline-controlled repressor; ZFN, zinc-finger nuclease. (B) Transgenic alleles generated to obtain hESCs expressing an EGFP reporter transgene that could be silenced using an inducible EGFP shRNA. (C) EGFP expression in the absence or presence of tetracycline for 5 days in hESCs targeted with the indicated combinations of inducible EGFP shRNA and tetR [wild-type standard tetR (STDtetR) or codon-optimized tetR (OPTtetR)]. Double-targeted hESCs that did not carry the EGFP shRNA were used as negative controls. Results are from two or three individual lines per condition (see Table S1). n.s., $P>0.05$ (non-significant), $^{**}P<0.01$, $^{***}P<0.001$ versus the same tetR line no tet and no shRNA (ANOVA with post-hoc Holm-Sidak comparisons). (D) Representative western blot for tetR in *ROSA26*-targeted hESCs expressing STDtetR or OPTtetR. HET, heterozygous targeting; HOM, homozygous targeting. hESCs with STDtetR random integration (Rand. Int.) are shown as a positive reference, while wild-type H9 hESCs are negative controls. Various amounts of protein were loaded to facilitate semi-quantitative comparison. TUBA4A (α-tubulin) provided a loading control. (E) EGFP knockdown and rescue kinetics in EGFP OPTiKD hESCs measured by flow cytometry (MFI) and qPCR (mRNA). Results are from two independent cultures per time point. (F) Tetracycline dose-response curve for EGFP knockdown in EGFP OPTiKD hESCs. The half-maximal inhibitory concentration (IC50) is reported. Results are from two independent cultures per dose, and the mean is shown. (G,H) qPCR analysis of *OCT4* OPTiKD hESCs in the absence of tetracycline, or following treatment with different doses of tetracycline for 5 days. (H) Genes are clustered by complete Euclidean distance, and genes specific for pluripotency or for the primary germ layers are in color-coded boxes: no color, hPSCs; red, mesoderm; green, endoderm; blue, neuroectoderm. Z-scores indicate differential expression measured in the number of standard deviations from the average level. Results are from three independent cultures.

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

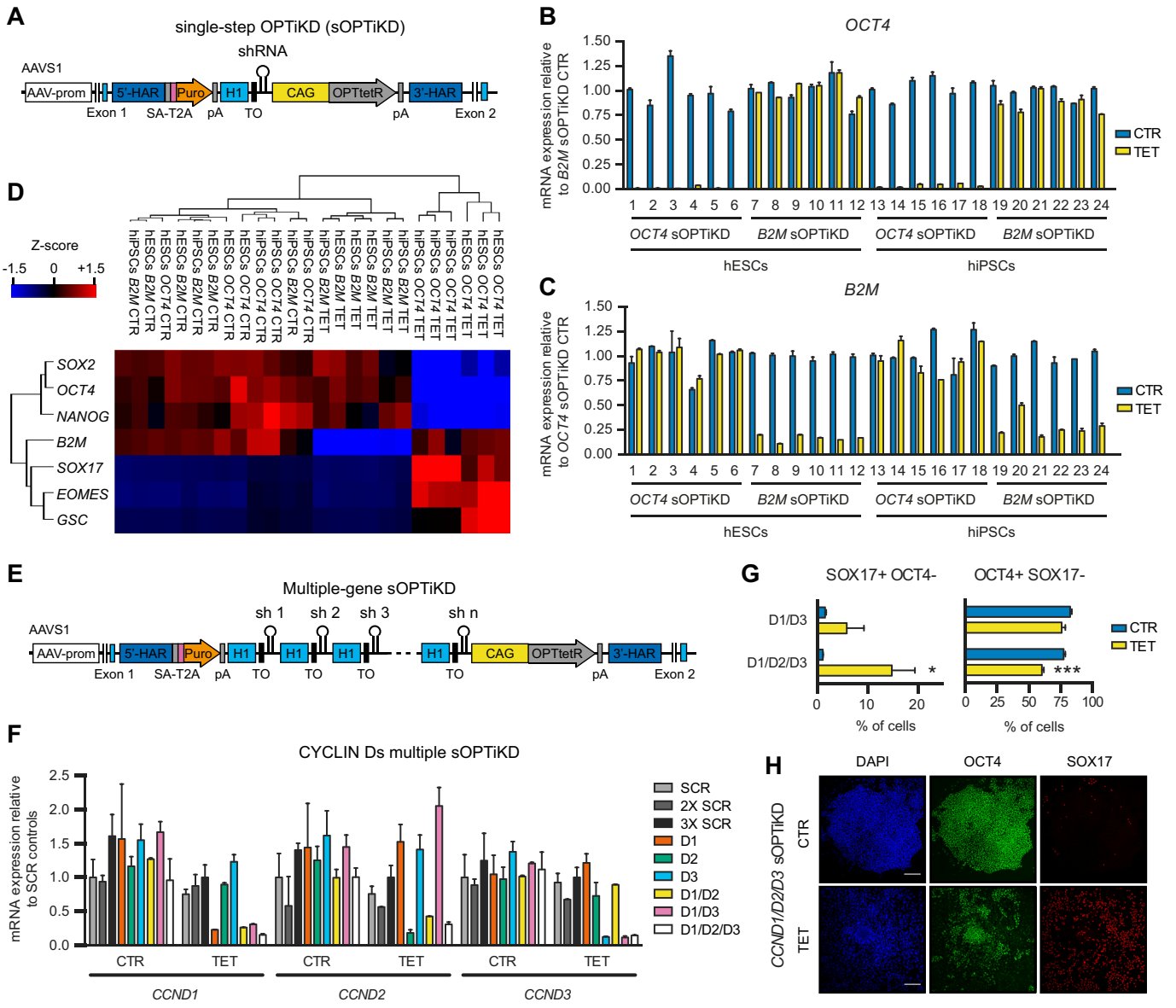


Fig. 3. Single-step optimized inducible knockdown (sOPTiKD) of individual and multiple genes in hESCs and hiPSCs. (A) The transgenic allele behind the single-step generation of OPTiKD hPSCs. (B,C) qPCR of *OCT4* and *B2M* sOPTiKD hESCs and hiPSCs in the absence (CTR) or presence of tetracycline for 5 days (TET). Individual clonal lines were analyzed in duplicate. (D) Heatmap summarizing qPCR analysis of cells treated as in B,C. Results are from three clonal lines per condition. Samples and genes were clustered by complete Euclidean distance, and Z-scores indicate differential expression measured in number of standard deviations from the average level. (E) The transgenic allele behind the generation of hPSCs with inducible knockdown of multiple genes. (F) qPCR analysis of sOPTiKD hESCs for individual or multiple cyclin D genes (D1, D2 and D3 in the key indicating inducible shRNAs against *CCND1*, *CCND2* and *CCND3*, respectively). Cells were analyzed in the absence or presence of tetracycline for 10 days, and sOPTiKD hESCs carrying one, two or three copies of a scrambled shRNA (SCR, 2X SCR and 3X SCR, respectively) were used as negative controls. For each condition, results are from two clonal pools obtained after gene targeting. (G,H) Flow cytometry quantifications (G) and representative immunostainings (H) for the pluripotency marker OCT4 and the definitive endoderm marker SOX17 in cells treated as in F. **P*<0.05, ****P*<0.001 versus CTR in the same line (ANOVA with post-hoc Holm-Sidak comparisons). DAPI shows nuclear staining. Scale bars: 100 µm.

resulted in definitive endoderm differentiation (Fig. 3G,H), as previously reported (Pauklin and Vallier, 2013; Pauklin et al., 2016). Collectively, these results demonstrate that sOPTiKD can be used to simultaneously decrease the expression of several genes with redundant functions.

Validation of the optimized inducible knockdown platforms in differentiated progenies of hPSCs

The capacity to knock down genes in a variety of differentiated cells would represent a significant advance over existing systems for inducible gene knockdown. To thoroughly test this possibility, we analyzed the efficacy of the OPTiKD and sOPTiKD platforms to knock down an EGFP transgene in hPSCs differentiated into the three germ layers, as well as in a panel of 13 fully differentiated cell types (Fig. 1A). For both methods, qPCR analyses demonstrated strong and inducible knockdown of EGFP transcripts in all lineages tested (Fig. 4A). Microscopy observations confirmed a robust decrease in EGFP protein expression (Fig. 4B), and flow cytometry showed a decrease in EGFP fluorescence of more than 70% for most lineages (Fig. S4A-G).

Interestingly, EGFP was less reduced in cell types with slower proliferation rates (Fig. S4A). Since EGFP has an extended half-life

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

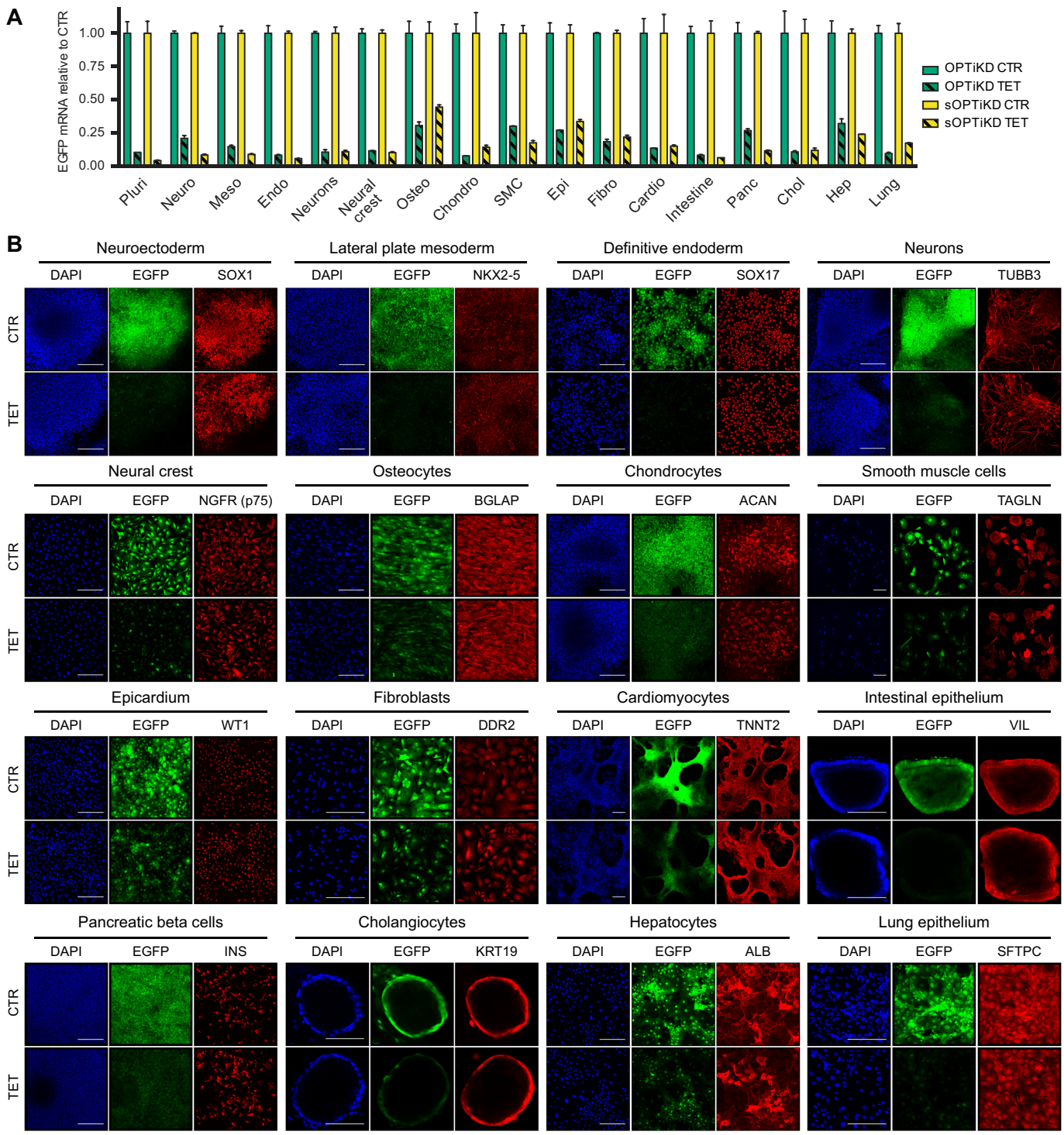


Fig. 4. Validation of the optimized inducible knockdown platforms following hPSC differentiation. (A) EGFP expression measured by qPCR in the absence (CTR) or presence of tetracycline for 5 days (TET) in the indicated cell types derived from EGFP OPTiKD and sOPTiKD hESCs. EGFP levels are reported relative to control conditions in the same line for each individual lineage. Abbreviations indicate the lineages described in Fig. 1A (pluri indicates undifferentiated). Results are from two independent cultures per condition. (B) Representative immunofluorescent stainings for lineage-specific markers in the three germ layers and in the indicated mature cell types derived from EGFP sOPTiKD hESCs and treated as in A. EGFP fluorescence is in green, and DAPI shows nuclear staining. Similar results supporting the same conclusions were obtained for EGFP OPTiKD hESCs (data not shown). Scale bars: 100 µm for intestinal epithelium and cholangiocytes; 200 µm for all other lineages.

of more than 24 h, protein loss upon transcriptional or post-transcriptional inhibition relies heavily upon its gradual dilution following cell division (Li, 1998). Considering the strong decrease in EGFP mRNA, we concluded that residual EGFP fluorescence was likely to be a consequence of the relatively short tetracycline treatment performed to trigger knockdown (5 days). To test this, we induced prolonged EGFP knockdown in postmitotic cardiomyocytes, and we indeed observed a slow but constant

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

decrease in protein expression for up to 20 days, at which point the EGFP was decreased by more than 75% (Fig. S4H,I). To reinforce these observations, we performed similar experiments using a *ROSA26*-targeted EGFP reporter transgene fused to a destabilization domain (EGFPd2; Li et al., 1998), which avoids confounding effects due to the long half-life of standard EGFP (Fig. S4J-M) (Wahlers et al., 2001). Remarkably, EGFPd2 inducible knockdown in cardiomyocytes using the sOPTiKD method resulted in >90% protein knockdown after 5 days of tetracycline treatment (Fig. S4N,O). Considered together, these results establish that OPTiKD and sOPTiKD allow efficient manipulation of gene expression even after differentiation of hPSCs.

Inducible knockdown of T (brachyury) during mesendoderm differentiation of hPSCs

We then sought to exemplify the use of OPTiKD and sOPTiKD to rapidly and efficiently evaluate endogenous gene function in a variety of cell types and at different stages of hPSC differentiation related to embryonic development. First, we focused on the master developmental regulator T (brachyury), which plays an essential role in mesoderm formation and, in particular, during the development of posterior mesoderm, notochord and somites (Martin, 2015; Papaioannou, 2014). Indeed, mice carrying a heterozygous mutation in *T* exhibit a short tail phenotype, while homozygous mutations are embryonic lethal at around 9.5 dpc (Dobrovolskaia-Zavadskaia, 1927; Gluecksohn-Schoenheimer, 1944). *T* mutants also present severe cardiovascular and placental defects (David et al., 2011; Inman and Downs, 2006; King et al., 1998). Furthermore, *T* was recently shown to specifically regulate mesoderm but not endoderm differentiation in hPSCs (Faial et al., 2015).

To investigate the role of *T* during the differentiation of hPSCs we combined *T* sOPTiKD sublines with culture conditions known to drive the differentiation of hPSCs into subpopulations that recapitulate different portions of the primitive streak and their derived lineages (Fig. 5A) (Bernardo et al., 2011; Cheung et al., 2012; Mendjan et al., 2014; Touboul et al., 2010). Inducible knockdown of *T* was robust in all cell types analyzed (Fig. 5B,C, Fig. S5A,B), confirming the efficiency of sOPTiKD to knock down developmental genes. Decrease in *T* expression did not affect definitive endoderm specification, while differentiation into posterior primitive streak cells, cardiac mesoderm and lateral plate mesoderm was mildly impaired (Fig. 5D, Fig. S5C). By contrast, the generation of late primitive streak progenitors (recapitulating the onset of somitogenesis) and their further specification into presomitic mesoderm and chondrocytes were severely affected following inducible knockdown of *T* (Fig. 5D-G). In particular, induction of *TBX6*, *MSGN1* and *MEOX1* was nearly abolished, in agreement with the established role of *T* in the expression of such genes (Chapman et al., 1996; Faial et al., 2015; Martin, 2015).

Collectively, these results strikingly recapitulate the known role of *T* during early embryonic development, thereby demonstrating the versatility of OPTiKD platforms to study the mechanisms of human development *in vitro*.

Inducible knockdown of *DPY30* at various stages of hPSC differentiation reveals stage- and lineage-specific functions

We then aimed to demonstrate the suitability of the OPTiKD platforms to investigate the function of genes that are not only expressed during early development, but also in differentiated cells. We focused on *DPY30*, a ubiquitously expressed co-factor of the COMPASS histone methyltransferase complexes required for histone H3 lysine 4 trimethylation (H3K4me3) (Jiang et al., 2011). This epigenetic modifier is necessary for mouse early embryonic development, as its knockout leads to impaired gastrulation associated with ectopic neuralization of the post-implantation epiblast (Bertero et al., 2015). Similarly, *DPY30* is required for hESC pluripotency (Bertero et al., 2015), and this early role had prevented further studies of its function during differentiation. Finally, *Dpy30* has been implicated in mouse ESC differentiation and in the proliferation and differentiation of hematopoietic progenitors (Jiang et al., 2011; Yang et al., 2014). Consequently, we decided to employ our inducible knockdown platform to bypass the early function of *DPY30* in hPSCs and specifically suppress its expression during differentiation (Fig. 6A).

First, we generated *DPY30* OPTiKD hESC sublines (Fig. S6A), and confirmed that inducible *DPY30* knockdown in undifferentiated hESCs impaired the expression of pluripotency genes and triggered neuroectoderm differentiation (Fig. 6B, Fig. S6B), as shown previously (Bertero et al., 2015). We then analyzed the function of *DPY30* during lineage specification by differentiating *DPY30* OPTiKD hESCs into five different cell types while inducing *DPY30* knockdown from the induction, specification or maturation stages (Fig. 6A). qPCR confirmed the decrease in *DPY30* expression in all the cells generated (Fig. S6C-H). Interestingly, phenotypic analyses demonstrated that *DPY30* knockdown from the early induction of cardiac specification impaired cardiomyocyte differentiation, as shown by the decrease in contractile markers (Fig. 6C). However, knockdown at later stages had no significant effects (Fig. 6C). A similar result was observed for the hepatocyte lineage, since decrease of *DPY30* expression in endoderm progenitors led to extensive cell death at the anterior foregut stage thereby preventing further differentiation (Fig. S6I). Similarly, specification of pancreatic endocrine cells was also impaired by knockdown of *DPY30* in the initial stage of differentiation (Fig. 6D). However, neither hepatocyte nor pancreatic endocrine cell specification was significantly affected by knockdown of *DPY30* in maturing progenitors or differentiated cells (Fig. 6D,E, Fig. S6I). By contrast, neuronal differentiation was promoted following *DPY30* knockdown during the induction of neuroepithelial progenitors (Fig. S6J). Finally, *DPY30* knockdown at any stage during smooth muscle cell differentiation had no effect on the expression of key lineage markers (Fig. S6K).

Considered together, these data confirm a key role for *DPY30* during germ layer specification while suggesting that the requirement for *DPY30* expression could vary during the differentiation and maturation of specific lineages (Fig. 6E). Overall, these experiments illustrate how the optimized inducible knockdown platform can be easily applied to acquire novel information about developmental mechanisms by performing functional studies at different steps of hPSC differentiation into multiple cell types.

Development of an optimized inducible CRISPR/Cas9 knockout platform in hPSCs

Having established an optimized inducible knockdown platform, we turned our attention to developing a complementary inducible knockout approach. Current inducible CRISPR/Cas9 methods rely on conditional overexpression of Cas9 in the presence of a constitutively expressed gRNA (González et al., 2014; Mandegar et al., 2016). In this case, control of Cas9 overexpression is achieved by a TET-ON method in which, following doxycycline treatment, a tetracycline-controlled reverse transactivator (rtTA) activates a Pol

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

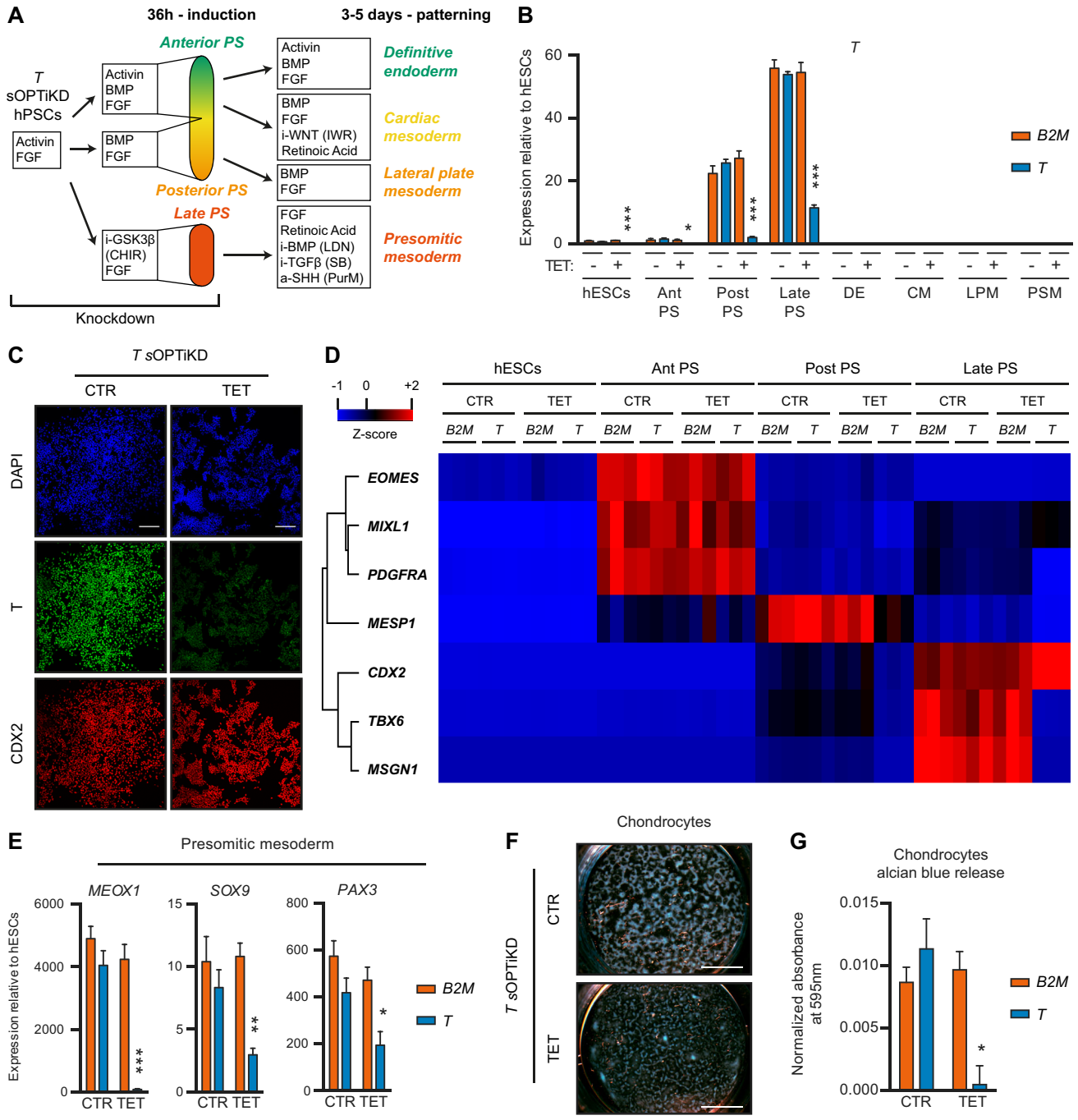
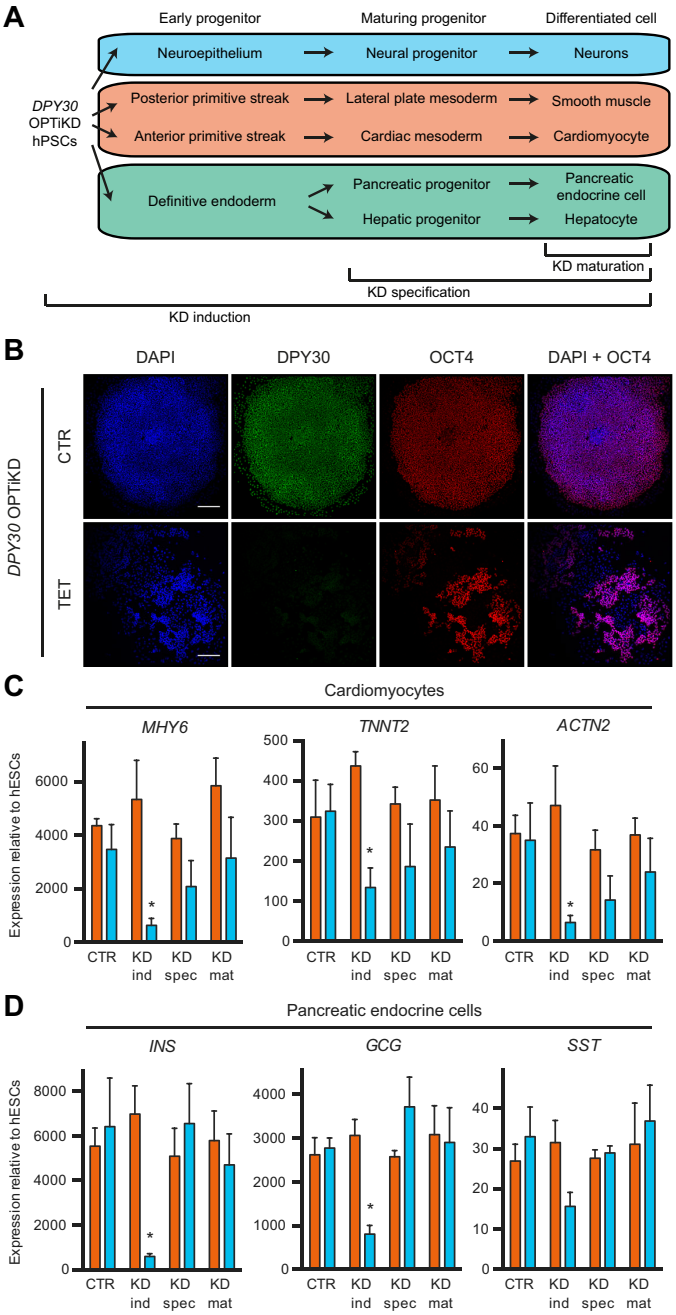


Fig. 5. Functional analysis of T (brachyury) during mesendoderm specification of hPSCs. (A) The experimental approach. *T* knockdown was induced for 2 days in pluripotent cells and maintained throughout differentiation. PS, primitive streak-like cells; i-, inhibitor; a-, activator; CHIR, CHIR99021; LDN, LDN193189; SB, SB431542; PurM, purmorphamine. (B) qPCR for *T* in the indicated lineages derived from *T* or *B2M* (control) sOPTiKD hESCs as described in A. Ant/Post/Late PS refer to anterior/posterior/late primitive streak; DE, definitive endoderm; CM, cardiac mesoderm; LPM, lateral plate mesoderm; PSM, presomitic mesoderm. (C) Representative immunofluorescent staining demonstrating inducible knockdown of T in late primitive streak cells expressing the marker CDX2. DAPI shows nuclear staining. (D) Heatmap summarizing qPCR results for various mesendoderm markers in primitive streak cells from the experiment described in A,B. Genes were clustered by complete Euclidean distance, and Z-scores indicate differential expression measured in number of standard deviations from the average level. (E) qPCR results for lineage-specific markers in presomitic mesoderm cells from the experiment described in A,B. (F,G) Representative Alcian Blue staining (F) and quantification of Alcian Blue release (G) in chondrocytes differentiated from presomitic mesoderm cells generated as described in A,B. (B,E,G) **P*<0.05, ***P*<0.01, ****P*<0.001 versus *B2M* in the same condition (two-way ANOVA with post-hoc Sidak comparisons), and results are from three independent clonal lines per condition. Scale bars: 100 µm in C; 4 mm in F.

II-dependent tetracycline-responsive element (TRE) promoter (a fusion between multiple TET operons and a minimal CMV promoter). Although this TET-ON platform has been successfully applied to certain human cell types (Qin et al., 2010), we observed that this inducible system is silenced during hPSC differentiation into multiple lineages (including cardiomyocytes, hepatocytes and smooth muscle cells), even after targeting into the *AAVS1* GSH (Fig. S7). These observations reinforce previous reports

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081



E

DPY30 KD:	Induction	Specification	Maturation
Neurons	Improved differentiation	No effect	No effect
Smooth muscle	No effect	No effect	No effect
Cardiomyocytes	Impaired differentation	No effect	No effect
Endocrine pancreas	Impaired differentiation	No effect	No effect
Hepatocytes	Differentiation failure	No effect	No effect

Fig. 6. Functional analysis of *DPY30* during hPSC differentiation. (A) The experimental design to investigate the role of *DPY30* at various stages of hPSC differentiation. (B) Representative immunofluorescent staining for DPY30 and the pluripotency marker OCT4 in undifferentiated *DPY30* OPTiKD hESCs in the absence (CTR) or presence of tetracycline for 10 days (TET). DAPI shows nuclear staining. Scale bars: 200 µm. (C,D) qPCR-based analyses of *DPY30* (blue) and *B2M* (orange) OPTiKD hESCs after differentiation into mature lineages. CTR, no knockdown; KD ind/spec/mat refer to knockdown from induction/specification/maturation, respectively. Results are from three independent cultures per condition. **P*<0.05 versus *B2M* in the same condition (two-way ANOVA with post-hoc Sidak comparisons). (E) Summary of the lineage- and stage-specific phenotypic effects following *DPY30* knockdown during hESC differentiation.

(Haenebalcke et al., 2013; Mandegar et al., 2016; Ordovas et al., 2015) and demonstrate that recently described systems for inducible CRISPR/Cas9 (González et al., 2014; Mandegar et al., 2016) are unlikely to work in a diversity of hPSC-derived cell types. For this reason, we explored the possibility of developing an alternative and improved method by combining a constitutively expressed CAG promoter-driven Cas9 with an inducible gRNA cassette based on that developed for inducible shRNA expression in sOPTiKD (Fig. 7A,B).

We generated hESC lines in which a fluorescent reporter gene could be knocked out in an inducible fashion (Fig. 7C). For this, we targeted *ROSA26*-EGFPd2 reporter hESCs (Fig. S4J,K) with both an inducible EGFP gRNA and a constitutive Cas9 in the *AAVS1* locus, each transgene being integrated into one of the two alleles (Fig. 7C,D). This dual targeting approach was rapid (<2 weeks) and efficient (>90% of lines containing both transgenes; Table S1). Remarkably, when individual clonal sublines were grown in the presence of tetracycline we observed decreased EGFPd2 expression in all of the targeted lines, and EGFPd2 homozygous cells showed near-homogeneous loss of at least one copy of the reporter gene as early as 5 days following tetracycline induction (as demonstrated by 50% reduction in EGFPd2 fluorescence, Fig. S8A). Prolonged treatment with tetracycline progressively led to the complete loss of EGFPd2 fluorescence in up to 75% of EGFPd2 homozygous cells (Fig. 7E,F, Fig. S8A,B). Interestingly, co-expression of either two or three copies of the same EGFP gRNA cassette from the same *AAVS1* locus was sufficient to significantly increase the speed and efficiency of inducible EGFPd2 knockout in all the clonal sublines analyzed (Fig. 7H,I, Fig. S8A). For instance, simultaneous induction of three copies of the same gRNA resulted in a remarkable 95% knockout efficiency following tetracycline treatment (Fig. 7I). Importantly, inducible EGFPd2 knockout hESCs did not show any significant decrease in the proportion of EGFPd2-positive cells nor in their fluorescence after prolonged culture in the absence of tetracycline, even when several gRNA copies were used (Fig. 7G, Fig. S8C,D). This demonstrated that the inducible gRNA expression was tightly controlled. Finally, testing of additional gRNAs against EGFPd2 revealed that the speed and efficiency of the inducible knockout strongly relied on the gRNA. Indeed, an optimal sequence allowed up to 90% knockout after only 2 days of induction (Fig. S8E,F,K). Of note, the most efficient gRNA also resulted in uncontrolled EGFPd2 knockout (Fig. S8G), but this limitation was avoided by simply adding a second TET operon to the inducible H1 promoter to ensure even more stringent transcriptional control (Fig. S8H-K).

Collectively, these results show that the sOPTiKD system could be readily repurposed to support inducible gRNA expression and allow tightly controlled activity of CRISPR/Cas9 over a broad range of gRNA potency (Fig. S8L). To the best of our knowledge, this is the first conditional CRISPR/Cas9 approach based on inducible gRNA expression. We named this method single-step optimized inducible gene knockout, or sOPTiKO.

Validation of the optimized inducible CRISPR/Cas9 platform in differentiated progenies of hPSCs

Having demonstrated that sOPTiKO allows efficient control of CRISPR/Cas9 activity in undifferentiated hPSCs, we thoroughly

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

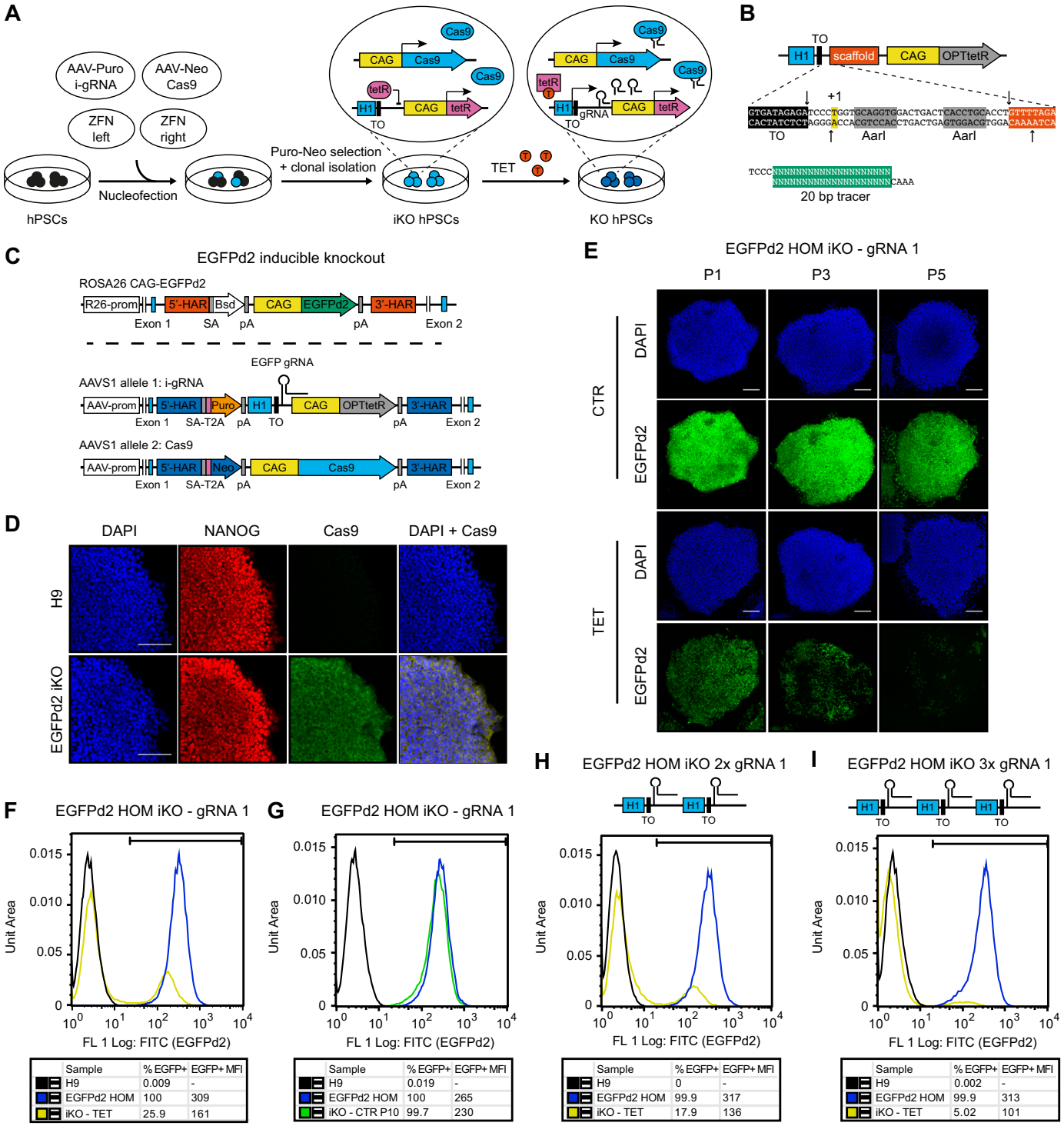


Fig. 7. Development of an optimized inducible CRISPR/Cas9 knockout platform in hPSCs. (A) Experimental approach for the generation of inducible knockout (iKO) hPSCs. (B) The cloning procedure to generate *AAVS1* targeting vectors with an inducible gRNA cassette. The arrows indicate the DNA cut sites induced by digestion with *Aar*I. (C) Transgenic alleles generated to obtain hESCs expressing an EGFPd2 reporter transgene that could be knocked out by CRISPR/Cas9 using an inducible EGFP gRNA (EGFP sOPTiKO hESCs). Bsd, blasticidin resistance; EGFPd2, destabilized EGFP. (D) Representative immunofluorescent stainings for Cas9 in EGFPd2 homozygous sOPTiKO hESCs. Wild-type hESCs (H9) were analyzed as a negative control. Cells were co-stained for the pluripotency factor NANOG, and DAPI shows nuclear staining. (E) Representative images depicting EGFPd2 fluorescence in EGFPd2 homozygous sOPTiKO hESCs in the absence (CTR) or presence of tetracycline (TET) for the indicated number of passages (P; cells were split every 5 days). (F) Representative flow cytometry for EGFPd2 expression in EGFPd2 homozygous sOPTiKO hESCs (iKO) following three passages in the presence of tetracycline. EGFPd2 homozygous cells that do not carry the inducible CRISPR/Cas9 system (EGFPd2 HOM) and wild-type hESCs were analyzed as positive and negative controls for EGFPd2 expression, respectively. The gate used to define EGFPd2-positive cells ($EGFP^+$) is shown, and the percentage of $EGFP^+$ cells and their MFI are reported. (G) As in F, but EGFPd2 homozygous sOPTiKO hESCs were analyzed following ten passages in the absence of tetracycline. (H,I) As in F, but EGFPd2 homozygous sOPTiKO hESCs were generated using an *AAVS1* targeting vector carrying two or three copies of the inducible EGFP gRNA cassette (2× and 3× gRNA, respectively). All results in this figure were obtained using EGFP gRNA 1. Scale bars: 100 µm.

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

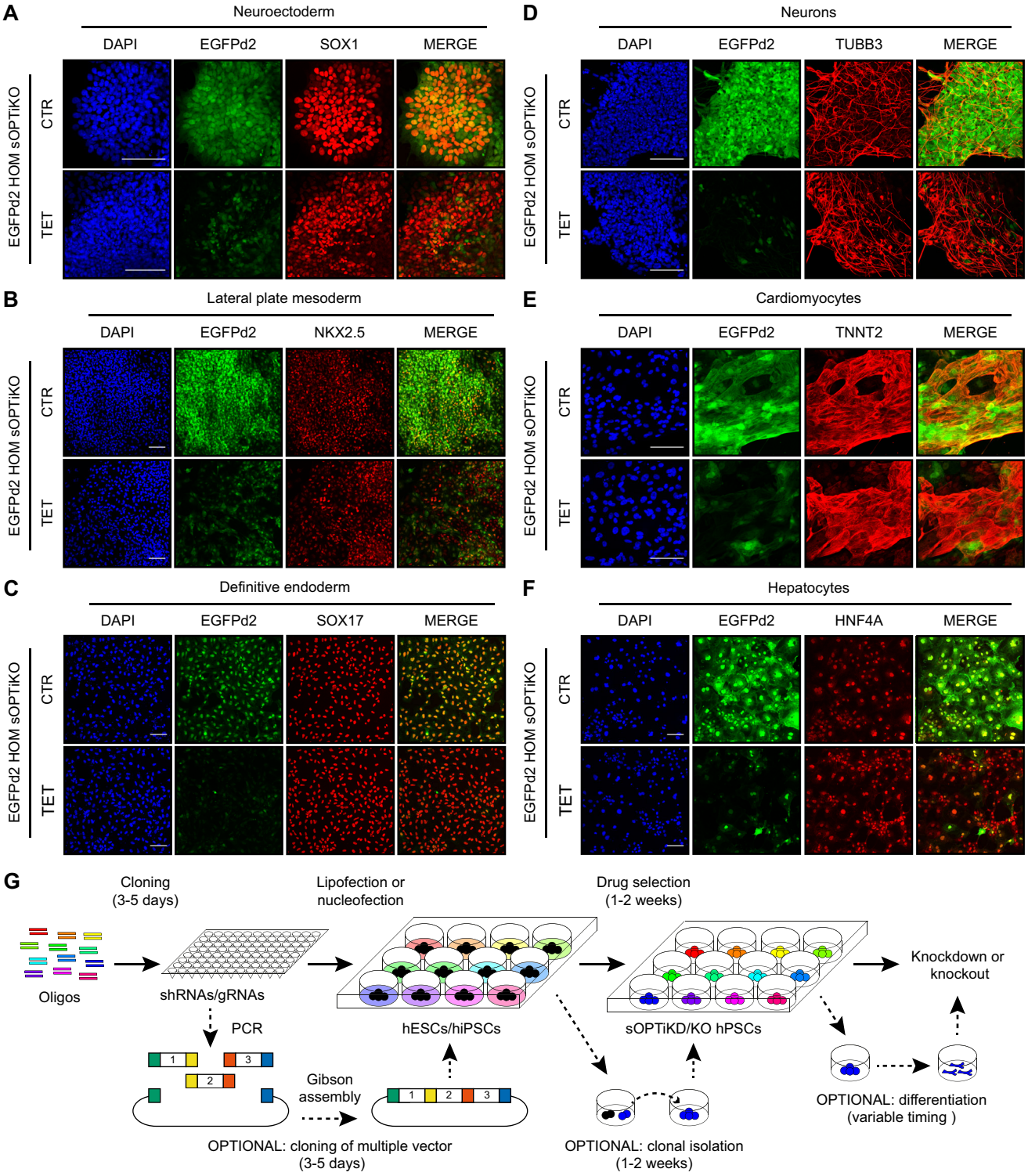


Fig. 8. Validation of the optimized inducible CRISPR/Cas9 platform following hPSC differentiation. (A-F) Representative immunofluorescent stainings for the indicated lineage-specific markers in cells derived from EGFPd2 sOPTiKO hESCs carrying a single EGFP inducible gRNA (gRNA 1). EGFPd2 knockout was induced with tetracycline for 6 days for the germ layers (A-C) and for 10 days for the mature cells (D-F). EGFPd2 fluorescence in control conditions (CTR) or after knockout (TET) is in green, and DAPI shows nuclear staining. Merged images of the EGFPd2 and lineage-specific markers are shown. Scale bars: 100 µm. (G) Summary of the experimental strategy behind the generation and application of inducible knockdown or knockout hPSCs using the sOPTiKD or sOPTiKO platforms.

tested its performance following differentiation. We differentiated homozygous EGFPd2 inducible knockout cells carrying a single copy of inducible EGFP gRNA into the three primary germ layers and into five cell types of clinical interest (neurons, cardiomyocytes, smooth muscle cells, hepatocytes and endocrine pancreatic cells). Immunostaining for lineage-specific markers demonstrated that

Development (2016) 143, 4405-4418 doi:10.1242/dev.138081

treatment with tetracycline resulted in strong loss of EGFPd2 expression (Fig. 8A-F, Fig. S9A,B) in all these cell types. Moreover, flow cytometry quantification confirmed that inducible knockout in differentiated cells was tightly controlled and efficient (Fig. S9C-H). For example, 85% of neuronal cells and 75% of smooth muscle cells completely lost EGFPd2 expression following tetracycline treatment (Fig. S9C,F). Considered together, these results validate that sOPTiKO allows efficient control of CRISPR/Cas9 activity not only in hPSCs, but also into a large panel of mature cell types (Fig. 8G).

DISCUSSION

This report describes sOPTiKD and sOPTiKO – two novel platforms for inducible knockdown or knockout of gene expression that address the limitations of previous methods. Compared with alternative approaches that rely on viral transduction or random integration of inducible shRNAs (Lambeth and Smith, 2013; Zafarana et al., 2009), sOPTiKD is simpler to use (plasmid based), quicker (2 weeks or less to generate stable lines following a single step of gene targeting by lipofection), more efficient (>95% of the resulting cells show inducible knockdown), more scalable (isolation of clonal sublines can be entirely bypassed) and significantly more robust (due to the use of GSHs and the lack of leakiness). Furthermore, sOPTiKO shares these same advantages, thus outperforming recent inducible CRISPR/Cas9 methods that rely on the conditional expression of Cas9 (González et al., 2014) or of a fusion protein between a catalytically inactive Cas9 and the transcriptional repressor KRAB (Mandegar et al., 2016). Indeed, these systems rely on the TRE promoter, which is heavily silenced upon hPSC differentiation into multiple lineages. Furthermore, these are lengthy two-step methods, and expression of the gRNA is achieved either by transient transfection, which can be poor in efficiency, or by random integration of the gRNA, which can result in mosaic expression. Finally, whereas CRISPR interference can only efficiently control gene promoter activity (Mandegar et al., 2016), sOPTiKO allows the deletion of a broader range of genomic targets, including regions outside of promoters that might not have a direct role in transcriptional regulation. Overall, sOPTiKD/KO are the first inducible shRNA and CRISPR/Cas9 technologies that enable streamlined functional studies of multiple genetic variants in hPSCs and in a diversity of differentiated cell types (Fig. 8G).

sOPTiKD and sOPTiKO each presents distinct advantages. On the one hand, the ability to control the level of knockdown using sOPTiKD allows the study of genes for which complete loss-of-function induces cell death, and facilitates the examination of gene dosage mechanisms. On the other hand, phenotypic studies following full gene knockout using sOPTiKO are more relevant in the case of genes that are still functional even when expressed at low levels. Moreover, sOPTiKO is applicable not only to genes, but also to non-coding genomic regulatory regions, which could represent a majority of disease-associated genetic traits (Cooper and Shendure, 2011).

Aside from the examples reported in this manuscript, we envision several other potential applications of the sOPTiKD/KO technologies. With regard to cellular and developmental biology, we anticipate that sOPTiKO could efficiently accommodate variants of the Cas9 gene with catalytically inactive domains (Dominguez et al., 2015). For instance, Cas9 fusion proteins with epigenetic modifiers could allow functional validations of putative genomic regulatory regions. Similarly, sOPTiKD could be repurposed to drive other types of inducible non-coding RNAs, such as antagomir or miRNA sponges to probe microRNA function (Ebert and Sharp, 2010). Remarkably, the high targeting efficiency and scalability of sOPTiKD/KO could allow high-throughput screenings by targeting inducible shRNA or gRNA pools. Compared with viral-based approaches (Chen et al., 2012), the isogenic integration of inducible shRNAs/gRNAs would reduce heterogeneity in the targeted population, hence increasing the screening sensitivity and specificity. With regard to disease modeling applications, sOPTiKD/KO could allow the simultaneous targeting of several hiPSC lines to probe gene function in different genetic backgrounds. Such an approach could facilitate the identification of genetic disease modifiers and the discovery of novel potential drug targets in the context of personalized medicine. Multiplex inducible gene knockdown or knockout could also be used to model complex genetic disorders. Finally, sOPTiKD/KO could be easily transferred to other cell types amenable to genetic manipulation, including established cell lines and adult stem cells (Drost et al., 2015; Mandal et al., 2014), thus allowing functional studies in a multitude of systems. In conclusion, we expect that sOPTiKD/KO technologies will have a broad impact on our ability to study human development, physiology and disease.

MATERIALS AND METHODS

hPSC culture and differentiation

Feeder- and serum-free hESC (H9 line; WiCell) and hiPSC ($A1AT^{R/R}$ line; Rashid et al., 2010) culture and differentiation were as previously described (Vallier, 2011). Details of media compositions and protocols are provided in the supplementary Materials and Methods.

Gene targeting

Sequences of all plasmids used in this study are provided in Appendix S1, and all cloning procedures and targeting experiments are described in detail in the supplementary Materials and Methods. Briefly, *AAVS1* targeting for OPTiKD and sOPTiKD was performed by lipofection, while *AAVS1* targeting for sOPTiKO was performed by nucleofection (Bertero et al., 2015; Vallier et al., 2004). Clonal lines were selected using 1 µg/ml puromycin (Sigma; for OPTiKD and sOPTiKD) or 25 µg/ml geneticin (G418 sulfate, Gibco) and 0.5 µg/ml puromycin (for sOPTiKO).

Inducible gene knockdown and knockout

Unless otherwise described in the results or figure legends, tetracycline hydrochloride (Sigma-Aldrich) was used at 1 µg/ml to induce gene knockdown or knockout. Refer to the supplementary Materials and Methods for details on the timing of *DPY30* inducible knockdown during hESC differentiation.

Analysis of RNA and protein expression

Quantitative real-time PCR (qPCR), western blot, flow cytometry and immunofluorescence were performed according to standard protocols as previously described (Bertero et al., 2015). Details, including the primer sequences and antibodies used, are provided in the supplementary Materials and Methods.

Statistical analysis

Statistical analyses were performed using GraphPad Prism 6. The type and number of replicates, the statistical test used, and the test results are described in the figure legends. All statistical tests employed were two-tailed. Unless stated otherwise in the figure legends, all graphical data are presented as mean±s.e.m. No experimental samples were excluded from the statistical analyses. Sample size was not pre-determined through power calculations, and no randomization or investigator blinding approaches were implemented during the experiments and data analyses. When representative results are presented, the experiments were reproduced in at least two independent cultures.

Acknowledgements

We thank Dr Kosuke Yusa for providing the AAVS1 ZFN plasmids and Dr Andrew Knights for technical support.

Competing interests
The authors declare no competing or financial interests.

Author contributions
A.B. conceived the study, designed, performed and analyzed experiments, and wrote the manuscript. M.P. and D.O. designed, performed and analyzed experiments. K.S., L.Y., M.C.d.B., S.B., W.G.B. and J.D.C. assisted experimental work. F.Se., E.G., L.G., N.F.R.H., D.I., F.Sa. and M.C.F.Z. (listed in alphabetical order) assisted with hESC differentiations into mature cell types. S.S. and M.K. supervised and supported the study. L.V. conceived, supervised and supported the study, wrote and gave final approval to the manuscript.

Funding
This work was supported by a European Research Council starting grant Relieve IMDs (281335 to L.V., D.O., N.R.F.H., M.C.F.Z., E.G.); the Cambridge University Hospitals National Institute for Health Research Biomedical Research Center (L.V., N.R.F.H., F.Sa.); the EU Seventh Framework Programme TISSUGEN (278955 to M.C.d.B.); the Wellcome Trust PhD program (PSAG/048 to L.Y.); a British Heart Foundation PhD Studentship (FS/11/77/39327 to A.B.); a research fellowship from the Deutsche Forschungsgemeinschaft [PA 2369/1-1 to M.P.]; and a core support grant from the Wellcome Trust and Medical Research Council to the Wellcome Trust – Medical Research Council Cambridge Stem Cell Institute (PSAG028). Deposited in PMC for immediate release.

Supplementary information
Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.138081.supplemental

References

Avior, Y., Sagi, I. and Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Mol. Cell Biol.* **17**, 170-182.
Bernardo, A. S., Faial, T., Gardner, L., Niakan, K. K., Ortmann, D., Senner, C. E., Callery, E. M., Trotter, M. W., Hemberger, M., Smith, J. C. et al. (2011). BRACHYURY and CDX2 mediate BMP-induced differentiation of human and mouse pluripotent stem cells into embryonic and extraembryonic lineages. *Cell Stem Cell* **9**, 144-155.
Bertero, A., Madrigal, P., Galli, A., Hubner, N. C., Moreno, I., Burks, D., Brown, S., Pedersen, R. A., Gaffney, D., Mendjan, S. et al. (2015). Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* **29**, 702-717.
Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. and Papaioannou, V. E. (1996). Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534-542.
Chen, T., Heller, E., Beronja, S., Oshimori, N., Stokes, N. and Fuchs, E. (2012). An RNA interference screen uncovers a new molecule in stem cell self-renewal and long-term regeneration. *Nature* **485**, 104-108.
Chen, Y., Cao, J., Xiong, M., Petersen, A. J., Dong, Y., Tao, Y., Huang, C. T.-L., Du, Z. and Zhang, S.-C. (2015). Engineering human stem cell lines with inducible gene knockout using CRISPR/Cas9. *Cell Stem Cell* **17**, 233-244.
Cheung, C., Bernardo, A. S., Trotter, M. W. B., Pedersen, R. A. and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat. Biotechnol.* **30**, 165-173.
Cooper, G. M. and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* **12**, 628-640.
David, R., Jarsch, V. B., Schwarz, F., Nathan, P., Gegg, M., Lickert, H. and Franz, W.-M. (2011). Induction of MesP1 by Brachyury(T) generates the common multipotent cardiovascular stem cell. *Cardiovasc. Res.* **92**, 115-122.
DeKelver, R. C., Choi, V. M., Moehle, E. A., Paschon, D. E., Hockemeyer, D., Meijsing, S. H., Sancak, Y., Cui, X., Steine, E. J., Miller, J. C. et al. (2010). Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* **20**, 1133-1142.
Dobrovolskaia-Zavadskaia, N. (1927). Regarding the spontaneous mortification of the tail of a new-born mouse and the existence of a hereditary characteristic (factor). *C. R. Seances Soc. Biol. Fil.* **97**, 114-116.
Dominguez, A. A., Lim, W. A. and Qi, L. S. (2015). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nat. Rev. Mol. Cell Biol.* **17**, 5-15.
Drost, J., van Jaarsveld, R. H., Ponsioen, B., Zimberlin, C., van Boxtel, R., Buijs, A., Sachs, N., Overmeer, R. M., Offerhaus, G. J., Begthel, H. et al. (2015). Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43-47.
Ebert, M. S. and Sharp, P. A. (2010). MicroRNA sponges: progress and possibilities. *RNA* **16**, 2043-2050.
Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum. Gene Ther.* **16**, 1241-1246.
Faial, T., Bernardo, A. S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G. E., Mascetti, V. L., Trotter, M. W. B., Smith, J. C. and Pedersen, R. A. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development* **142**, 2121-2135.
Fath, S., Bauer, A. P., Liss, M., Spriestersbach, A., Maertens, B., Hahn, P., Ludwig, C., Schäfer, F., Graf, M. and Wagner, R. (2011). Multiparameter RNA and codon optimization: a standardized tool to assess and enhance autologous mammalian gene expression. *PLoS ONE* **6**, e17596.
Gluecksohn-Schoenheimer, S. (1944). The development of normal and homozygous brachy (T/T) mouse embryos in the extraembryonic coelom of the chick. *Proc. Natl. Acad. Sci. USA* **30**, 134-140.
González, F., Zhu, Z., Shi, Z.-D., Lelli, K., Verma, N., Li, Q. V. and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215-226.
Haenebalcke, L., Goossens, S., Naessens, M., Kruse, N., Farhang Ghahremani, M., Bartunkova, S., Haigh, K., Pieters, T., Dierickx, P., Drogat, B. et al. (2013). Efficient ROSA26-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem Cell Rev.* **9**, 774-785.
Henriksen, J. R., Løkke, C., Hammerø, M., Geerts, D., Versteeg, R., Flaegstad, T. and Einvik, C. (2007). Comparison of RNAi efficiency mediated by tetracycline-responsive H1 and U6 promoter variants in mammalian cell lines. *Nucleic Acids Res.* **35**, e67.
Herbst, F., Ball, C. R., Tuorto, F., Nowrouzi, A., Wang, W., Zavidij, O., Dieter, S. M., Fessler, S., van der Hoeven, F., Kloz, U. et al. (2012). Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo. *Mol. Ther.* **20**, 1014-1021.
Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R. C., Katibah, G. E., Amora, R., Boydston, E. A., Zeitler, B. et al. (2009). Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851-857.
Inman, K. E. and Downs, K. M. (2006). Brachyury is required for elongation and vasculogenesis in the murine allantois. *Development* **133**, 2947-2959.
Irion, S., Luche, H., Gadue, P., Fehling, H. J., Kennedy, M. and Keller, G. (2007). Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat. Biotechnol.* **25**, 1477-1482.
Jiang, H., Shukla, A., Wang, X., Chen, W., Bernstein, B. E. and Roeder, R. G. (2011). Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell* **144**, 513-525.
King, T., Beddington, R. S. P. and Brown, N. A. (1998). The role of the brachyury gene in heart development and left-right specification in the mouse. *Mech. Dev.* **79**, 29-37.
Lambeth, L. S. and Smith, C. A. (2013). Short hairpin RNA-mediated gene silencing. *Methods Mol. Biol.* **942**, 205-232.
Li, X., Zhao, X., Fang, Y., Jiang, X., Duong, T., Fan, C., Huang, C.-C. and Kain, S. R. (1998). Generation of destabilized green fluorescent protein as a transcription reporter. *J. Biol. Chem.* **273**, 34970-34975.
Mandal, P. K., Ferreira, L. M. R., Collins, R., Meissner, T. B., Boutwell, C. L., Friesen, M., Vrbanac, V., Garrison, B. S., Stortchevoi, A., Bryder, D. et al. (2014). Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643-652.
Mandegar, M. A., Huebsch, N., Frolov, E. B., Shin, E., Truong, A., Olvera, M. P., Chan, A. H., Miyaoka, Y., Holmes, K., Spencer, C. I. et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* **18**, 541-553.
Martin, B. L. (2015). Factors that coordinate mesoderm specification from neuromesodermal progenitors with segmentation during vertebrate axial extension. *Semin. Cell Dev. Biol.* **49**, 59-67.
Mendjan, S., Mascetti, V. L., Ortmann, D., Ortiz, M., Karjosukarso, D. W., Ng, Y., Moreau, T. and Pedersen, R. A. (2014). NANOG and CDX2 pattern distinct subtypes of human mesoderm during exit from pluripotency. *Cell Stem Cell* **15**, 310-325.
Ordovás, L., Boon, R., Pistoni, M., Chen, Y., Wolfs, E., Guo, W., Sambathkumar, R., Bobis-Wozowicz, S., Helsen, N., Vanhove, J. et al. (2015). Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition. *Stem Cell Rep.* **5**, 918-931.
Papaioannou, V. E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. *Development* **141**, 3819-3833.
Pauklin, S. and Vallier, L. (2013). The cell-cycle state of stem cells determines cell fate propensity. *Cell* **155**, 135-147.
Pauklin, S., Madrigal, P., Bertero, A. and Vallier, L. (2016). Initiation of stem cell differentiation involves cell cycle-dependent regulation of developmental genes by Cyclin D. *Genes Dev.* **30**, 421-433.
Pourquié, O., Bruneau, B., Keller, G. and Smith, A. (2015). Looking inwards: opening a window onto human development. *Development* **142**, 1-2.
Qin, J. Y., Zhang, L., Clift, K. L., Hulur, I., Xiang, A. P., Ren, B.-Z. and Lahn, B. T. (2010). Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS ONE* **5**, e10611.
Ramachandra, C. J. A., Shahbazi, M., Kwang, T. W. X., Choudhury, Y., Bak, X. Y., Yang, J. and Wang, S. (2011). Efficient recombinase-mediated cassette

exchange at the AAVS1 locus in human embryonic stem cells using baculoviral vectors. *Nucleic Acids Res.* **39**, e107.

Rashid, S. T., Corbineau, S., Hannan, N., Marciniak, S. J., Miranda, E., Alexander, G., Huang-doran, I., Griffin, J., Ahrlund-richter, L., Skepper, J. et al. (2010). Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127-3136.

Raya, A., Rodríguez-Pizà, I., Guenechea, G., Vassena, R., Navarro, S., Barrero, M. J., Consiglio, A., Castellà, M., Río, P., Sleep, E. et al. (2009). Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells. *Nature* **460**, 53-59.

Sadelain, M., Papapetrou, E. P. and Bushman, F. D. (2012). Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer* **12**, 51-58.

Shearwin, K. E., Callen, B. P. and Egan, J. B. (2005). Transcriptional interference–a crash course. *Trends Genet.* **21**, 339-345.

Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A. and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875-889.

Touboul, T., Hannan, N. R. F., Corbineau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J. et al. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754-1765.

Vallier, L. (2011). Serum-free and feeder-free culture conditions for human embryonic stem cells. *Springer Protoc.* **690**, 57-66.

Vallier, L., Rugg-Gunn, P. J., Bouhon, I. A., Andersson, F. K., Sadler, A. J. and Pedersen, R. A. (2004). Enhancing and diminishing gene function in human embryonic stem cells. *Stem Cells* **22**, 2-11.

Wahlers, A., Schwieger, M., Li, Z., Meier-Tackmann, D., Lindemann, C., Eckert, H.-G., von Laer, D. and Baum, C. (2001). Influence of multiplicity of infection and protein stability on retroviral vector-mediated gene expression in hematopoietic cells. *Gene Ther.* **8**, 477-486.

Wang, Z., Oron, E., Nelson, B., Razis, S. and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440-454.

Wright, A. V., Nuñez, J. K. and Doudna, J. A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell* **164**, 29-44.

Yang, Z., Augustin, J., Chang, C., Hu, J., Shah, K., Chang, C.-W., Townes, T. and Jiang, H. (2014). The DPY30 subunit in SET1/MLL complexes regulates the proliferation and differentiation of hematopoietic progenitor cells. *Blood* **124**, 2025-2022.

Yao, S., Sukonnik, T., Kean, T., Bharadwaj, R. R., Pasceri, P. and Ellis, J. (2004). Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. *Mol. Ther.* **10**, 27-36.

Zafarana, G., Avery, S. R., Avery, K., Moore, H. D. and Andrews, P. W. (2009). Specific knockdown of OCT4 in human embryonic stem cells by inducible short hairpin RNA interference. *Stem Cells* **27**, 776-782.

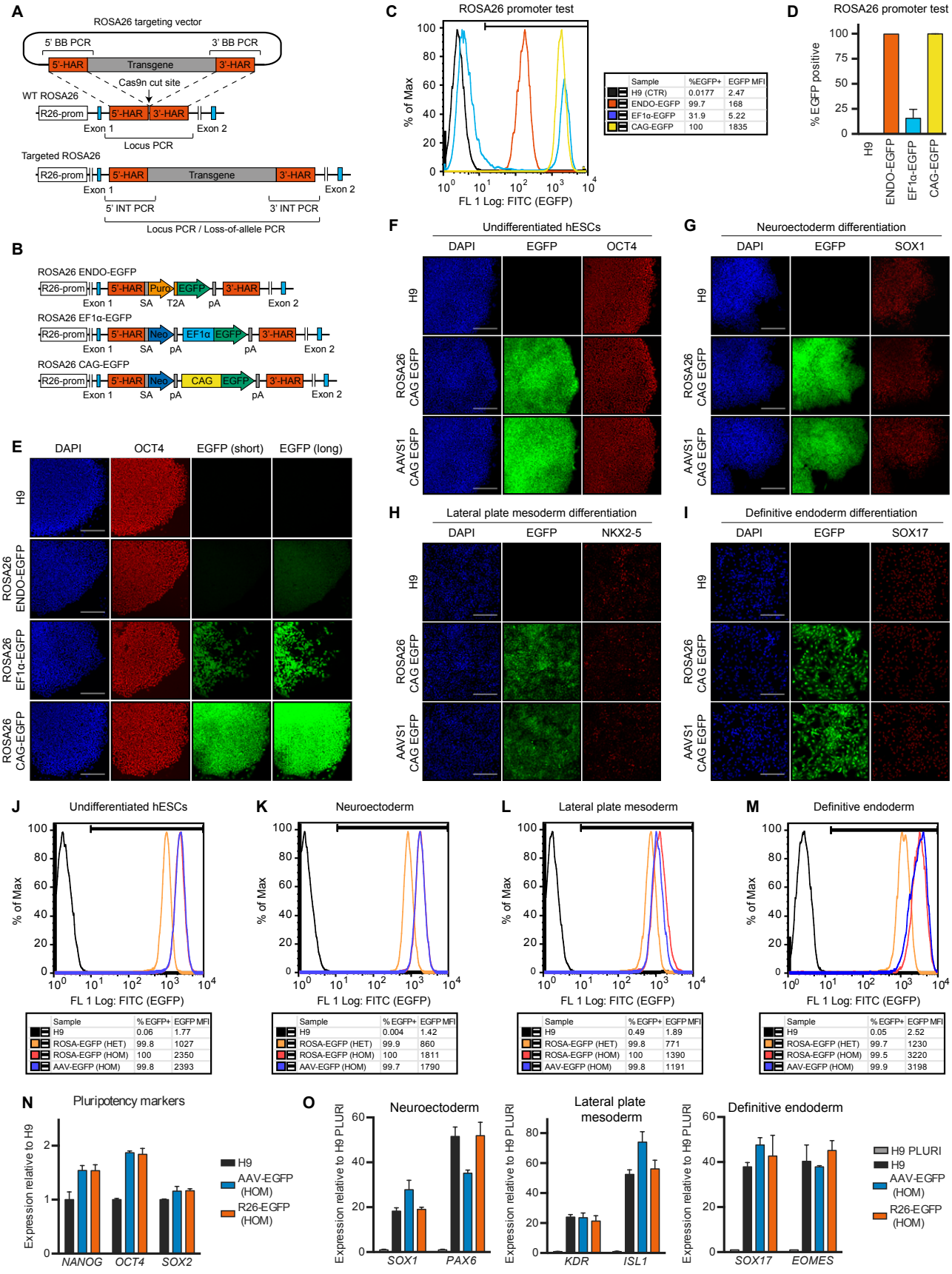
Zhu, Z. and Huangfu, D. (2013). Human pluripotent stem cells: an emerging model in developmental biology. *Development* **140**, 705-717.

Supplementary Information

Optimized inducible shRNA and CRISPR/Cas9 platforms for *in vitro* studies of human development using hPSCs

Bertero A. et al.,

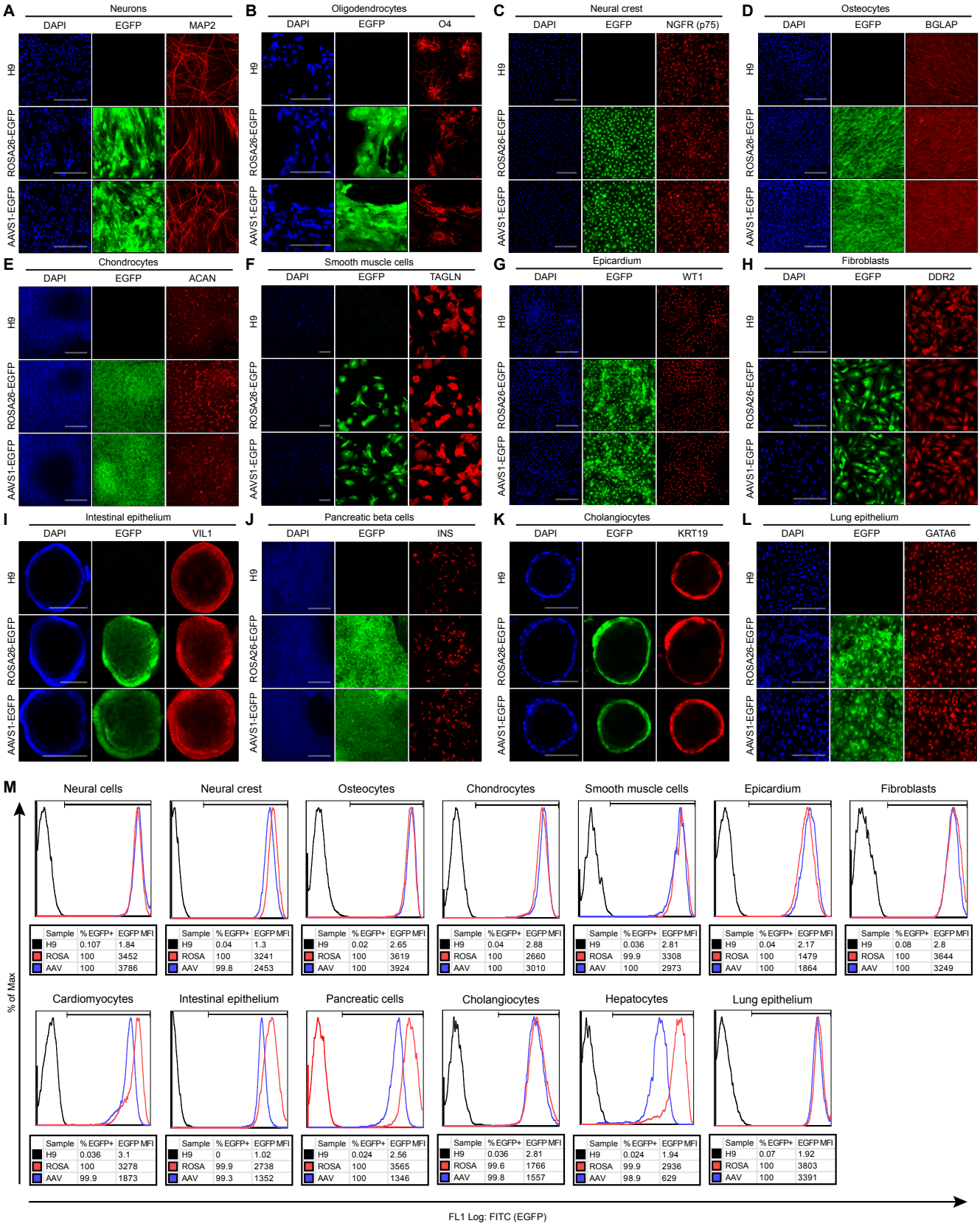
Supplementary Figures



Supplementary Figure 1. Generation of ROSA26 and AAVS1 EGFP reporter hESCs.

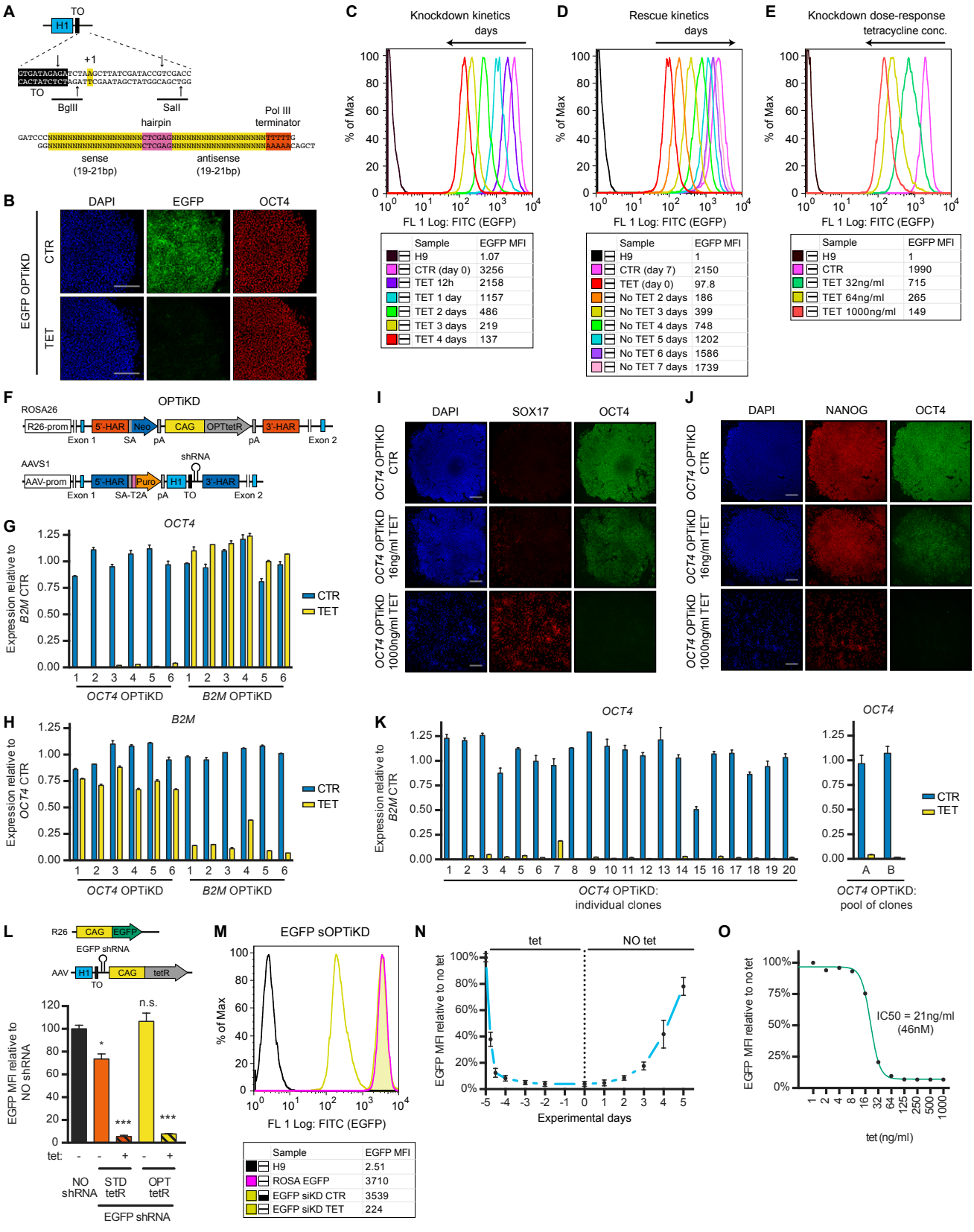
(A) Schematic of the ROSA26 targeting approach and of the genotyping strategies used to identify correctly targeted lines (see Supplementary Methods for additional details). Cas9n: D10A nickase mutant Cas9 endonuclease from *S. Pyogenes*.

R26-prom: ROSA26 locus promoter (THUMPD3-AS1 gene); 5'-HAR/3'-HAR: upstream/downstream homology arm; Transgene: region integrated following gene targeting; Locus PCR: PCR product of wild-type ROSA26 locus (indicating a non-targeted allele); Locus PCR/Loss-of-allele: PCR product of targeted allele/PCR that fails if the transgene contains the GC-rich CAG promoter (indicative of expected transgene targeting); 5' INT/3' INT PCR: PCR product of transgene 5'-end/3'-end integration region (indicative of expected transgene targeting); 5' BB/3' BB PCR: PCR product of vector backbone 5'-end/3'-end (indicative of non-specific off-target plasmid integration). Note that similar targeting and genotyping strategies were applied for the AAVS1 locus targeting. **(B)** Schematic of the ROSA26 transgenic alleles generated to test the best strategy for constitutive EGFP (enhanced green fluorescent protein) expression. ENDO-EGFP: EGFP driven by the endogenous ROSA26 promoter (R26-prom; targeting vector pR26-Puro_ENDO-EGFP); EF1α-EGFP: EGFP driven by the elongation factor 1α promoter (targeting vector pR26-Neo_EF1α-EGFP); CAG-EGFP: EGFP driven by the CAG promoter (targeting vector pR26-Neo_CAG-EGFP); SA: splice acceptor; Puro: puromycin resistance (puromycin *N*-acetyltransferase); Neo: neomycin resistance (neomycin phosphotransferase II); pA: polyadenylation signal. **(C)** Flow cytometry quantification of the percentage of EGFP positive cells (EGFP+; the gate is shown), and of the EGFP median fluorescence intensity (MFI) in representative ROSA26-EGFP reporter hESC clonal lines, or wild-type H9 hESCs. **(D)** Percentage of EGFP positive cells in ROSA26-EGFP reporter hESCs. Results are for 3 clones with heterozygous ROSA26 targeting per condition (see Table S1). **(E)** Representative images showing EGFP fluorescence for the same ROSA26-EGFP reporter hESC clonal lines from panel C. Cells were fluorescently immunostained for the pluripotency marker OCT4 (red), and DAPI (blue) shows nuclear staining. Scale bars: 200 μm. **(F-I)** Representative immunofluorescence of the indicated lineage-specific markers (red) in undifferentiated hESCs (F) and in the three germ layers (G-I) obtained from wild-type H9 hESCs and homozygous ROSA26 and AAVS1 CAG-EGFP reporter hESCs (targeting vectors pR26-Neo_CAG-EGFP and pAAV-Puro_CAG-EGFP, respectively). EGFP fluorescence is in green, and DAPI (blue) shows nuclear staining. Scale bars: 200 μm. **(J-M)** Representative flow cytometry EGFP quantifications in ROSA26 (ROSA) and AAVS1 (AAV) CAG-EGFP reporter lines in undifferentiated conditions or following germ layer differentiation. H9 hESCs (H9) were used as negative control. HET: heterozygous targeting; HOM: homozygous targeting. **(N-O)** Characterization of the differentiation potential of ROSA26 (R26) and AAVS1 (AAV) CAG-EGFP reporter hESCs by qPCR for lineage-specific markers in undifferentiated cells (N) and in the three germ layers (O). H9 PLURI: undifferentiated H9 hESCs. Results are for 2 independent cultures per condition.



Supplementary Figure 2. Validation of the ROSA26 and AAVS1 loci as *bona fide* genomic safe harbors.
(A-L) Representative immunofluorescence of the indicated lineage-specific markers (red) in mature cells derived from ROSA26 and AAVS1 homozygous CAG-EGFP reporter hESCs, or wild-type H9 hESCs. EGFP fluorescence is in green, and DAPI (blue) shows nuclear staining. Scale bars: 100 μm for oligodendrocytes, intestinal epithelium and cholangiocytes, and 200 μm for all other lineages. **(M)** Representative flow cytometry quantification of the percentage of EGFP positive cells

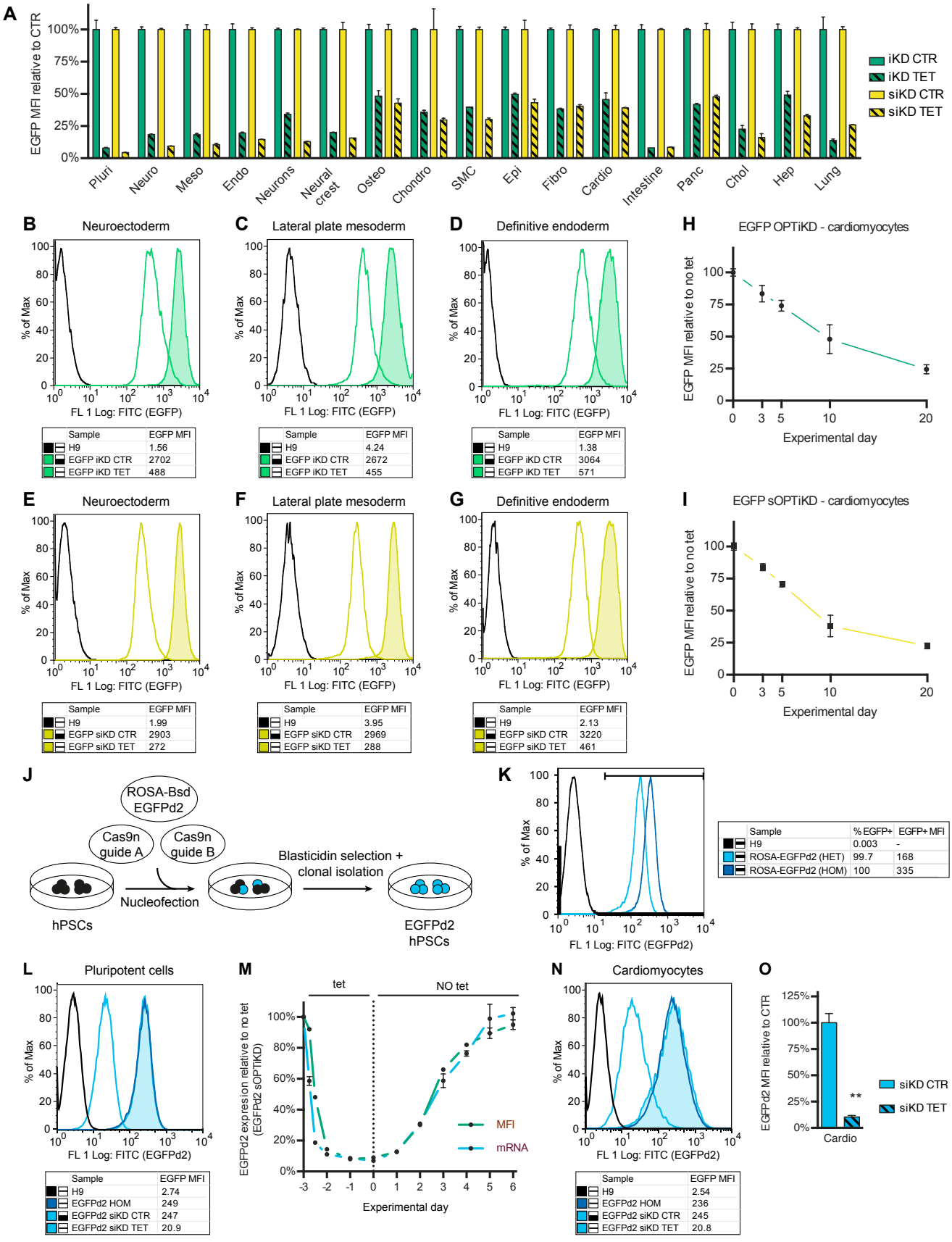
(EGFP+; the gate is indicated), and of the EGFP median fluorescence intensity (MFI) in mature cells derived from ROSA26 (ROSA) and AAVS1 (AAV) homozygous CAG-EGFP reporter hESCs, or wild-type H9 hESCs.



Supplementary Figure 3. Development of optimized inducible knockdown systems based on double- or single-step GSH targeting.

(A) Schematic of the cloning strategy to generate AAVS1 targeting vectors carrying inducible shRNAs. The vector contains a H1 Pol III promoter containing a single tet operon (TO). Following digestion with BglII and SalI, a double-stranded fragment is ligated to introduce the desired shRNA sequence. This fragment is obtained by annealing two single-stranded oligonucleotides, and includes the shRNA proper (sense target sequence, hairpin, and antisense target sequence) followed by

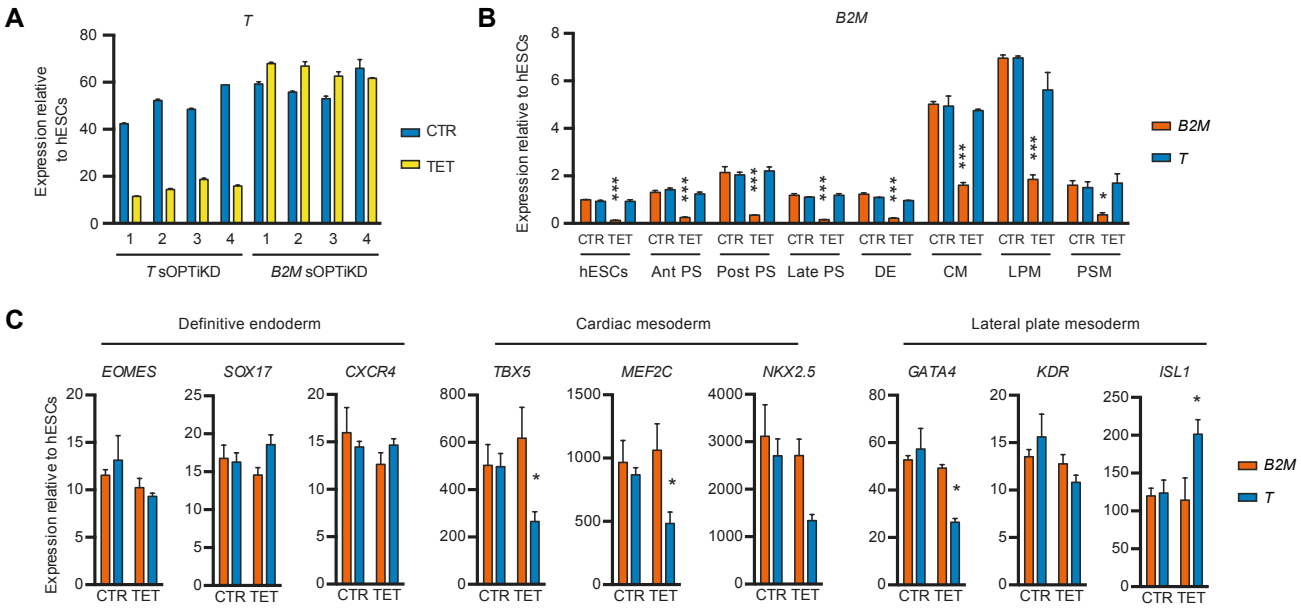
a Pol III terminator. **(B)** Representative images of EGFP fluorescence (green) in EGFP OPTimized inducible KnockDown (OPTiKD) hESCs (cells sequentially targeted with pR26-Neo_CAG-OPTtetR and pAAV-Puro_EGFPiKD, see Fig. 2B) in absence (CTR) or presence (TET) of tetracycline for 5 days. Cells were fluorescently immunostained for the pluripotency marker OCT4 (red), and DAPI (blue) shows nuclear staining. Scale bars: 200 μm. **(C-D)** Representative flow cytometry EGFP quantifications during knockdown and rescue kinetics in EGFP OPTiKD hESCs. Cells were exposed to tetracycline for 5 days, followed by tetracycline withdrawal for 7 days (day 5 of TET was considered as day 0 of the rescue kinetics). Samples were collected at the indicated time points. CTR: cells maintained in the absence of tetracycline throughout the experiment and collected at the indicated experimental day. Wild-type H9 hESCs were used as negative control. **(E)** Representative flow cytometry quantifications of EGFP knockdown dose-response to tetracycline in EGFP OPTiKD hESCs. Cells were cultured for 5 days with the indicated doses of tetracycline. CTR: no tetracycline. **(F)** Strategy for the generation of OPTiKD hPSCs by dual targeting of the ROSA26 and AAVS1 genomic safe harbors. Targeting vectors: pR26-Neo_CAG-OPTtetR; pAAV-Puro_iKD. R26-prom: ROSA26 locus promoter; AAV-prom: AAVS1 locus promoter; 5'-HAR/3'-HAR: upstream/downstream homology arm; Neo: neomycin resistance; Puro: puromycin resistance; SA: splice acceptor; T2A: self-cleaving T2A peptide; pA: polyadenylation signal; CAG: CMV early enhancer, chicken β-actin and rabbit β-globin hybrid promoter; OPTtetR: codon-optimized tetracycline repressor; H1: H1 Pol III promoter; TO: tet operon; shRNA: short hairpin RNA. **(G-H)** qPCR validation of *OCT4* and *B2M* (negative control gene) OPTiKD hESCs (AAVS1 targeting vectors pAAV-Puro_iKD-OCT4 and pAAV-Puro_iKD-B2M) in absence (CTR) or presence of tetracycline for 5 days (TET). Individual clones were analyzed in duplicate. The expression is shown as normalized on the average level in *B2M* or *OCT4* OPTiKD hESCs in control conditions, as indicated. **(I-J)** Immunofluorescence for the pluripotency genes OCT4 and NANOG, and of the definitive endoderm marker SOX17 in undifferentiated *OCT4* OPTiKD lines maintained in absence of tetracycline (CTR) or cultured for 5 days with different doses of tetracycline that induced intermediate (16 ng ml^{-1}) or full (1000 ng ml^{-1}) *OCT4* knockdown. DAPI (blue) shows nuclear staining. Scale bars: 200 μm. **(K)** qPCR validation of *OCT4* OPTiKD hESCs in clonally- or non clonally-isolated gene targeted cells. Individual clones or pools were analyzed in duplicate. The expression is normalized on the average level in *B2M* OPTiKD hESCs in control conditions. **(L)** Quantification of EGFP mean fluorescence intensity (MFI) in EGFP inducible knockdown cells generated with a single gene targeting step. Homozygous ROSA26 CAG-EGFP hESCs were re-targeted in the AAVS1 locus with an inducible EGFP shRNA and CAG-tetR cassette (homozygous targeting). AAVS1 targeting vectors used: pAAV-Puro_EGFPsiKD-STD (STDtetR, or STD); pAAV-Puro_EGFPsiKD-OPT (OPTtetR, or OPT). Cells were analyzed in absence (-) or presence (+) of tetracycline for 5 days. EGFP levels were compared to those in homozygous ROSA26 CAG-EGFP hESCs (NO shRNA). Results are from 2-3 individual lines per condition (see Table S1). n.s.=$p>0.05$ (non-significant), *=$p<0.05$, ***=$p<0.001$ VS NO shRNA (ANOVA with post-hoc Holm-Sidak comparisons). The condition expressing the OPTtetR was named EGFP single-step OPTtimized inducible KnockDown (sOPTiKD). **(M)** Representative flow cytometry EGFP quantification in EGFP sOPTiKD hESCs treated as described in panel L. **(N)** EGFP mRNA knockdown and rescue kinetics in EGFP sOPTiKD hESCs. Results are from 2 independent cultures per time point. **(O)** Tetracycline dose-response curve for EGFP knockdown in EGFP sOPTiKD hESCs. The half-maximal inhibitory concentration (IC50) is reported. Results are from 2 independent cultures per dose, and the mean is shown.



Supplementary Figure 4. Validation of EGFP OPTiKD and sOPTiKD in hESCs-derived cells.

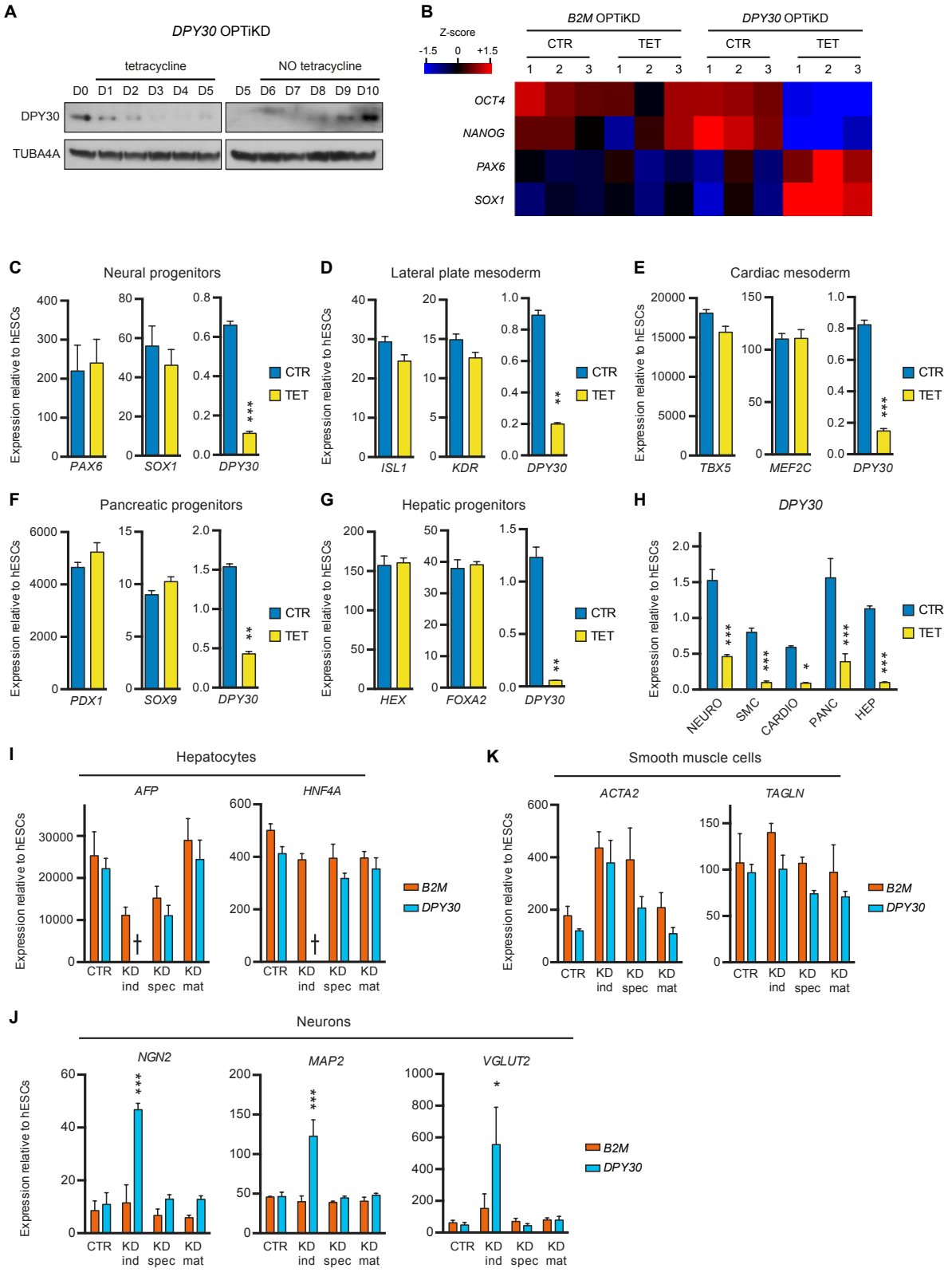
(A) EGFP expression measured by flow cytometry in absence (CTR) or presence of tetracycline for 5 days (TET) in the indicated cell types derived from EGFP OPTiKD (iKD) or sOPTiKD (siKD) hESCs. EGFP levels are reported relative to control conditions in the same line for each individual lineage. Pluri: undifferentiated hESCs; Neuro: neuroectoderm; Meso:

lateral plate mesoderm; Endo: definitive endoderm; Osteo: osteocytes; Chondro: chondrocytes; SMC: smooth muscle cells; Epi: epicardium; Fibro: cardiac fibroblasts; Intestine: intestinal organoids; Panc: pancreatic cells; Chol: cholangiocytes; Hep: hepatocytes. **(B-G)** Representative flow cytometry EGFP quantification in the three germ layers derived from EGFP OPTiKD (iKD) and sOPTiKD (siKD) hESCs, and cultured in absence or presence of tetracycline for 5 days. H9 hESCs were used as a negative control. **(H-I)** Kinetics of EGFP loss following prolonged treatment with tetracycline of cardiomyocytes derived from EGFP OPTiKD and sOPTiKD hESCs. (**J**) Experimental approach to generate destabilized EGFP (EGFPd2) ROSA26 reporter hPSCs (targeting vector: pR26-Bsd_CAG-EGFPd2). (**K**) Representative flow cytometry EGFP quantification in heterozygous (HET) and homozygous (HOM) ROSA26 CAG-EGFPd2 reporter hESCs. H9 hESCs (H9) were used as negative control. (**L**) Representative flow cytometry EGFP quantification in homozygous ROSA26 CAG-EGFPd2 hESCs re-targeted with pAAV-Puro_EGFPsiKD-OPT (EGFPd2 sOPTiKD, homozygous targeting). Cells were cultured in absence (CTR) or presence of tetracycline for 5 days (TET). Homozygous ROSA26 CAG-EGFPd2 hESCs and wild-type H9 hESCs were used as positive and negative controls, respectively. (**M**) EGFP knockdown and rescue kinetics in EGFPd2 sOPTiKD hESCs measured by flow cytometry (MFI) and qPCR (mRNA). (**N-O**) Flow cytometry quantification of EGFPd2 expression in cardiomyocytes generated from EGFPd2 sOPTiKD hESCs and cultured in absence (CTR) or presence of tetracycline for 5 days (TET). *=$p<0.05$ (t-test). For all the graphs in this figure results are from two independent cultures per condition.



Supplementary Figure 5. *T* sOPTiKD in hESCs and during mesendoderm differentiation.

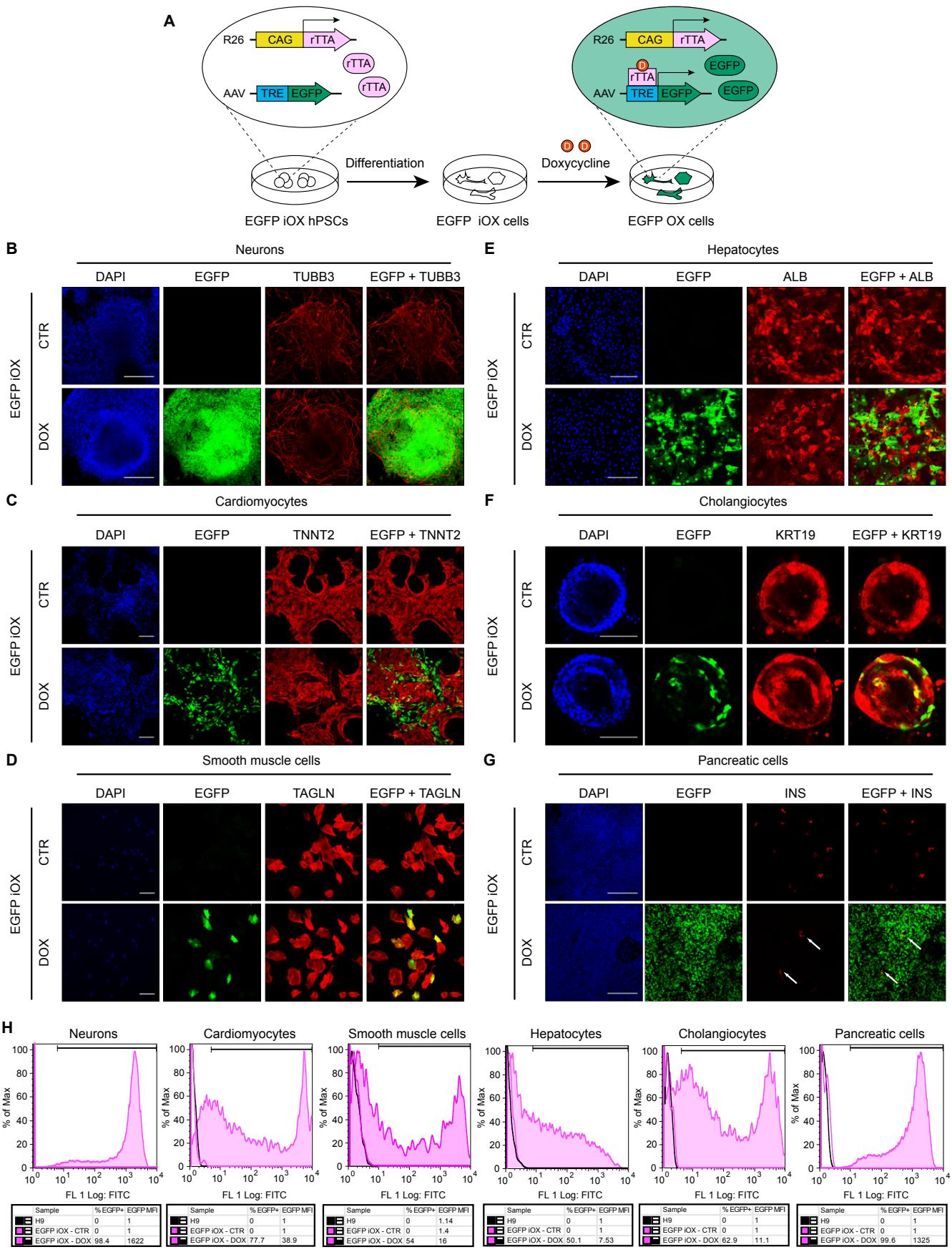
(A) qPCR validation of *T* sOPTiKD hESCs (AAVS1 targeting vector pAAV-Puro_iKD-T). Cells were cultured in absence (CTR) or presence (TET) of tetracycline both for 2 days in the pluripotent state and during differentiation into late primitive streak cells (Fig. 5A). Individual clones were analyzed in duplicate, and the expression is shown as normalized on the average level in undifferentiated hESCs. *B2M* sOPTiKD hESCs were compared as negative controls. (**B**) qPCR for *B2M* in the indicated lineages derived from *T* or *B2M* control sOPTiKD hESCs. Knockdown was induced for two days in the pluripotent state and maintained throughout the differentiation (Fig. 5A,B; CTR: no knockdown; TET: knockdown; Ant/Post/Late PS: anterior/posterior/late primitive streak; DE: definitive endoderm; CM: cardiac mesoderm; LPM: lateral plate mesoderm; PSM: presomitic mesoderm). *=p<0.05, ***=p<0.001 VS *T* in the same condition (2-way ANOVA with post-hoc Sidak comparisons), and results are from 3 independent clonal lines per condition. (**C**) qPCR results for lineage-specific markers in the indicated cell types from the experiment described in B. *=p<0.05 VS *B2M* in the same condition (2-way ANOVA with post-hoc Sidak comparisons), and results are from 3 independent clonal lines per condition.



Supplementary Figure 6. *DPY30* OPTiKD in hESCs and during hESCs differentiation.

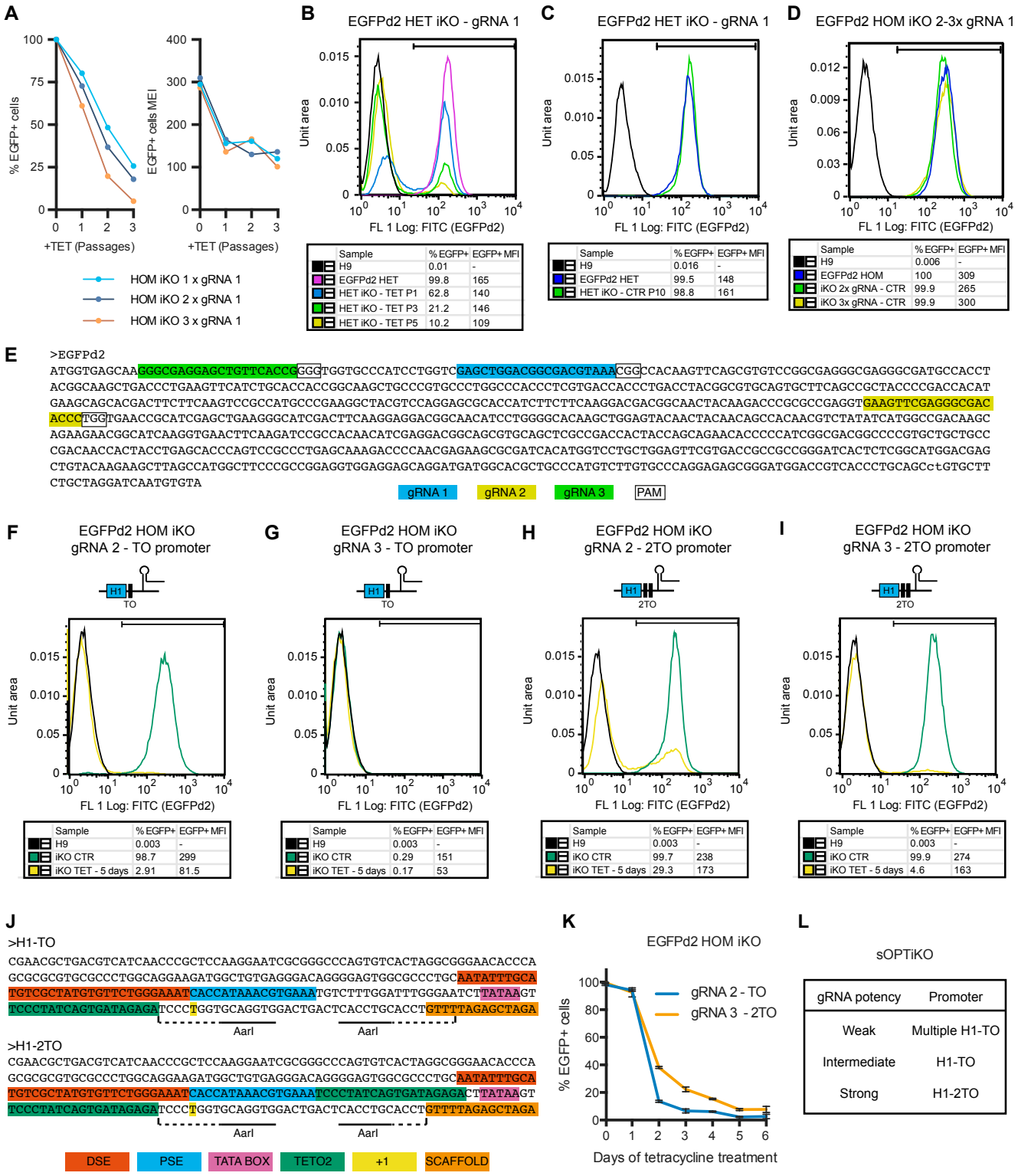
(A) Western blots to determine the kinetics of DPY30 protein knockdown and rescue in *DPY30* OPTiKD hESCs. Cells were exposed to tetracycline for 5 days, followed by tetracycline withdrawal for 5 days. Samples were collected at the indicated days (D1-D10). D0: cells maintained in the absence of tetracycline and collected at the start of the experiment. TUBA4A (α-tubulin): loading control. **(B)** Heatmap summarizing qPCR results for *DPY30* and *B2M* OPTiKD hESCs in control conditions (CTR) or following 10 days of knockdown (TET). Z-scores indicate differential expression measured in number of standard deviations from the average level. Three independent cultures per condition were analyzed, as indicated. **(C-G)** qPCR for *DPY30* and the indicated lineage specific markers in maturing progenitors during specification of *DPY30* OPTiKD hESCs

into differentiated cell types (see Fig. 6A). Cells were cultured in the absence (CTR) or presence of tetracycline (TET) from the third day before being analyzed at the indicated developmental stages (knockdown specifically in maturing progenitors). **(H)** *DPY30* expression in differentiated cell types derived from *DPY30* OPTiKD hESCs. Maturing cells were cultured in the absence or presence of tetracycline for 7 days before being analyzed. Refer to panels I-K and Fig. 6C,D for the expression of lineage markers specific for the cell types analyzed. NEURO: neurons; SMC: smooth muscle cells; CARDIO: cardiomyocytes; PANC: pancreatic endocrine cells; HEP: hepatocytes. For panels C-H: *=p<0.05; **=p<0.01; ***=p<0.001 VS CTR (t-test). **(I-K)** qPCR-based phenotypic analyses of *DPY30* and *B2M* (negative control gene) OPTiKD hESCs after differentiation into the indicated mature lineages. Cross symbols indicate that cells died during the differentiation and could not be analyzed by qPCR. *=p<0.05; ***=p<0.001 vs *B2M* in the same condition (2way ANOVA with post-hoc Sidak comparisons). CTR: no knockdown; KD ind/spec/mat: knockdown from the induction, specification or maturation (see Fig. 6A). For panels C-K, results are from 3 independent cultures per condition, and the expression is reported relative to the average level in undifferentiated hESCs.



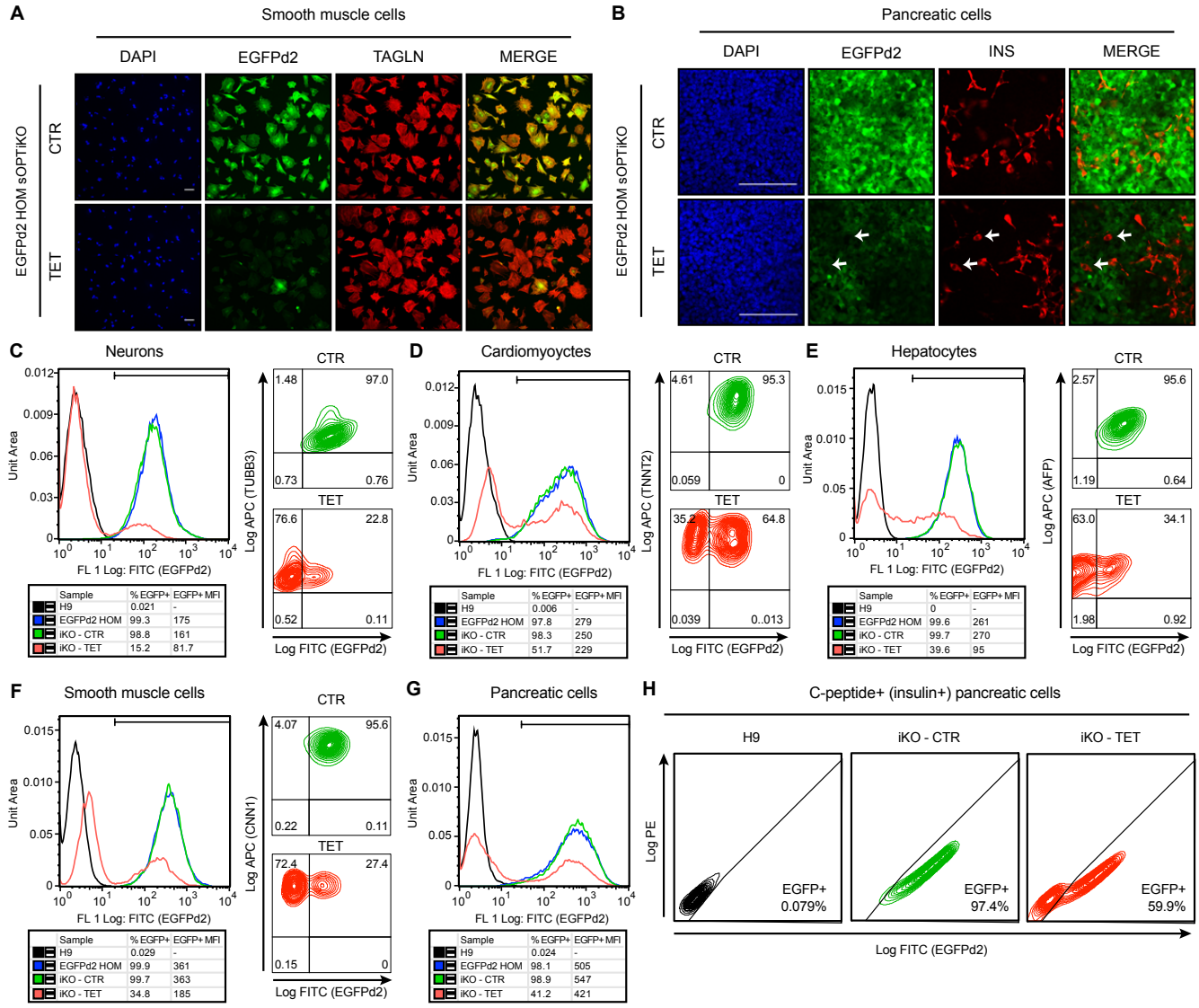
Supplementary Figure 7. Efficiency of TRE-mediated inducible gene overexpression in hESCs-derived mature cells.
(A) Schematic of the experimental approach. EGFP inducible overexpression (iOX) hPSCs were generated by two gene targeting steps: (1) in the ROSA26 (R26) locus, homozygous targeting of CAG-driven constitutive reverse tetracycline-

controlled TransActivator (rtTA) protein; (2) in the AAVS1 locus, homozygous targeting of tet-responsive element (TRE)-driven EGFP transgene (described in MP, DO, AB, et al., manuscript in revision). In the absence of the drug doxycycline, the rtTA is transcriptionally inactive and EGFP is not expressed. iOX hESCs were differentiated into several types of mature cells (see panels B-G), at which point cells were treated with 1 μg ml^{-1} doxycycline for 5 days. This induces transcriptional activity of the rtTA and EGFP overexpression (OX). **(B-G)** Representative immunofluorescent stainings for lineage specific markers (red) in the indicated cell types derived from EGFP iOX hESCs and cultured in absence (CTR) or presence of doxycycline (DOX) for 5 days. EGFP fluorescence is in green, and DAPI (blue) shows nuclear staining. Merged images facilitate identification of cells where inducible EGFP overexpression failed due to silencing of the TRE promoter. Scale bars: 100 μm for cholangiocytes, and 200 μm for all other lineages. **(H)** Representative EGFP quantification by flow cytometry in the same cells described for panels B-G. Wild-type H9 hESCs were used as negative controls. The percentage of EGFP positive cells (%EGFP+, the gates are shown) and the EGFP median fluorescence intensity (MFI) are reported. Heavy silencing of the inducible EGFP transgene is evident in cardiomyocytes, smooth muscle cells, hepatocytes, and cholangyocytes, while inducible overexpression is only moderately heterogeneous in neurons and pancreatic cells.



Supplementary Figure 8. Development of an optimized inducible CRISPR/Cas9 knockout platform in hESCs.
(**A**) Expression of EGFPd2 in homozygous (HOM) ROSA26-EGFPd2 hESCs gene targeted with EGFP inducible knockout systems (iKO, see Fig. 7) expressing one, two or three copies of the EGFP inducible gRNA 1 (targeting vectors: pAAV-Neo_CAG-Cas9 in combination with either pAAV-Puro_siKO-EGFP-1, pAAV-Puro_MsiKO-EGFP-1x2, or pAAV-Puro_MsiKO-EGFP-1x3, respectively). Cells were analyzed following tetracycline (TET) treatment for the indicated number of passages (every 5 days). The percentage of EGFP positive cells (%EGFP+) and the median fluorescence intensity (MFI) of EGFP positive cells are reported. (**B**) Representative flow cytometry for EGFPd2 expression in heterozygous ROSA26-EGFPd2 sOPTiKO hESCs (iKO) following the indicated number of passages in the presence of tetracycline (TET). EGFPd2 heterozygous cells not carrying the inducible CRISPR/Cas9 system (EGFPd2 HET) and wild-type hESCs (H9) were

analyzed as positive and negative controls for EGFPd2 expression, respectively. (**C**) As in panel B, but EGFPd2 heterozygous sOPTiKO hESCs were analyzed following ten passages in absence of tetracycline to monitor the potential leakiness of the system. (**D**) As in C, but EGFPd2 homozygous sOPTiKO hESCs expressing 2 or 3 inducible gRNAs were analyzed. (**E**) EGFPd2 coding sequence showing the location of the three EGFP gRNAs tested in this study. (**F-I**) Representative flow cytometry for EGFPd2 expression in EGFPd2 homozygous sOPTiKO hESCs carrying the indicated combinations of gRNA (2 or 3) and inducible promoter (TO or 2TO, see panel J). Targeting vectors: pAAV-Puro_siKO-EGFP-2 (F), pAAV-Puro_siKO-2TO-EGFP-2 (G), pAAV-Puro_siKO-EGFP-3 (H), pAAV-Puro_siKO-2TO-EGFP-3 (I). Cells were cultured in presence of tetracycline (TET) for 5 days, or maintained in control (CTR) conditions in the absence of tetracycline. Note that for panels B-D and F-I the histograms have been normalized so that the area under the curve equals to 1 (100%) for all samples presented, in order to facilitate direct visual comparison. (**J**) Nucleotide sequences of inducible H1 Pol III promoters for the sOPTiKO system containing one or two tet operons (H1-TO and H1-2TO, respectively). Key sequence features are highlighted with various colours. The restriction enzyme cut sites used for gRNA cloning are shown (Fig. 7B). DSE: distal sequence element; PSE: proximal sequence element; TETO2: tet operon; +1: start position of RNA transcription. (**K**) Flow cytometry quantification of EGFPd2 inducible knockout kinetics in sOPTiKO cells from panels F (gRNA 2 – TO) and I (gRNA 3 – 2TO). The percentage of EGFP positive cells was monitored daily following addition of tetracycline. Results are from 2 independent cultures. (**L**) Summary of the inducible H1 promoters for the sOPTiKD system recommended to obtain potent and tightly controlled inducible gene knockout according to the potency of a given gRNA.



Supplementary Figure 9. Validation of the optimized inducible CRISPR/Cas9 platform following hESCs differentiation.

(A-B) Representative immunofluorescent staining for the indicated lineage specific markers in cells derived from EGFPd2 homozygous sOPTiKO hESCs carrying a single inducible EGFP gRNA (EGFP gRNA 1). EGFPd2 fluorescence in control conditions (CTR) or after knockout (TET) is in green, and DAPI shows nuclear staining. Knockout was induced for 10 days. Merged images of the EGFPd2 and lineage specific markers are shown. Scale bars: 100 µm. **(C-G)** Representative flow cytometry quantifications for EGFPd2 expression in EGFPd2 homozygous sOPTiKD hESCs (iKO) differentiated into the lineages indicated, and treated as described for panels A-B. For each panel, the histogram on the left reports analysis of live cells, in which EGFPd2 expression was compared with ROSA26-EGFPd2 homozygous hESCs (EGFPd2 HOM, positive control) and wild-type H9 hESCs (H9, negative control). The percentage of EGFPd2 positive cells (%EGFP+, the gate used is indicated) and the median fluorescence intensity (MFI) of EGFPd2 positive cells are reported. Note that the histograms have been normalized so that the area under the curve equals to 1 (100%) for all samples, in order to facilitate direct visual comparison. For all panels but G, the plots on the right describe analysis of cells stained for the indicated lineage-specific markers. The FITC (EGFPd2) gating was set based on FITC autofluorescence in wild-type hESCs-derived cells, and the APC (lineage marker) gating was set based on secondary antibody-only staining. **(H)** Expression of EGFPd2 in EGFPd2 homozygous sOPTiKD hESCs differentiated into pancreatic endocrine cells. Cells were stained for c-peptide (insulin) and gated according to secondary antibody-only staining. Due to the high autofluorescence of fixed c-peptide positive cells, EGFPd2 expression was measured by comparing fluorescence in the PE channel (indicative of non-specific autofluorescence) with the one in the FITC channel (indicative of specific EGFPd2 expression). C-peptide positive cells derived from wild-type H9 hESCs were used as negative control for this gating strategy, as indicated.

Supplementary Tables

Supplementary Table 1. Summary of gene targeting experiments

Donor plasmid	Picked clones	Incorr. target.[a]	Rand. Int.[b]	Het + extra[c]	Homo + extra[c]	Het	Homo	Correct targeting efficiency (%)[d]	Total targeting efficiency (%)[e]
pR26-Puro_ENDO-EGFP	12	0	0	9	0	3	0	25	100
pR26-Neo_EF1a-EGFP	12	0	1	8	0	3	0	25	92
pR26-Neo_CAG-EGFP	11/11[f]	3/1[f]	0/0[f]	2/2[f]	0/1[f]	6/5[f]	0/2[f]	54/64[f]	73/91[f]
pR26-Neo_CAG-STDtetR	28	6	1	8	3	8	2	36	75
pR26-Neo_CAG-OPTtetR	28/27[f]	3/11[f]	0/0[f]	12/9[f]	3/2[f]	10/4[f]	0/1[f]	36/19[f]	89/59[f]
pR26-Bsd_CAG-EGFPd2	71	11	12	17	6	21	2	32	67
pAAV-Puro_CAG-EGFP	8/4/8[g]	0/0/0[g]	0/0/1[g]	0/0/0[g]	3/1/2[g]	0/0/0[g]	5/3/5[g]	62/75/62[g]	100/100/87[g]
pAAV-Puro_EGFPiKD	8/4/8[g]	0/0/0[g]	0/0/0[g]	0/0/0[g]	6/2/5[g]	0/0/0[g]	2/2/3[g]	25/50/37[g]	100/100/100[g]
pAAV-Puro_iKD-OCT4	6	0	0	1	2	0	3	50	100
pAAV-Puro_iKD-B2M	6	0	1	0	4	0	1	17	83
pAAV-Puro_iKD-DPY30	6	0	1	0	4	0	1	17	83
pAAV-Puro_EGFPsiKD-STD	6	0	0	0	3	0	3	50	100
pAAV-Puro_EGFPsiKD-OPT	6/12[h]	0/1[h]	0/3[h]	0/0[h]	3/5[h]	0/1[h]	3/2[h]	50/25[h]	100/67[h]
pAAV-Puro_siKD-OCT4	6/6[i]	0/0[i]	0/0[i]	1/5[i]	3/0[i]	1/1[i]	1/0[i]	33/17[i]	100/100[i]
pAAV-Puro_siKD-B2M	6/6[i]	0/0[i]	0/0[i]	4/4[i]	2/0[i]	0/2[i]	0/0[i]	0/33[i]	100/100[i]
pAAV-Puro_siKD-T	8	0	0	1	3	1	3	50	100
pAAV_Neo_CAG-Cas9 pAAV_Puro_siKO-EGFP-1[j]	11/13[k]	0/0[k]	0/1[k]	1/3[k]	6/8[k]	0/0[k]	4/1[k]	36/8[k]	91/69[k]
pAAV_Neo_CAG-Cas9 pAAV_Puro_MsiKO-EGFP-1x2[j]	12	0	0	0	11	0	1	8	100
pAAV_Neo_CAG-Cas9 pAAV_Puro_MsiKO-EGFP-1x3[j]	12	0	0	1	7	0	4	33	92
pAAV_Neo_CAG-Cas9 pAAV_Puro_siKO-EGFP-2[j]	4	0	0	0	3	0	1	25	100
pAAV_Neo_CAG-Cas9 pAAV_Puro_siKO-2TO-EGFP-2[j]	7	0	0	0	5	0	2	29	100
pAAV_Neo_CAG-Cas9 pAAV_Puro_siKO-EGFP-3[j]	6	0	0	0	3	2	1	17	67
pAAV_Neo_CAG-Cas9 pAAV_Puro_siKO-2TO-EGFP-3[j]	15	0	0	0	10	0	5	33	100

[a] Evidence of targeting, but incorrect size of 5'- or 3'-integration PCR (see Fig. S1A for a schematic of the PCR genotyping strategies used, and refer to Supplementary Methods).

[b] No evidence of targeting (lack of bands in 5'- and 3'-integration PCR, and presence of WT band in locus PCR).

[c] Correct targeting, but with additional random integration of the targeting plasmid (bands in 5'- and/or 3'-backbone PCR).

[d] Het + homo only.

[e] Including clones with additional random integration of the plasmid (het + extra and homo + extra).

[f] The two figures are from two different targeting experiments in hESCs.

[g] The three figures are from targeting in three different hESCs: ROSA26 HOMO STDtetR, ROSA26 HET OPTtetR, and ROSA26 HOMO OPTtetR, respectively.

[h] The two figures are from targeting in two different hESCs: ROSA HOMO EGFP and ROSA HOMO EGFPd2, respectively.

[i] The first figure is from hESCs targeting; the second figure is from hiPSCs targeting.

[j] Simultaneous targeting of both vectors followed by Neo+Puro drug selection. For these experiments, het indicates that only one of the two vectors was correctly targeted, while hom clones were targeted with each vector in one of the alleles of the AAVS1 locus. Accordingly, both correct and total targeting efficiency was calculated only on hom and hom+extra clones.

[k] The two figures are from targeting in two different hESCs: ROSA HOMO EGFPd2 and ROSA HET EGFPd2, respectively.

Supplementary Material and Methods

Differentiation of hESCs

Differentiation was initiated in adherent cultures of hESCs 48 h following passaging. Unless otherwise specified, media changes were performed daily, and volumes were adjusted to cell density. See the below for media compositions not specified in the text. hPSCs were routinely monitored for absence of mycoplasma contamination and chromosomal stability.

Composition of media used for hPSC culture and passaging

Media	Components	Concentration	Supplier
CDM BSA/PVA	IMDM : F-12 (1:1)	-	Gibco
	Chemically defined concentrated lipids	1%	Gibco
	1-thioglycerol	450 mM	Sigma-Aldrich
	Insulin	7 μg ml^{-1}	Roche
	Transferrin	15 μg ml^{-1}	Roche
	Penicillin-streptomycin (optional)	1%	Gibco
	Bovine serum albumin (BSA)	5 mg ml^{-1}	Europa Bio Product
	OR polyvinyl alcohol (PVA)	or 1 mg ml^{-1}	Sigma-Aldrich
KSR	Advanced DMEM F12	-	Gibco
	Knockout Serum Replacer	20%	Gibco
	L-Glutamine	1%	Gibco
	B-mercaptoethanol	100 μM	Sigma-Aldrich
	Penicillin-streptomycin (optional)	1%	Gibco
MEF Media	Advanced DMEM F12	-	Gibco
	Fetal bovine serum	10%	Biosera
	L-glutamine	1%	Gibco
	β-mercaptoethanol	100 μM	Sigma-Aldrich
	Penicillin-streptomycin (optional)	1%	Gibco
Gelatine	Embryo transfer water	-	Gibco
	Gelatine from porcine skin	0.1%	Sigma-Aldrich
Collagenase	Advanced DMEM F12	-	Gibco
	Knockout serum replacer	20%	Gibco
	L-Glutamine	1%	Gibco
Dispase	DMEM/F12	-	Gibco
	Dispase	1 mg ml^{-1}	Gibco

Germ layers: neuroectoderm was induced for 6 days in CDM-BSA with 12 ng ml^{-1} FGF2 (Dr. Marko Hyvonen, University of Cambridge) and 10 μM SB-431542 (Activin/Nodal/TGFβ signalling inhibitor; Tocris), as described previously (Vallier et al., 2009). Mesoderm specification was obtained in two steps: (1) induction of early mesoderm/primitive streak-like cells (late streak, posterior streak, and anterior streak, respectively for presomitic, lateral plate, and cardiac mesoderm) for 36 h, a single media change; (2) mesoderm patterning for 3.5 days, two media changes. Presomitic mesoderm: (1) CDM-BSA with 40 ng ml^{-1} FGF2 and 8 μM CHIR99021 (WNT signalling activator; Tocris); (2) CDM-BSA with 4 ng ml^{-1} FGF2, 1 μM Retinoic Acid (Sigma-Aldrich), 0.1 μM LDN193189 (BMP signalling inhibitor; Biovision), 10 μM SB-431542, and 1 μM Purmorphamine (SHH signalling activator; Tocris; Mendjan et al., 2014). Lateral plate mesoderm: (1) CDM-PVA with 20 ng ml^{-1} FGF2, 10 μM LY294002 (PI3K inhibitor; Promega), and 10 ng ml^{-1} BMP4 (R&D); (2) CDM-PVA with 20 ng ml^{-1} FGF2, and 50 ng ml^{-1} BMP4 (Cheung et al., 2012). Cardiac mesoderm: (1) CDM-BSA (without insulin) with 20 ng ml^{-1} FGF2, 10 μM LY294002, 50 ng ml^{-1} Activin-A (Dr. Marko Hyvonen, University of Cambridge), and 10 ng ml^{-1} BMP4; (2) CDM-BSA (without insulin) with 8 ng ml^{-1} FGF2, 10 ng ml^{-1} BMP4, 1 μM IWR1 (WNT signalling inhibitor, Sigma-Aldrich), and 0.5 μM Retinoic Acid (Mendjan et al., 2014). Definitive endoderm was differentiated for 3 days in CDM-PVA (without insulin) with 20 ng ml^{-1} FGF2, 10 μM LY294002, 100 ng ml^{-1} Activin-A, and 10 ng ml^{-1} BMP4, as previously described (Touboul et al., 2010).

Neural cells: neural cell types were obtained using two different protocols. For experiments presented in Fig. 1, Fig. 4, Fig. 8, Fig. S2, Fig. S4, Fig. S7 and Fig. S9 (GSH EGFP reporter and EGFP inducible knockdown/knockout validation), a mixed bulk culture of neurons, astrocytes and oligodendrocytes were obtained according to a published protocol (Douvaras et al., 2014), with minor modifications. hESCs were cultured in CDM-BSA with 10 µM SB431542, 0.1 µM LDN193189 for 8 days to generate neuroepithelial cells. During this stage, 0.1 µM Retinoic Acid and 1 µM purmorphamine were respectively added from day 2 and day 4 onwards to promote neuroepithelial patterning towards a caudal and ventral fate. From day 8 to 12, media was switched to N2 media (Thermo) with 0.1 µM Retinoic Acid and 1 µM Purmorphamine (SHH agonist). At day 12, adherent neural progenitors were mechanically lifted and transferred into low-attachment plates to form neurospheres. During this stage, cells were initially cultured in N2B27 media (Thermo) supplemented with 0.1 µM Retinoic Acid and 1 µM Purmorphamine for 8 days. Subsequently, neuronal and glial specification was initiated at day 20 of differentiation by culturing neurospheres in differentiation medium consisting of N2B27 media (Thermo) supplemented with 60 ng ml^{-1} T3 (Sigma-Aldrich), 100 ng ml^{-1} biotin (Sigma-Aldrich), 1 µM dbcAMP (Sigma-Aldrich), 10 µg ml^{-1} insulin (Roche), 10 ng ml^{-1} FGF2, 10 ng mlv PDGF-AA (Peprotech), and 10 ng ml^{-1} NT3 (R&D). From this point onwards media changes were performed every other day. At day 30, spheres were plated onto poly-D-lysine(PDL)/laminin-coated dishes to promote neuronal and glial outgrowth, and cultured in differentiation medium without FGF2. From day 70 onwards, the concentration of PDGFaa and NT3 was reduced to half of the previous levels to promote cell maturation. Differentiation concluded at day 95, and resulted into a mixed culture of mostly astrocytes and neurons, and a minority of oligodendrocytes.

For the experiments presented in Fig. 6 and Fig. S6 (*DPY30* inducible knockdown during neuronal differentiation) we used a shorter differentiation protocol modified from that described above. This resulted in a culture containing predominantly (>70%) neurons, but not astrocytes or oligodendrocytes. Neuroepithelial induction was initiated in CDM-BSA supplemented with 20 ng ml^{-1} FGF2, 100 nM LDN193189, 10 µM SB431542 and 3 µM CHIR99021. The medium was refreshed with the same composition after 1 day. From day 2 onwards, the medium was changed to N2B27, a 1:1 mixture of DMEM/F12 and Neurobasal medium that contained 1% B27 and 0.5% N2 supplements, 1% Glutamine, 1% Penicillin/Streptomycin and 100 µM β-mercaptoethanol. From day 2 to day 4 SB431542 was added to this medium. Thereafter, the differentiation was continued without the addition of SB431542. After 10 days of differentiation neural rosettes containing progenitors of the neural lineages were obtained. These differentiated into neurons over the following days, and cultures were analyzed after 21 days of differentiation. *DPY30* inducible knockdown (KD) was induced from: (1) three days before the start of differentiation, KD induction (in neuroepithelium); (2) day 7, KD specification (in neural progenitors); (3) day 14, KD maturation (in neurons).

Osteocytes: neural crest differentiation was performed following modifications of a previously published protocol for neuroectoderm differentiation (Cheung et al., 2014; Felipe Serrano, William G. Bernard, et al., manuscript in preparation). Briefly, cells were differentiated into neuroectoderm as described earlier, and neural crest were enriched and expanded from this population. Additional details of this method are available upon request to Dr. Sanjay Sinha. Osteocytes were differentiated from neural crest cells using StemPro Osteogenesis Differentiation Kit (Gibco) according to manufacturer's instructions. Cells were cultured in these conditions for two weeks performing media changes every other day.

Chondrocytes: chondrocyte differentiation was performed as previously described (Mendjan et al., 2014). Following presomitic mesoderm differentiation for 5 days as described above, chondrogenic specification was induced in CDM-BSA with 8 ng ml^{-1} FGF2 and 10 ng ml^{-1} BMP4 for 10 days and performing media changes every other day.

For alcian blue staining, chondrocytes were fixed in 100% methanol at -20°C for 15 minutes, washed twice with ice-cold PBS, once with 0.5 N HCl, and finally stained overnight (16 h) in alcian blue solution (0.25% alcian blue in 0.5 N HCl) at room temperature and under gentle shaking. Cells where then washed 3 times for 15 minutes at room temperature with PBS before phase-contrast imaging of the staining. Alcian blue staining was then quantified by extracting the dye overnight (16 h) in 8 M guanidine-HCl (1 ml for a well of a 12-well tissue culture plate) at room temperature and under gentle shaking. The absorbance at 595 nm was then measured using a NanoDrop 1000 (NanoDrop). Undifferentiated hESCs were used as negative controls for the staining, and the alcian blue absorbance observed in such controls was subtracted from that of chondrocytes to quantify specific staining.

Smooth muscle: smooth muscle cells were obtained as previously reported (Cheung et al., 2012; Cheung et al., 2014). Following lateral plate mesoderm differentiation for 5 days as described above, cells were dissociated in TrypLE Express (Gibco) for 5 minutes at 37°C, washed with CDM-PVA and centrifuged at 200 g for 3 minutes at room temperature, and

seeded at a density of 20,000 cells per cm^2 onto gelatin/MEF media-coated plates in CDM-PVA supplemented with 10 ng ml^{-1} PDGF-BB (Peprotech) and 2 ng ml^{-1} TGFβ (Peprotech). Cells were cultured in these conditions for 12 days to generate smooth muscle cells. Media was changed every other day, and cells were split after the first 6 days of culture as just described, but seeded at a 1:2 ratio.

DPY30 inducible knockdown (KD) during smooth muscle differentiation was induced from: (1) three days before the start of differentiation, KD induction (in posterior primitive streak-like cells, also referred to as posterior early mesoderm in the text and figures); (2) day 2, KD specification (in lateral plate mesoderm); (3) day 10, KD maturation (in smooth muscle cells).

Cardiac fibroblasts: differentiation into epicardium and cardiac fibroblasts was done as previously reported (Iyer et al., 2015). Following lateral plate mesoderm differentiation for 5 days as described above, cells were dissociated in TrypLE Express (Gibco) for 5 minutes at 37°C, washed with CDM-PVA and centrifuged at 200g for 3 minutes at room temperature, and seeded at a density of 25,000 cells per cm^2 onto gelatin/MEF media-coated plates in CDM-PVA supplemented with 25 ng ml^{-1} WNT3A (R&D), 50 ng ml^{-1} BMP4 and 4 µM Retinoic Acid. Cells were cultured in these conditions for 10 days with media changes every 4 days to generate epicardial cells. Following this, cells were dissociated in TrypLE Express for 5 minutes at 37°C, washed with CDM-PVA by centrifuging at 300 g for 3 minutes at room temperature, and seeded at a density of 30,000 cells per cm^2 onto gelatin/MEF media-coated plates in CDM-PVA supplemented with 50 ng ml^{-1} VEGF-B (Peprotech) and 50 ng ml^{-1} FGF2. Cells were cultured in these conditions for 12 days with media changes every other day to generate cardiac fibroblasts.

Cardiomyocytes: cardiac differentiation was performed as previously described (Mendjan et al., 2014). Following cardiac mesoderm differentiation for 5 days as described above, cardiac maturation was initiated by culturing cells in CDM-PVA (without insulin) with 8 ng ml^{-1} FGF2 and 10 ng ml^{-1} BMP4 for 2 days (one media change). Following this, cells were cultured in CDM-PVA (without insulin) until beating clusters appeared (between day 8 and 10 of differentiation). Cardiomyocytes were then matured in CDM-PVA containing insulin. Media changes were performed every other day. Cells were analyzed at day 17 for all experiments apart from the prolonged EGFP inducible knockdown treatment (Supplementary Fig. 4H-I), for which knockdown was induced starting from day 15 (D0 of tetracycline treatment), and cells were collected at different time points until day 35 (D20 of tetracycline treatment).

DPY30 inducible knockdown (KD) during cardiac differentiation was induced from: (1) three days before the start of differentiation, KD induction (in anterior primitive streak-like cells, also referred to as anterior early mesoderm in the text and figures); (2) day 2, KD specification (in cardiac mesoderm); (3) day 10, KD maturation (in cardiomyocytes).

Intestine: intestinal organoids generation followed a previously published method (Fordham et al., 2013; Hannan et al., 2013a), with some minor changes. Following definitive endoderm differentiation for 3 days as described above, cells were cultured in RPMI/B27 media (consisting of RPMI Medium 1640 with GlutaMAX, 2% B27 supplement, 1% non essential amino acids, 100 U ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin; all from Gibco). This was first supplemented with 50 ng ml^{-1} Activin-A for 1 day, then with 6 µM CHIR99021 and 3 µM Retinoic Acid for 4 days, in order to pattern definitive endoderm into a posterior fate and obtain hindgut cells. Monolayer cultures were then transferred to three-dimensional conditions. For this, cells were dissociated to small clumps using collagenase IV for 20 minutes at 37°C, washed twice with basal growth medium (consisting of Advanced DMEM/F12, 10 mM HEPES pH 7.4, 2% B27 serum-free supplement, 1% N2 serum-free supplement, and 20 mM L-Glutamine) by centrifuging at 200 g for 3 minutes at room temperature, and finally resuspended at a density of 100,000 cells per ml in a mixture of 70% Matrigel (BD Biosciences) and 30% basal growth medium supplemented with 500 ng ml^{-1} R-Spondin 1 (R&D), 3 µM CHIR99021, 100 ng ml^{-1} Noggin (BMP signalling inhibitor, R&D), 2.5 µM Prostaglandin E2 (Cayman Chemicals), 100 ng ml^{-1} EGF (R&D), and 0.5 µM A83-01 (TGFβ signalling inhibitor, Tocris). The cell-medium-matrigel solution was distributed in 50 µl droplets, one for each well of 24-well plates. The droplets were allowed 30 minutes for the gel to solidify, following which 1 ml of basal media with growth factors was added per well. 10 µM Y-27632 was also added only for the first 24 h to promote cell survival. Cells were cultured in these conditions for 10 days with media changes every other day to generate intestinal organoids.

Pancreas: pancreatic differentiation was done according to what previously reported (Cho et al., 2012), with minor changes. Following definitive endoderm differentiation for 3 days as described above, cells were cultured in Adv-BSA media (consisting of Advanced DMEM/F12, 5 mg ml^{-1} BSA, 20 mM L-Glutamine, 100 U ml^{-1} penicillin, and 100 µg ml^{-1} streptomycin). This was supplemented with 3 µM Retinoic Acid, 50 ng ml^{-1} FGF10 (Autogen Bioclear), 150 ng ml^{-1} Noggin,

and 10 μM SB431542 for 3 days to generate dorsal foregut, then with 3 μM Retinoic Acid, 50 ng ml^{-1} FGF10, 150 ng ml^{-1} Noggin, and 0.25 μM KAAD-Cyclopamine (SHH signalling inhibitor, Toronto Research Chemicals) for 3 days, and finally with 3 μM Retinoic Acid, 50 ng ml^{-1} FGF10, and 0.25 μM KAAD-Cyclopamine for 3 days. Pancreatic progenitors were obtained at this stage (day 12 of differentiation). Pancreatic specification was then initiated in Adv-BSA with 3 μM Retinoic Acid, 1% B27 supplement, and 1 μM DAPT (Notch signalling inhibitor, Sigma-Aldrich) for 3 days. During this stage, 0.1 mM 6-Bnz-cAMP (PKA activator, Sigma-Aldrich) was also added only for the first 48 h. Cells were then cultured in Adv-BSA with 3 μM Retinoic Acid, 1% B27 supplement, and 0.25 μM KAAD-Cyclopamine for 3 days (one media change) in order to obtain immature pancreatic endocrine cells. These were further matured for 6 days (media changes every 72 h) under the same culture conditions to generate mature pancreatic endocrine cells (predominantly β-cells, some α-cells, and few δ-cells) at day 24 of differentiation.

DPY30 inducible knockdown (KD) during pancreatic differentiation was induced from: (1) three days before the start of differentiation, KD induction (in definitive endoderm); (2) day 9, KD specification (in pancreatic progenitors); (3) day 17, KD maturation (in pancreatic endocrine cells).

Hepatocytes: hepatocytes were generated according to previous reports (Hannan et al., 2013b; Touboul et al., 2010), with minor modifications. Following definitive endoderm differentiation for 3 days as described above, cells were cultured in RPMI/B27 media supplemented with 50 ng ml^{-1} Activin-A for 5 days to generate anterior foregut cells. Cells were then cultured in Hepatozyme (Gibco) supplemented with 2% non essential amino-acids, 2% chemically defined concentrated lipids (Gibco), 20 mM L-Glutamine, 14 μg ml^{-1} insulin, 15 μg ml^{-1} transferrin (Roche), 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin, 50 ng ml^{-1} HGF (R&D), and 20 ng ml^{-1} Oncostatin M (R&D), with media changes performed every other day. After 3 days into these conditions (day 11 of differentiation), hepatic progenitors were obtained. These were further matured under the same culture conditions for 11 more days to generate mature hepatocytes at day 22.

DPY30 inducible knockdown (KD) during hepatic differentiation was induced from: (1) three days before the start of differentiation, KD induction (in definitive endoderm); (2) day 8, KD specification (in hepatocyte progenitors); (3) day 15, KD maturation (in hepatocytes).

Cholangiocytes: cholangiocytes generation was previously described (Sampaziotis et al., 2015). Following definitive endoderm differentiation for 3 days as described above, cells were cultured in RPMI/B27 media. This was first supplemented with 50 ng ml^{-1} Activin-A for 5 days to generate foregut progenitors, then with 10 μM SB431542 and 50 ng ml^{-1} BMP4 for 4 days to obtain bipotent hepatoblasts, and finally with 50 ng ml^{-1} FGF10 (Peprotech), 50 ng ml^{-1} Activin-A, 3 μM Retinoic Acid for 4 days to derive cholangiocyte progenitors. Monolayer cultures were then transferred to three-dimensional conditions. For this, cells were incubated with cell dissociation buffer (Gibco) for 10 minutes at 37°C, mechanically dissociated into small clumps, washed twice with RPMI medium by centrifuging at 200 g for 3 minutes at room temperature, and finally resuspended at a density of 200,000 cells per ml in a mixture of 60% Matrigel and 40% William's E Medium (Gibco) supplemented with 10 mM nicotinamide (Sigma-Aldrich), 17 mM sodium bicarbonate (Sigma-Aldrich), 0.2 mM 2-Phospho-L-ascorbic acid trisodium salt (Sigma-Aldrich), 6.3 mM sodium pyruvate (Invitrogen), 14 mM glucose (Sigma-Aldrich), 20 mM HEPES pH 7.4 (Invitrogen), ITS+ premix (BD Biosciences), 0.1 uM dexamethasone (R&D Systems), 2 mM Glutamax (Invitrogen), 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin and 20 ng ml^{-1} EGF. The cell-medium-matrigel solution was distributed in 50 μl droplets, one for each well of 24-well plates. The droplets were allowed 30 minutes for the gel to solidify, following which 1 ml of William's E medium with supplements was added per well. 10 μM Y-27632 was also added only for the first 24 h to promote cell survival. Cells were cultured in these conditions for 10 days with media changes every other day to obtain cholangiocytes.

Lung: generation of lung epithelium was recently described (Hannan et al., 2015). Following definitive endoderm differentiation for 3 days, cells were cultured in RPMI/B27 media supplemented with 50 ng ml^{-1} Activin-A for 5 days to generate anterior foregut cells. These cells were cultured in RPMI which was first supplemented with 100 ng ml^{-1} FGF10, and 1 μM Retinoic Acid for 5 days to pattern lung endoderm, then supplemented with 100 ng ml^{-1} FGF10 and 50 ng ml^{-1} HGF for 10 days to generate lung progenitors, and finally supplemented with 100 ng ml^{-1} FGF10 for 15 days, in order to mature lung epithelial cells.

Molecular cloning

Unless otherwise indicated, traditional cloning was performed using restriction enzymes from NEB and T4 DNA ligase from Promega, and Gibson cloning was performed using Gibson Assembly Master Mix (NEB). PCR were performed using Q5 Hot Start High-Fidelity DNA Polymerase (NEB), DNA blunting was done with DNA Polymerase I, Large (Klenow) Fragment (NEB), and site-directed mutagenesis was achieved using QuickChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Whenever possible, vectors were dephosphorylated before ligation using Alkaline Phosphatase Calf Intestinal (CIP) from NEB. All oligonucleotides were ordered from Sigma-Aldrich as desalted lyophilized products. QIAEX II Gel Extraction Kit and QIAquick PCR Purification Kits (QIAGEN) were used for DNA extraction from agarose gels and purification of PCR products, respectively. Recombinant plasmids were transformed into DH5α E. Coli (Alpha-Select Gold Efficiency, Bioline). QIAGEN Plasmid Mini, Midi and Maxi Kits were used for plasmid preparations. All these procedures were performed according to manufacturer's instructions. Additional molecular biology procedures (such as DNA electrophoresis and E. Coli culture) were performed according to standard protocols. All the plasmids were sequence verified by Sanger sequencing through Beckman Coulter Genomics.

ROSA26 CRISPR/Cas9n (pSpCas9n(BB)_R26-L and pSpCas9n(BB)_R26-R): the pSpCas9n(BB) (a gift of Feng Zhang, Addgene plasmid #48873) was used to generate two plasmid each expressing a small guide RNAs (gRNAs) and the Cas9n (D10A nickase mutant) from *S. Pyogenes*. Left and right guide RNAs specific for the human ROSA26 locus were designed using the online CRISPR Design Tool from the Zhang-lab, MIT, Boston, MA (http://tools.genome-engineering.org): gRNA-L 5'-GTCGAGTCGCTTCTCGATTA(TGG)-3'; gRNA-R 5'-GGCGATGACGAGATCACGCG(AGG)-3' (PAM sites are reported in parenthesis). The score of the resulting pair of gRNAs was 97 (high quality), with no predicted off target effects. The gRNAs were cloned as previously described (Ran et al., 2013) using the following single stranded oligonucleotide pairs: gRNA-L-top 5'-CACCGTCGAGTCGCTTCTCGATTA-3'; gRNA-L-bot: 5'-AAACTAATCGAGAAGCGACTCGAC-3'; gRNA-R-top 5'-CACCGGCGATGACGAGATCACGCG-3'; gRNA-R-bot 5'-AAACCGCGTGATCTCGTCATCGCC-3'. The combination of the two resulting plasmids (pSpCas9n(BB)_R26-L and pSpCas9n(BB)_R26-R) is predicted to induce a specific double strand break in the intron between exons 1 and 2 of THUMPDS3-AS1 on chromosome 3 (ROSA26 locus).

pR26-Puro_ENDO-EGFP: the targeting vector for the human ROSA26 locus was constructed starting from a pUC plasmid. First, left (5', 904bp) and right (3', 869bp) homology arms (HAR) were generated by PCR from H9 hESCs genomic DNA using primers that inserted restriction sites on the 5' and 3' ends of each amplicon (MfeI/KpnI for 5'-HAR, and SalI/HindIII for 3'-HAR): 5'-HAR_fw 5'-GACTCAATTGGCTCGAAACCGGACGGAGCCATTGCTC-3'; 5-HAR_rev 5'-GCATGGT-ACCGATCACGCGAGGAGGAAAGGAGGGAGG-3'; 3'-HAR_fw 5'-GACTGTCGACGCTTCTCGATTATGGGCGGG-ATTCTTTTGC-3'; 3'HAR_rev 5'-GCATAAGCTTGGAAGCTATCACACAGGCATCTGAGATCAG-3', These amplicons were inserted sequentially into the multiple cloning site of the pUC19 by restriction digestion: first, the 5'-HAR was ligated into the EcoRI/KpnI sites; secondly, the 3'-HAR was ligated into the SalI/HindIII sites. The resulting vector was termed pR26. Next, a promoterless gene-trap vector was constructed in which expression of a bicistronic SA-puromycin-T2A-EGFP-pA cassette is under control of the endogenous ROSA26 promoter (THUMPD3-AS1 gene) following correct genomic integration. For this a Gibson Assembly was performed in which three inserts were cloned into the KpnI/SalI sites of pR26. The first insert encoded for the adenoviral splice acceptor (SA), which was first synthesised *de novo* and cloned into an in house shuttle vector. The Gibson Assembly fragment was produced from this by PCR amplification using primer fw 5'-TTTCCTCCTCGCGTGATCGGTACCTAGGGCGCAGTAGTCCAGG-3' and primer rev 5'-CTCGGTCATGGT-GGCCGGTCCGGGATTCTCCTC-3'). The second insert encoded for the puromycin resistance gene, which was PCR amplified from pTRE-TIGHT-EGFP (a gift from Rudolf Jaenisch. Addgene plasmid #22074) using primer fw 5'-TCCCGGACCGGCCACCATGACCGAGTACAAGCCCACGGTG-3' and primer rev 5'-CTCCACTGCCCTTAAGGGCA-CCGGGCTTGCGGGT-3'. The third insert encoded the T2A-EGFP-pA cassette, which was PCR amplified from pSpCas9n(BB)-2A-GFP (a gift from Feng Zhang, Addgene plasmid #48140) using primer fw 5'-GCCCGGTGCCCTTAAGGGCAGTGGAGAGGGCAGA-3' and primer rev 5'-CCGCCCATAATCGAGAAGCGT-CGACCCCCAGCATGCCTGCTAT-3'. The resulting vector was termed pR26-Puro_ENDO-EGFP.

pR26-Neo_EF1α-EGFP: a gene-trap vector was constructed in which expression of EGFP was under control of an exogenous EF1a promoter. For this, a Gibson Assembly was performed in which three inserts were cloned into the BglII/SacI sites of pR26-Puro_ENDO-EGFP (thus substituting the puromycin-2A-EGFP sequence). The first insert encoded for the neomycin resistance gene and was PCR amplified using primer fw 5'-CTTTCCAGTTTCGAACGGGAG-ATCTGCCACCATGGGATCGGCCATTGAA-3' and primer rev 5'-CGGAGCCAATCCATAGAGCCCACCGCAT-3'. The

second insert encoded the EF1α promoter which was PCR amplified from pLVX-EtO (a gift from Oliver Brüstle; Ladewig et al., 2012) using primer fw 5'-GGCTCTATGGATTGGCTCCGGTGCCCGT-3' and primer rev 5'-ATGGTGGCG-GCGGATCCGGGTCGAAATTCC-3'. The third insert encoded the ORF of EGFP and was PCR amplified from pTRE-TIGHT-EGFP (see above) using primer fw 5'-ACCCGGATCCGCCGCCACCATGGTGAGCA-3' and primer rev 5'-TCGAGGCTGATCAGCGAGCTACGCGTTTATCTAGACTTGTACAGCTCGTCCATGCCG-3'. The resulting vector was termed pR26-Neo_EF1α-EGFP.

pR26-Neo_CAG-EGFP: a gene-trap vector was constructed in which expression of EGFP was under control of an exogenous CAG promoter. For this the EF1α promoter in pR26-Neo_EF1α-EGFP was removed with SacI and BamHI and exchanged with the CAG promoter from an in house vector (pAAV-Neo_CAG) cut by a SacI-digestion, following blunt-ending of both the backbone and insert. The resulting vector was termed pR26-Neo_CAG-EGFP.

pR26-Neo_CAG-STDtetR: the wild-type bacterial tetR sequence containing a 5'-terminus SV40 nuclear localization sequence (tetR-nls, which we named standard tetR, or STDtetR) was amplified by PCR from pCAGTetRnls (a gift of Peter Andrews, Addgene plasmid #26599) using primers fw 5'-CATTTTGGCAAAGAATTAATTCGGATCCACCATGCC-AAAAAAGAAGAGGAAGGTATC-3' and rev 5'- CGAGGCTGATCAGCGAGCTACGCGTTTACCGCGGAGACCCAC-TTTCAC-3'. The resulting product was cloned by Gibson Assembly into pR26-Neo_CAG-EGFP following EGFP removal by BamHI-MluI digestion.

pR26-Neo_CAG-OPTtetR: the STDtetR cDNA sequence was used as template for gene synthesis following multi-parameter codon and RNA optimization (Fath et al., 2011) (GeneArt GeneOptimizer service, Invitrogen; refer to the plasmid sequences in Appendix S1 for details of the resulting cDNA). The resulting synthetic gene (which we named optimized tetR, or OPTtetR) was amplified by PCR using primer fw 5'-CATTTTGGCAAAGAATTAATTCGGATCCACCATGCCCAAG-AAAAAGCGG-3' and rev 5'-CGAGGCTGATCAGCGAGCTACGCGTTCATCTGGGGGAGCCGC-3'). The resulting product was cloned by Gibson assembly into pR26-Neo_CAG-EGFP following EGFP removal by BamHI-MluI digestion.

pR26-Bsd_CAG-EGFPd2: first, pR26-Neo_CAG-EGFP was digested with BglII and SpeI in order to remove the Puro-polyA, and ligated via Gibson assembly with a Bst-polyA fragment. This last was obtained as synthetic double-stranded DNA from Integrated DNA Technologies, and amplified by PCR using primers fw 5'- CCAGTTTCGA-ACGGGAGATCTGCCACCATGGCCA-3' and rev 5'- GTAATTGATTACTATTAATAACTAGTCCAGCTGGTTCTT-TCC-3'. The resulting product (pR26-Bsd_CAG-EGFP) was then digested with BamHI and MluI to remove the EGFP, and ligated via Gibson assembly with an EGFPd2 cDNA. This last was obtained by PCR from pHes1-GFPd2 (a gift from Connie Cepko, Addgene plasmid #14808) following PCR using primers fw 5'-TAATTCGGATCCCGCCACCATGGTGAGCAAG-3' and rev 5'- AGCTAGACGCGTCTACACATTGATCCTAGCAGAAGCACAG-3'.

AAVS1 ZFN (pZFN_AAVS1-L-ELD and pZFN_AAVS1-R-KKR): these plasmids were a generous gift of Dr. Kosuke Yusa (Wellcome Trust Sanger Institute, Hinxton, UK). Previously described AAVS1 left and right ZFNs amino acid sequences (Hockemeyer et al., 2009) were used as basis for artificial gene synthesis following codon optimization for mammalian expression (GeneArt, Invitrogen; refer to the plasmid sequences in Appendix S1 for details of the resulting cDNA), and with insertion of a 5' EcoRI site and a 3' XhoI site. Following digestion with these two enzymes, the ZFNs were cloned into pVAX1 (Invitrogen) using EcoRI-XhoI (generating pZFN_AAVS1-L and pZFN_AAVS1-R). These plasmids were then modified to generate obligate heterodimer ZFNs by introducing mutated FokI domains (Doyon et al., 2011). The FokI-ELD and FokI-KKR domains amino acid sequences were used as basis for artificial gene synthesis following codon optimization for mammailian expression (GeneArt, Invitrogen; refer to the plasmid sequences in Appendix S1 for details of the resulting cDNA), and with insertion of a 5' BamHI site and a 3' XhoI site. Following digestion with these two enzymes, the mutated FokI domains were cloned in ZFNs plasmid in which the normal FokI domains had been removed by BamHI-XhoI digestion. FokI-ELD was used for AAVS1 left ZFN (pZFN_AAVS1-L-ELD) and FoKI-KKR was used for AAVS1 right ZFN (pZFN_AAVS1-R-KKR). The combination of these two plasmids is predicted to induce a specific double strand break between exons 1 and 2 of PPR1R12C on chromosome 19 (AAVS1 locus).

pAAV-Puro_iKD (and related shRNA-containing plasmids): AAVS1 SA-2A-puro-pA donor (a gift from Rudolf Jaenisch, Addgene plasmid # 22075) was used to construct these plasmids. First, the single BglII site was removed by site-directed mutagenesis using top primer 5'-CACAGGGCCTCGAGAGTTCTGGCAGCGGAGAGG-3' and bottom primer 5'-CCTCTCCGCTGCCAGAACTCTCGAGGCCCTGTG-3'. This step was required as BglII had to be subsequently used to

clone the shRNAs. Secondly, the H1-TO promoter from pSUPERIOR_Neo (Oligoengine) was inserted in the HincII site following EcoRI-KpnI digestion and blunt-ending. The resulting plasmid was named pAAV-Puro_iKD. Finally, shRNAs were inserted to generate the final targeting vectors. For this last step, pAAV-Puro_iKD was digested with 25 U μg^{-1} of BglII and SalI-HF in 1x NEB3.1 at 37°C for 3 h, gel extracted following DNA electrophoresis, and reconstituted at 50 ng μl^{-1} for subsequent ligations. Validated shRNA sequences were obtained either from published reports (Zafarana et al., 2009; *OCT4/POU5F1* and *B2M* shRNAs), from the RNAi Consortium TRC library (Moffat et al., 2006; *DPY30*: clone TRCN0000131112; *CCND1*: clone TRCN0000010317; *CCND2*: clone TRCN0000045294; *CCND3*: clone TRCN0000003828; *T*: clone TRCN0000005484), or from Sigma-Aldrich (EGFP: vector SHC005; SCR: vector SHC016). A single G (guanine) was added at the 5' end of the shRNAs not starting with G or A, as the H1 promoter is more effective for sequences starting with such nucleotides. All other sequences were used without any modification. Complementary single-stand oligonucleotides containing the shRNAs, the Pol III terminator sequence, and appropriate overhangs (see table below) were mixed at a 50 µM concentration of each oligo in a 20 µl reaction containing 10 mM Tris-HCl pH8, 1 mM EDTA, 100 mM NaCl (annealing buffer), and annealed in a thermocycler with the following protocol: (1) 94°C for 5'; (2) 93°C for 20'', reduce by 1°C/cycle and repeat for 12 cycles; (3) 80°C for 4'; (4) 79°C for 20'', reduce by 1°C/cycle and repeat for 3 cycles; (5) 75°C for 4'; (6) 74°C for 20', reduce by 1°C/cycle and repeat for 3 cycles; (7) 70°C for 4'; (8) 69°C for 20'', reduce by 1°C/cycle and repeat for 60 cycles; (9) hold at 10°C. Annealed oligos were diluted 1:500 in annealing buffer, and 4 µl were used for ligation with 50 ng of pAAV-Puro_iKD (prepared by restriction digestion as described above) for 2 h at room temperature. Following transformation, bacterial colonies were screened by colony PCR using primers fw 5'-GTGTCATTCTATTCTGGGGGGTG-3' and rev 5'-GTGGGGGTTAGACCCAATATCAG-3', in a 12.5 µl mix containing 400 nM of each primer, 400 µM dNTPs, 3 mM $MgCl_2$, 1 x NH_4 Reaction Buffer, and 0.625 U of BIOTAQ DNA Polymerase (Bioline), and using the following PCR program: (1) 95°C for 5'; (2) 95°C for 30''; (3) 60°C for 30''; (4) 72°C for 1'; (5) repeat steps 2 to 4 for 34 times; (6) hold at 10°C. Positive clones carrying a band of approximately 500 bp (depending on the size of the shRNA) instead of the 464 bp band from the parental pAAV-Puro_siKD vector were expanded, and the shRNA sequence was confirmed by Sanger sequencing of isolated plasmids using the same primers employed for colony PCR.

Oligonucleotide sequences used for shRNAs cloning

Gene	Top oligo[d]	Bottom Oligo[d]
OCT4/POU5F1[a]	GATCCCGGATGTGGTCCGAGTGTGG**TTCAAGAGA** CCACACTCGGACCACATCC*TTTTTT*G	TCGACAAAAAAGGATGTGGTCCGAGTGTGG**TCTCTTGAA** CCACACTCGGACCACATCCGG
B2M[a]	GATCCCGGACTGGTCTTTCTATCTC**TTCAAGAGA** GAGATAGAAAGACCAGTCC*TTTTTT*G	TCGACAAAAAAGGACTGGTCTTTCTATCTC**TCTCTTGAA** GAGATAGAAAGACCAGTCCGG
SCR[b]	GATCCCGCGCGATAGCGCTAATAATTT**CTCGAG** AAATTATTAGCGCTATCGCGC*TTTTTT*G	TCGACAAAAAAGCGCGATAGCGCTAATAATTT**CTCGAG** AAATTATTAGCGCTATCGCGCGG
EGFP[b]	GATCCCGTACAACAGCCACAACGTCTAT**CTCGAG** ATAGACGTTGTGGCTGTTGTA*TTTTTT*G	TCGACAAAAAATACAACAGCCACAACGTCTAT**CTCGAG** ATAGACGTTGTGGCTGTTGTACGG
DPY30[c]	GATCCCGTCTCACAGACAACGTTGAGA**CTCGAG** TCTCAACGTTGTCTGTGAGAC*TTTTTT*G	TCGACAAAAAAGTCTCACAGACAACGTTGAGA**CTCGAG** TCTCAACGTTGTCTGTGAGACGG
CCND1[c]	GATCCCGATTGGAATAGCTTCTGGAAT**CTCGAG** ATTCCAGAAGCTATTCCAATC*TTTTTT*G	TCGACAAAAAAGATTGGAATAGCTTCTGGAAT**CTCGAG** ATTCCAGAAGCTATTCCAATCGG
CCND2[c]	GATCCCGTCACCAACACAGACGTGGATT**CTCGAG** AATCCACGTCTGTGTTGGTGA*TTTTTT*G	TCGACAAAAAATCACCAACACAGACGTGGATT**CTCGAG** AATCCACGTCTGTGTTGGTGACGG
CCND3[c]	GATCCCGCACATGATTTCCTGGCCTTC**CTCGAG** GAAGGCCAGGAAATCATGTGC*TTTTTT*G	TCGACAAAAAAGCACATGATTTCCTGGCCTTC**CTCGAG** GAAGGCCAGGAAATCATGTGCGG
T[c]	GATCCCGCGAGGAGATCACAGCTCTTAA**CTCGAG** TTAAGAGCTGTGATCTCCTCG*TTTTTT*G	TCGACAAAAAACGAGGAGATCACAGCTCTTAA**CTCGAG** TTAAGAGCTGTGATCTCCTCGCGG

[a] shRNA from (Zafarana et al., 2009).

[b] shRNA from Sigma-Aldrich.

[c] shRNA from the RNAi Consortium TRC library; Moffat et al., 2006; https://www.broadinstitute.org/rnai/public/).

[d] Sense and anti-sense shRNA strands are underlined; the hairpin loop is in **bold**; Pol III terminator is in *italic*; the remaining sequences provided the necessary overhangs for cloning into pAAV-Puro_iKD or pAAV-Puro_siKD. Note that the 5' BglII site is purposely lost after cloning of the shRNA to facilitate potential screening of recombinant bacteria by restriction digestion.

pAAV-Puro_CAG-EGFP: AAVS1-CAGGS-EGFP (a gift from Rudolf Jaenisch, Addgene plasmid # 22212) was renamed pAAV-Puro_CAG-EGFP to provide consistent nomenclature with the rest of the plasmids used in the study.

pAAV-Puro_EGFPiKD: a fragment containing the CAG-EGFP-pA cassette was cut from pR26-Neo_CAG-EGFP with SpeI and HincII, blunt-ended, and inserted in pAAV-Puro_iKD-EGFP following digestion with PspXI and blunt ending.

pAAV-Puro_EGFPsiKD-STD: a fragment containing the CAG-STDtetR-pA cassette was cut from pR26-Neo_CAG-STDtetR with SpeI and HincII, blunt-ended, and inserted in pAAV-Puro_iKD-EGFP following digestion with PspXI and blunt ending.

pAAV-Puro_EGFPsiKD-OPT: a fragment containing the CAG-OPTtetR-pA cassette was cut from pR26-Neo_CAG-OPTtetR with SpeI and HincII, blunt-ended, and inserted in pAAV-Puro_iKD-EGFP following digestion with PspXI and blunt ending.

pAAV-Puro_siKD (and related shRNA-containing plasmids): a fragment containing the CAG-OPTtetR-pA cassette was cut from pR26-Neo_CAG-OPTtetR with SpeI and HincII, blunt-ended, and inserted in pAAV-Puro_iKD following digestion with PspXI and blunt ending to generate the pAAV-Puro_siKD vector. This plasmid was used to clone shRNAs as described above for pAAV-Puro_iKD (see table above for the sequences used), with the exception that colony PCR screening of the recombinant bacteria was performed using primers fw 5'-CGAACGCTGACGTCATCAACC-3' and rev 5'-GGGCTATGAACTAATGACCCCG-3' with a PCR annealing temperature of 60°C and generating a band of 350 bp (instead of 295 bp in the parental pAAV-Puro_siKD vector).

pAAV-Puro_MsiKD plasmids: vectors carrying multiple inducible shRNAs were generated by Gibson assembly of PCR products of individual shRNA cassettes that had previously been cloned in pAAV-Puro_iKD. For this, primers were designed to introduce short spacers to separate the shRNA cassettes, with each spacer ending with a unique sequence to allow for directional Gibson assembly (see table below). The 5'-most cassette included an overlap (named L, for left hand side overlap) to the 3'-end of the linearized vector, and the 3'-most cassette included an overlap (named R, for right hand side overlap) to the 5'-end of the linearized vector. For double shRNA assembly, the cassettes were linked to each other by overlap BL1 (for block 1), while three shRNA cassettes were orderly linked by overlap BL1 and BL2 (for block 2), used in this order from the 5'- to 3'- end. As a result, the PCR products for a double shRNA assembly were constructed as: 5'-L_shRNA1_BL1-3' and 5'-BL1_shRNA2_R-3', while PCR products for triple shRNA assembly were: 5'-L_shRNA1_BL1-3', 5'-BL1_shRNA2_BL2-3', and 5'-BL2_shRNA3_R-3' (see table below and Figure 8G for a visual representation). PCR were performed using KAPA HiFi HotStart (Kapa Biosystems) according to manufacturer's instructions and using the following protocol: (1) 95°C for 3'; (2) 98°C for 20''; (3) 72°C for 30''; (4) repeat steps 2-3 for 24 cycles; (5) 72°C for 1'. PCR products were gel extracted, and 20 ng of each fragment were assembled with 100 ng of linearized pAAV_siKD following sequential digestion with BstBI (2 h at 50°C) and HincII (2 h at 37°C) to remove the pre-existing inducible H1 promoter. For this, Gibson Assembly Master Mix (NEB) was used according to manufacturer's instructions. Following transformation, bacterial colonies were screened by colony PCR as described above for pAAV_Puro_siKD, with the exception that the extension time was increased to 1'30" to accommodate for the larger insert, and that the screening generated bands of approximately 730 bp and 1080 bp for double and triple shRNA constructs, respectively. Correct assembly was confirmed by Sanger sequencing of isolated plasmids using the same primers employed for colony PCR, plus primers specific for BL1 (5'-GCTGTGTCTTGACAGCAGAC-3') and BL2 (5'-ACACAAGTACTGTCGGCAAC-3'). pAAV-Puro_MsiKD plasmids were generated for shRNAs against *CCND1* and *CCND2* (pAAV-Puro_MsiKD-D1-D2); *CCND1* and *CCND3* (pAAV-Puro_MsiKD-D1-D3); *CCND1*, *CCND2* and *CCND3* (pAAV-Puro_MsiKD-D1-D2-D3); and 2 or 3 scramble controls (pAAV-Puro_MsiKD-SCRx2 and pAAV-Puro_MsiKD-SCRx3, respectively).

Spacers used for Gibson assembly of multiple shRNA vectors

Spacer	**Primer FW**	**Primer REV**
5'-L_shRNA1_BL1-3'	TGCGGTGGGCTCTATGGGTCAATTCGAACGCTGACGTCATCAAC	GTCTGCTGTCAAGACACAGCATAGTCCTAGTAAAGCTTAGTACTGTCCG
5'-BL1_shRNA2_R-3'	GTCTGCTGTCAAGACACAGCATAGTCCTAGTAAAGCTTAGTACTGTCCG	ATTGATTACTATTAATAACTAGTCGAGGTCATCCCTAGTAAAGCTTAGTACTGTCCG
5'-BL1_shRNA2_BL2-3'	GTCTGCTGTCAAGACACAGCATAGTCCTAGTAAAGCTTAGTACTGTCCG	GTTGCCGACAGTACTTGTGTGTCCACCTAGTAAAGCTTAGTACTGTCCG
5'-BL2_shRNA3_R-3'	ACACAAGTACTGTCGGCAACCACACCGAACGCTGACGTCATCAAC	ATTGATTACTATTAATAACTAGTCGAGGTCATCCCTAGTAAAGCTTAGTACTGTCCG

pAAV-Neo_Cas9: the RTTA3 cDNA in AAVS1-SA-2A-NEO-CAG-RTTA3 (a gift of Paul Gadue, Addgene plasmid #60431) was removed by digestion with MluI and AflII, and replaced with the Cas9 cDNA from pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene #62988). For this, Cas9 was amplified by PCR using primers fw 5'-CTCATCATTTTGGCA-AAGAATTCCGCCACCATGGACTATAAGGACCACGA-3' and 5'-AGCCTGCACCTGAGGAGTGAATTCATTACTT-TTTCTTTTTTGCCTGGCCG-3', and cloned by Gibson assembly.

pAAV-Puro_siKO (and related gRNA-containing plasmids): pAAV-Puro_siKD was used as starting material. First, the only AarI site on this plasmid was removed to allow later introduction of twin AarI sites to facilitate tracer RNA cloning. pAAV-Puro_siKD was digested with AarI and MreI to relase a short fragment containing the AarI site, which was replaced exactly but with a single point mutation disrupting the AarI site (note that AarI is a type II restriction enzyme that cuts outside of its recognition sequence). For this, a double stranded DNA fragment was generated by annealing oligos 5'-CCGGCGCTCCGGGGGCCGCCGCGCCCCTCCCCCGAGCCCTCCCCGGCCCGAGGCGGCCCCGCCCCGCCCGGCA CCCCCACCAGCCGCC-3' and 5'- GGGTGGCGGCTGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCC-GGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCG-3', and then ligated. The resulting vector was digested with BglII and SalI, and ligated to a fragment containing twin AarI sites followed by a gRNA scaffold, so that this was placed following the H1-TO promoter. This fragment was obtained by annealing oligos 5'-GATCCCTGGTGCAGGTGG-ACTGACTCACCTGCACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTG-3' and 5'-TCGACAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGT-TGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGCAGGTGAGTCAGTCCACCTG-CACCAGG-3'. The resulting vector was named pAAV-Puro_siKO. Finally, tracer RNA sequences were cloned between the twin AarI sites to generate a full gRNA. For this, pAAV-Puro_siKO was digested with 2 U of AarI (Thermo Fisher Scientific) per µg of plasmid in 1x AarI buffer supplemented with 0.5 µM AarI oligo at 37°C for 16 h (overnight), gel extracted following DNA electrophoresis, and reconstituted at 50 ng μl^{-1} for subsequent ligation. Validated EGFP tracer RNA sequences were previously described (Shalem et al., 2014), and complementary single-stand oligonucleotides containing the tracer and appropriate overhangs (see table below) were mixed, annealed, and ligated to the prepared pAAV-Puro_siKO as described above for the generation of pAAV-Puro_iKD vectors. Following transformation, bacterial colonies were screened by colony PCR using the forward primer 5'-CGAACGCTGACGTCATCAACC-3', and the bottom tracer RNA oligo as reverse primer. This reaction was performed as described above for pAAV-Puro_iKD vectors, and resulted in a band of approximately 250 bp in correct cloning products. These were expanded, and the gRNA sequence was confirmed by Sanger sequencing of isolated plasmids using the same forward primer employed for colony PCR.

pAAV-Puro_siKO-2TO: the H1-2TO promoter was designed by replacing the 19bp sequence before the TATA box of the H1-TO promoter with an additional tet operon (see Fig. S8J), as previously described (Henriksen et al. 2007). This sequence was obtained as double a stranded synthetic DNA molecule also containing the gRNA cloning site, the gRNA scaffold (Integrated DNA Technologies), and up- and downstream homology regions to allow subsequent Gibson cloning (5'-GTGGGCTCTATGGGTCAATT-3' and 5'-GACCTCGACTAGTTATTAATAGTAA-3', respectively). The resulting product was Gibson cloned in the pAAV-Puro_siKO following BstBI and HincII digestion. This pAAV-Puro_siKO-2TO vector was then used to clone the EGFP gRNAs exactly as just described for pAAV-Puro_siKO.

Oligonucleotide sequences used for gRNA cloning

gRNA	Top oligo[a]	Bottom Oligo[a]
EGFP-1	TCCC**GAGCTGGACGGCGACGTAAA**	AAAC**TTTACGTCGCCGTCCAGCTC**
EGFP-2	TCCC**GAAGTTCGAGGGCGACACCC**	AAAC**GGGTGTCGCCCTCGAACTTC**
EGFP-3	TCCC**GGGCGAGGAGCTGTTCACCG**	AAAC**CGGTGAACAGCTCCTCGCCC**

[a] The 20 base-pairs gRNA sequence is in bold, while other sequences provide the necessary overhangs for the cloning.

pAAV-Puro_MsiKO plasmids: vectors carrying multiple inducible gRNAs were generated by Gibson assembly of PCR products of individual gRNA cassettes that had previously been cloned in pAAV-Puro_siKO. This was performed using a strategy analogous to the one described above for pAAV-Puro_MsiKD vectors, with the following few exceptions. First, some of the primers used to generate fragments for Gibson assembly were different (see table below). Second, pAAV-Puro_siKO was used for cloning following sequential digestion with BstBI (2 h at 50°C) and HincII (2 h at 37°C) to remove the pre-existing inducible H1 promoter and scaffold RNA. Finally, colony PCR was performed using primers fw 5'-GTGTCATTCTATTCTGGGGGGTG-3' and rev 5'-GGGCTATGAACTAATGACCCCG-3' and resulted in bands of approximately 890 bp and 1288 bp for vectors containing 2 or 3 gRNAs, respectively.

Spacers used for Gibson assembly of multiple gRNA vectors

Spacer	Primer FW	Primer REV
5'-L_gRNA1_BL1-3'	TGCGGTGGGCTCTATGGGTCAATTCGAACGCTG ACGTCATCAAC	GTCTGCTGTCAAGACACAGCATAGTCCTAGTAA AGCTTAGTACTGTCCGGGCCCCCCCTCGAGGTC GACAAAAAAGCACCGACTCGG
5'-BL1_gRNA2_R-3'	GCTGTGTCTTGACAGCAGACCTCGTCGAACGCT GACGTCATCAAC	ATTGATTACTATTAATAACTAGTCGAGGTCGAC AAAAAAGCACCGACTCGG
5'-BL1_gRNA2_BL2-3'	GCTGTGTCTTGACAGCAGACCTCGTCGAACGCT GACGTCATCAAC	GTTGCCGACAGTACTTGTGTGTCCACCTAGTAA AGCTTAGTACTGTCCGGGCCCCCCCTCGAGGTC GACAAAAAAGCACCGACTCGG
5'-BL2_gRNA3_R-3'	ACACAAGTACTGTCGGCAACCACACCGAACGCT GACGTCATCAAC	ATTGATTACTATTAATAACTAGTCGAGGTCGAC AAAAAAGCACCGACTCGG

Gene targeting

For OPTiKD and sOPTiKD, AAVS1 targeting was performed by lipofection as previously described (Bertero et al., 2015; Vallier et al., 2004). Briefly, hPSCs were seeded feeder-free in 6-well plates, and transfected 48 h following cell passaging with 4 µg of DNA (equally divided between the two AAVS1 ZFN plasmids and the targeting vector) using 10 µl per well of Lipofectamine 2000 in Opti-MEM media (Gibco) for 24 h, all according to manufacturer's instructions. After 4 days, 1 µg ml^{-1} of Puromycin was added to the culture media, and individual clones were picked and expanded following 7-10 days of selection.

For sOPTiKO, AAVS1 targeting was performed by nucleofection. hESCs pre-treated for 16 h with 10 µM Y-27632 (Tocris) were dissociated to clumps of 2-8 cells using Accutase (Gibco), and 2 x 10^6 cells were nucleofected in 100 µl with a total of 12 µg of DNA (4 µg each for the two ZFN plasmids, and 2 µg each for the two targeting vectors) using the Lonza P3 Primary Cell 4D-Nucleofector X Kit and the cycle CA-137 on a Lonza 4D-Nucleofector System, all according to manufacturer's instructions. Nucleofected hESCs were plated onto a feeder layer of irradiated DR4 (puromycin and neomycin resistant) mouse embryonic fibroblasts and cultured in KSR media supplemented with 4 ng ml^{-1} FGF2 and 10 µM Y-27632 (this last only for the first 24 h). After 4 days, hPSC colonies carrying both puromycin and neomycin resistance gene were selected for 7-10 days with 25 µg ml^{-1} of Geneticin (G418 Sulfate, Gibco) and 0.5 µg ml^{-1} Puromycin. Individual clones were then picked and expanded in feeder-free conditions as described above.

AAVS1-EGFP, ROSA26-EGFP, ROSA26-STDtetR, ROSA26-OPTtetR, and ROSA26-EGFPd2 hESCs were generated by lipofection (AAVS1 locus) or nucleofection (ROSA26 locus) of the targeting vectors with AAVS1 ZFN or ROSA26 CRISPR/Cas9n pairs (as described above). 2 µg ml^{-1} Blasticidin S-HCl (Gibco) was used for pR26-Bsd_CAG-EGFPd2 plasmid. Generation of inducible EGFP overexpression hESCs carrying ROSA26-rtTA and AAVS1-TRE-EGFP transgenes is described in detail elsewhere (MP, DO, AB, et al., manuscript in revision). Briefly, cells were sequentially gene targeted first by nucleofection of pR26-Neo_CAG-rtTA with ROSA26 CRISPR/Cas9n plasmids, then by lipofection of pAAV-Puro_TRE-EGFP with AAVS1 ZFN plasmids.

Gene targeted hPSC clonal lines were screened by genomic PCR to verify site-specific targeting, determine the number of alleles targeted, and exclude off-target integrations of the targeting plasmid (see Fig. S1A). The following section will clarify the primer combinations and PCR conditions used for the various targeting vectors. The results of all targeting experiments are summarized in Table S1, and only fully correctly targeted lines (no off target integrations) were used for the experiments presented. Karyotype analysis was performed by standard G banding techniques to confirm euploidy of targeted lines (Medical Genetics Service, Cambridge University Hospitals).

Genotyping

The three tables below show the primer pairs and PCR conditions used for genotyping of targeted hPSCs.

Targeting vector(s)	PCR type	Primer location	Primer sequence	Amplicon wild-type (bp)	Amplicon transgene (bp)	Amplicon plasmid[a] (bp)	Temp. ann.[b] (°C)	Ext. time[b]
All pR26 vectors	LOCUS	Genomic; 5' to 5'-HAR	GAGAAGAGGCTGTGCTTCGG	2186	Variable[c] (can be loA)	No band	63	>3''[c]
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
All pR26 vectors	5'-INT	Genomic; 5' to 5'-HAR	GAGAAGAGGCTGTGCTTCGG	No band	1264	No band	60	1'30"
		Splice acceptor	AAGACCGCGAAGAGTTTGTCC					
pR26-Puro_ENDO-EGFP pR26-Neo_EF1α -EGFP pR26-Neo_CAG-EGFP pR26-Bsd_CAG-EGFPd2	3'-INT	EGFP	CTACCCCGACCACATGAAGC	No band	1746	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
pR26-Neo_CAG-STDtetR	3'-INT	STDtetR	CGACGCCTTAGCCATTGAGA	No band	1706	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
pR26-Neo_CAG-OPTtetR	3'-INT	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	1548	No band	60	2'
		Genomic; 3' to 3'-HAR	ACAGTACAAGCCAGTAATGGAG					
All pR26 vectors	5'-BB	Backbone; 5' to 5'-HAR	CGTTGTAAAACGACGGCCAG	No band	No band	1148	60	1'30"
		Neomycin	GTGCCCAGTCATAGCCGAAT					
pR26-Puro_ENDO-EGFP pR26-Neo_EF1α -EGFP pR26-Neo_CAG-EGFP pR26-Bsd_CAG-EGFPd2	3'-BB	EGFP	ACCACTACCAGCAGAACAC	No band	No band	1315	60	1'30"
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					
pR26-Neo_CAG-STDtetR	3'BB	STDtetR	CCCGTAAACTCGCCCAGAAG	No band	No band	1681	60	2'
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					
pR26-Neo_CAG-OPTtetR	3'BB	OPTtetR	AGAAACTGGGCGTGGAACAG	No band	No band	1736	60	2'
		Backbone; 3' to 3'-HAR	TGACCATGATTACGCCAAGC					

[a] Result of PCR on targeting vector (positive control for off-target plasmid integration).

[b] Variable parameters in PCR program: (1) 94° 5'; (2) 94° 15''; (3) Temp. ann. 30"; (4) 65° Ext. time; (5) Repeat 2 to 4 for 34 cycles; (6) 65° 5'; (7) 10° hold.

[c] Size depending on transgene size; the extension time was set accordingly. For CAG promoter-containing transgenes the PCR failed due to high GC-content and resulted in loss-of-allele (loA).

Targeting vector(s)	PCR type	Primer location	Primer sequence	Amplicon wild-type (bp)	Amplicon transgene (bp)	Amplicon plasmid[a] (bp)	Temp. ann.[b] (°C)	Ext. time[b]
All pAAV vectors	LOCUS	Genomic; 5' to 5'-HAR	CTGTTTCCCCTTCCCAGGCAGGTCC	1692	Variable[c] (can be loA)	No band	65	> 3'[c]
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					
All pAAV-Puro vectors	5'-INT	Genomic; 5' to 5'-HAR	CTGTTTCCCCTTCCCAGGCAGGTCC	No band	1103	No band	65	1'30"
		Puromycin	TCGTCGCGGGTGGCGAGGCGCACCG					
pAAV-Neo_CAG-Cas9	5'-INT	Genomic; 5' to 5'-HAR	CTGTTTCCCCTTCCCAGGCAGGTCC	No band	1032	No band	60	1'30"
		Neomycin	GTGCCCAGTCATAGCCGAAT					
pAAV-Puro_CAG-EGFP pAAV-Puro_EGFPiKD	3'-INT	EGFP	GGATCACTCTCGGCATGGAC	No band	1656	No band	60	2'
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					
All pAAV-Puro_iKD vectors	3'-INT	After H1 promoter	GCCCGGACAGTACTAAGCTTTACTAGGG	No band	884	No band	60	1'
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					
All pAAV-Puro_siKD vectors pAAV-Puro_EGFPsiKD-OPT All pAAV-Puro_siKO vectors All pAAV-Puro_MsiKO vectors	3'-INT	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	1447	No band	60	1' 30"
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					
pAAV-Puro_EGFPsiKD-STD	3'-INT	STDtetR	CGACGCCTTAGCCATTGAGA	No band	1605	No band	60	2'
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					
pAAV-Neo_CAG-Cas9	3'-INT	Cas9	AAACGGCCGGAAGAGAATG	No band	2151	No band	60	2'30"
		Genomic; 3' to 3'-HAR	TGCAGGGGAACGGGGCTCAGTCTGA					

Development 143: doi:10.1242/dev.138081: Supplementary information

Targeting vector(s)	PCR type	Primer location	Primer sequence	Amplicon wild-type (bp)	Amplicon Transgene (bp)	Amplicon plasmid[a] (bp)	Temp. ann.[b] (°C)	Ext. time[b]
All pAAV-Puro vectors	5’-BB	Backbone; 5' to 5’-HAR	ATGCTTCCGGCTCGTATGTT	No band	No band	1227	60	1’30”
		Puromycin	TGAGGAAGAGTTCTTGCAGCTC					
pAAV-Neo_CAG-Cas9	5’-BB	Backbone; 5' to 5’-HAR	ATGCTTCCGGCTCGTATGTT	No band	No band	1249	60	1’30”
		Neomycin	GTGCCCAGTCATAGCCGAAT					
pAAV-Puro_CAG-EGFP pAAV-Puro_EGFPiKD	3’-BB	EGFP	GGATCACTCTCGGCATGGAC	No band	No band	2011 (EGFP) 1526	60	2’ 30”
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAGTT					
All pAAV-Puro_iKD vectors	3’-BB	After H1 promoter	GCCCGGACAGTACTAAGCTTTACTAGGG	No band	No band	1239	60	1’ 30”
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAGTT					
All pAAV-Puro_siKD vectors pAAV-Puro_EGFPsiKD-OPT All pAAV-Puro_siKO vectors All pAAV-Puro_MsiKO vectors	3’-BB	OPTtetR	CCACCGAGAAGCAGTACGAG	No band	No band	1802	60	2’
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAGTT					
pAAV-Puro_EGFPsiKD-STD	3’-BB	STDtetR	CGACGCCTTAGCCATTGAGA	No band	No band	1960	60	2’
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAGTT					
pAAV-Neo_CAG-Cas9	3’-BB	Cas9	AAACGGCCGGAAGAGAATG	No band	No band	2500	60	2’30”
		Backbone; 3' to 3'-HAR	ATGCACCACCGGGTAAAGTT					

Development 143: doi:10.1242/dev.138081: Supplementary information

Quantitative real-time PCR (qPCR)

Analysis of gene expression was performed as previously described (Bertero et al., 2015). Briefly, RNA was extracted using GenElute Mammailan Total RNA Miniprep Kit and the On-Column DNAse I Digestion Set (Sigma-Aldrich) and used for cDNA synthesis with SuperScript II (Invitrogen), all following manufacturer's recommendations. SensiMix SYBR low-ROX (Bioline) was used for qPCR, and reactions were run on a Stratagene Mx-3005P (Agilent). See the table below for primer sequences. Results were analyzed with the ΔΔCt method (Schmittgen and Livak, 2008) using *RPLP0* as housekeeping gene. Perseus software (MaxQuant) was used to generate heatmaps summarizing the qPCR data.

Primers used for quantitative PCR

Gene	Primer	Sequence
ACTA2	FW	GTGTTGCCCCTGAAGAGCAT
	REV	GCTGGGACATTGAAAGTCTCA
ACTN2	FW	CAAACCTGACCGGGGAAAAAT
	REV	CTGAATAGCAAAGCGAAGGATGA
AFP	FW	AGAACCTGTCACAAGCTGTG
	REV	GACAGCAAGCTGAGGATGTC
B2M	FW	ATGTCTCGCTCCGTGGCCTTAGCT
	REV	CCTGAATCTTTGGAGTACGCTGGATAGC
CCND1	FW	GATGCCAACCTCCTCAACGA
	REV	TCTGTTCCTCGCAGACCTCC
CCND2	FW	AGCTGTCTCTGATCCGCAAG
	REV	TGGCAAACTTAAAGTCGGTGG
CCND3	FW	GGCCCTCTGTGCTACAGATTA
	REV	CAGTCCACTTCAGTGCCAGT
CDX2	FW	GGCAGCCAAGTGAAAACCAG
	REV	TTCCTCTCCTTTGCTCTGCG
CER1	FW	TTCTCAGGGGGTCATCTTGC
	REV	ATGAACAGACCCGCATTTCC
CXCR4	FW	CACCGCATCTGGAGAACCA
	REV	GCCCATTTCCTCGGTGTAGTT
DPY30	FW	TGCTGGAGGGACAAACGCAGG
	REV	AGGCACGAGTTGCAAAGACTGG
EGFP	FW	CCCGACAACCACTACCTGAG
	REV	GTCCATGCCGAGAGTGATCC
EOMES	FW	ATCATTACGAAACAGGGCAGGC
	REV	CGGGGTTGGTATTTGTGTAAGG
FOXA2	FW	GGGAGCGGTGAAGATGGA
	REV	TCATGTTGCTCACGGAGGAGTA
GATA4	FW	TCCCTCTTCCCTCCTCAAAT
	REV	TCAGCGTGTAAAGGCATCTG
GCG	FW	AAGCATTTACTTTGTGGCTGGATT
	REV	TGATCTGGATTTCTCCTCTGTGTCT
GSC	FW	GAGGAGAAAGTGGAGGTCTGGTT
	REV	CTCTGATGAGGACCGCTTCTG
HEX	FW	CTGCAGCTCAGCGAGAGACA
	REV	CAGGGGAGGGCGAACATGGA
HFN4A	FW	CATGGCCAAGATTGACAACCT
	REV	TTCCCATATGTTCCTGCATCAG
HLXB9	FW	CACCGCGGGCATGATC
	REV	ACTTCCCCAGGAGGTTCGA
HNF1B	FW	GCACCCCTATGAAGACCCAG
	REV	GGACTGTCTGGTTGAATTGTCG
INS	FW	CAGGAGGCGCATCCACA
	REV	AAGAGGCCATCAAGCAGATCA

Gene	Primer	Sequence
ISL1	FW	GCAAATGGCAGCGGAGCCCA
	REV	AGCAGGTCCGCAAGGTGTGC
KDR	FW	TTTTTGCCCTTGTTCTGTCC
	REV	TCATTGTTCCCAGCATTTCA
MAP2	FW	AGACTGCAGCTCTGCCTTTAG
	REV	AGGCTGTAAGTAAATCTTCCTCC
MEF2C	FW	CAGACATCGTGGAGGCATT
	REV	GGGGTGAGTGCATAAGAGGA
MESP1	FW	GAAGTGGTTCCTTGGCAGAC
	REV	TCCTGCTTGCCTCAAAGTGT
MEOX1	FW	AAAGTGTCCCCTGCATTCTG
	REV	CACTCCAGGGTTCCACATCT
MHY6	FW	CTCCCGCTTTGGGAAATT
	REV	GGACTTCTCCAGCAGGTAGGT
MIXL1	FW	GGTACCCCGACATCCACTTG
	REV	TAATCTCCGGCCTAGCCAAA
MSGN1	FW	GGAGAAGCTCAGGATGAGGA
	REV	GTCTGTGAGTTCCCCGATGT
NANOG	FW	CATGAGTGTGGATCCAGCTTG
	REV	CCTGAATAAGCAGATCCATGG
NKX2.5	FW	GAGCCGAAAAGAAAGCCTGAA
	REV	CACCGACACGTCTCACTCAG
NGN2	FW	TGTTCGTCAAATCCGAGACCT
	REV	CGATCCGAGCAGCACTAACA
NODAL	FW	TGAGCCAACAAGAGGATCTG
	REV	TGGAAAATCTCAATGGCAAG
OCT4/POU5F1	FW	AGTGAGAGGCAACCTGGAGA
	REV	ACACTCGGACCACATCCTTC
PAX3	FW	GTGCCGTCAGTGAGTTCCATC
	REV	AAGTCACCCAGCAAGTGCG
PAX6	FW	CTTTGCTTGGGAAATCCGAG
	REV	AGCCAGGTTGCGAAGAACTC
PDGFRA	FW	TTTTGTTGTAGAGGTGCGGG
	REV	TCCTTAGCACGGATCAGCTT
PDX1	FW	GATTGGCGTTGTTTGTGGCT
	REV	GCCGGCTTCTCTAAACAGGT
RPLP0	FW	GGCGTCCTCGTGGAAGTGAC
	REV	GCCTTGCGCATCATGGTGTT
SOX1	FW+ REV	Quantitec primers (Qiagen): QT00215299
SOX17	FW	CGCACGGAATTTGAACAGTA
	REV	GGATCAGGGACCTGTCACAC
SOX2	FW	TGGACAGTTACGCGCACAT
	REV	CGAGTAGGACATGCTGTAGGT
SOX9	FW+ REV	Quantitec primers (Qiagen): QT00001498
SST	FW	CCCCAGACTCCGTCAGTTTC
	REV	TCCGTCTGGTTGGGTTCAG
T	FW	TGCTTCCCTGAGACCCAGTT
	REV	GATCACTTCTTTCCTTTGCATCAAG
TAGLN	FW	TCTTTGAAGGCAAAGACATGG
	REV	TTATGCTCCTGCGCTTTCTT
TBX5	FW	GCTGGAAGGCGGATGTTT
	REV	GATCGTCGGCAGGTACAATG
TBX6	FW	AAGTACCAACCCCGCATACA
	REV	TAGGCTGTCACGGAGATGAA
tetR	FW	CGACGCCTTAGCCATTGAGA
	REV	TTTCTGTAGGCCGTGTACCT

Gene	Primer	Sequence
TNTT2	FW	ACAGAGCGGAAAAGTGGGAAG
	REV	TCGTTGATCCTGTTTCGGAGA
VGLUT2/SLC17A6	FW	GTAGACTGGCAACCACCTCC
	REV	CCATTCCAAAGCTTCCGTAGAC

Flow cytometry

For EGFP quantification in live cells, cultures were incubated with TrypLE Select (Gibco) for 5-20 minutes at 37° C to obtain a single cell suspension. Following a wash in PBS, cells were resuspended in ice-cold PBS 1% BSA with 5 µg ml^{-1} DAPI, incubated for 5 minutes on ice, and filtered through a 100 µm cell strainer. Cells were analyzed using a Cyan ADP flow cytometer to determine the EGFP median fluorescence intensity (MFI) of viable cells (DAPI negative). Flow cytometry analysis was performed using FlowJo.

For intracellular staining, single cell suspensions were prepared as described above, washed and then fixed for 20 minutes at 4°C with PBS 4% PFA. After three washes with PBS, cells were first permeabilized for 20 minutes at room temperature with PBS 0.1% Triton X-100, then blocked for 30 minutes at room temperature with PBS 10% Donkey Serum. Primary antibodies (see table below) and Alexa Fluor 647 Donkey secondary antibodies were incubated for 1 hour each at room temperature in PBS 1% Donkey Serum 0.1% Triton X-100, and cells were washed three times with this same buffer after each incubation. Flow cytometry was performed using a Cyan ADP flow cytometer and at 10,000-50,000 events were recorded. Secondary only stains were used as negative controls for all experiments.

Western blot

Analysis of protein expression was performed as previously described (Bertero et al., 2015). Briefly, protein lysates were prepared using CellLytic M (Sigma-Aldrich) supplemented with cOmplete Protease Inhibitor (Roche), and quantified using Protein Quantification Kit-Rapid (Sigma-Aldrich), all according to manufacturer's instructions. Protein electrophoresis was performed using NuPAGE LDS Sample Buffer and 4-12% NuPAGE Bis-Tris Precast Gels (Invitrogen). Following proteins transfer on PVDF using NuPAGE Transfer Buffer (Invitrogen), membranes were blocked with PBS 0.05% Tween (PBST) 4% milk for 1 h at room temperature, and incubated overnight with primary antibodies (see table below) in PBST 4% milk. Membranes were washed three times with PBST, incubated with HRP-conjugated secondary antibodies (Sigma-Aldrich) in PBST 4% milk, further washed three times with PBST, incubated with Pierce ECL2 Western Blotting Substrate (Thermo), and exposed to X-Ray Super RX Films (Fujifilm).

Immunofluorescence

For most cell types analyzed, immunostaining was performed as previously described (Bertero et al., 2015). Briefly, cells were fixed in PBS 4% PFA for 20 minutes at room temperature, rinsed three times with PBS, blocked and permeabilized with PBS 10% Donkey Serum (Biorad) 0.1% Triton X-100 for 30 minutes at room temperature, and incubated with primary antibodies diluted in PBS 1% Donkey Serum 0.1% Triton X-100 overnight at 4°C (see table below). Following three washes for 5 minutes with PBS at room temperature, cells were incubated with appropriate Alexa Fluor 568 and/or 647 Donkey secondary antibodies diluted in PBS 1% Donkey Serum 0.1% Triton X-100 for 1 h at room temperature, and finally further washed three times for 5 minutes with PBS at room temperature (5 µg ml^{-1} DAPI was added to the first wash to stain nuclei). The immunostaining protocol was modified for cells growing in three-dimensional cultures (cholangiocytes and intestinal organoids) by extending each wash step to 1 h and by performing the incubation with secondary antibodies overnight at 4°C. For all cell types analyzed, EGFP expression could be monitored directly and without the use of antibodies, as the immunostaining procedure largely preserved EGFP fluorescence. Immunostainings were imaged using an LSM 700 confocal microscope (Leica).

Antibodies

Antigen	Supplier	Product	WB	IF	Flow
ACAN/aggrecan	R&D	AF1220	-	1:50	-
ACTN2/cardiac alpha actinin	Sigma-Aldrich	A7811	-	1:500	-
AFP/alpha fetoprotein	Dako	A0008	-	-	1:100
ALB	Bethyl Laboratories	A80-229A-3	-	1:100	-
BGLAP/osteocalcin	R&D	MAB1419	-	1:25	-
Cas9	Cell Signalling	14697	-	1:50	-
CDX2	BioGenex	MU392A-UC	-	1:200	-
CNN1/calponin	Sigma	C2687	-	-	1:5000
DDR2	Santa Cruz	sc-7555	-	1:50	-
DPY30	Sigma-Aldrich	HPA043761	1:100	1:200	-
GFAP	DAKO	Z0334	-	1:1000	-
HNF4A	Santa Cruz	sc-8987	-	1:100	-
INS/c-peptide (pro-insulin)	Acris	BM270S	-	1:300	1:200
ISL1	Abcam	ab23345	-	-	-
KRT19/citokeratin 19	Abcam	ab7754	-	1:100	-
MAP2	Sigma-Aldrich	M4403	-	1:200	-
NGFR/p75	Santa Cruz	sc-6188	-	1:100	-
NKX2.5	Santa Cruz	sc-14033	-	1:200	-
O4	Sigma-Aldrich	O7139	-	1:250	-
OCT4/POU5F1	Santa Cruz	sc-5279	-	1:200	1:200
SFPTC/Surfactant protein c	Santa Cruz	sc-7705	-	1:50	-
SOX1	R&D	AF3369	-	1:100	-
SOX17	R&D	AF1924	-	1:100	1:200
TAGLN/SM22 alpha	Abcam	ab14106	-	1:1000	-
T/Brachyury	R&D	AF2085	-	1:200	-
tetR	Clonetech	631131	1:1000	-	-
tetR	Mobitec	TET01	-	1:4000	-
TNTT2/cardiac troponin T	Abcam	ab45932	-	1:500	1:300
TUBA4A/alpha 4 tubulin	Sigma-Aldrich	T6199	1:10000	-	-
TUBB3/beta 3 tubulin	Millipore	MAB1637	-	1:1000	1:200
VIL1/villin	Santa Cruz	sc-58897	-	1:100	-
WT1	Abcam	ab89901	-	1:50	-

Appendix S1. Sequences of plasmids used in this study

Nucleotide sequences are provided (supplemental archive.zip) in each of two formats. (1) Genbank, which can be opened with virtually any nucleotide sequence analysis software (e.g. SnapGene Viewer, http://www.snapgene.com/products/snapgene_viewer/). The user will be able to visualize a number of features associated with the sequences. (2) Raw sequence in FASTA format, which can be opened with any text editor (but does not contain feature annotation).

Click here to Download Supplemental Archive.zip