4 Development of an inducible lineage tracing system to study *in vitro* differentiation

Marloes Blotenburg, Buys de Barbanson, Peter Zeller,
Reinier van der Linden, Anna Alemany, Alexander van Oudenaarden

Abstract

One of the key questions in developmental biology is how one totipotent cell can grow out into a whole complex organism. To understand this process, efforts have been made to follow lineage specification of an embryo as it develops. Classically, these approaches have been mostly imaging-based, thereby providing spatial resolution and detailed tracking of cell divisions and migration, but are hindered by lack of detailed cell typing. More recently, sequencing-based approaches have been developed, where complex mutations are introduced as barcodes to determine which cells are derived from a similar origin. Within the same single cell, this barcode read-out can be coupled with identification of the transcriptome, thus allowing for detailed cell typing. ScarTrace is one such technique, in which CRISPR-Cas9 has been used to introduce complex mutations, or scars, in the developing zebrafish embryo, and read out together with the transcriptome at the single cell level. Here, we introduce an novel design for more dynamic labelling based on the original ScarTrace protocol. We adapted the system to be inducible, to target endogenous loci with single-molecule resolution and to function in a mammalian model. In addition, activation can be easily observed through the integration of a colour switch coupled to Cas9-expression. We applied this design to an *in vitro* mammalian system termed gastruloids. These are three-dimensional aggregates of mouse embryonic stem cells (mESCs) that over the course of five days develop cell types derived from all three germ layers, recapitulate key aspects such as axial organisation and symmetry breaking and resemble the mouse embryo at E8.5. We detect cell types derived from all three germ layers, allowing for complex lineage tracing of cell differentiation from pluripotency until the formation of a body plan. We show that upon induction of this lineage tracing system, scars build up over time, thus allowing us to identify cellular barcodes and define lineage relationships. However, we encountered multiple technical challenges, including silencing upon differentiation, leaky activation before induction, and the accumulation of pre-scars. Therefore, we conclude that the system in its current form is unsuitable for a dynamic differentiation system and propose several alternatives that should produce improved results.

Introduction

During embryonic development, the totipotent zygote eventually develops into a fully differentiated complex multicellular organism. How one cell can give rise to all different lineages is a fundamental question in developmental biology, and the mechanisms involved in this process are still being investigated. Perhaps the most striking is the fact that one DNA sequence is able to generate hundreds of distinct cell types. To understand how one totipotent cell generates all differentiated cells that make up an organism, efforts have been made to track this process over time. For deterministic organisms with few cells in total such as *C. elegans*, it is possible to track all cell divisions during its development and build a complete lineage tree[1]. However, higher order organisms follow a less deterministic path, complicating the quest for generating one unified lineage tree. Furthermore, in mammals the study of this process is technically obscured by their development *in utero*. Historically, parts of this progress have been unveiled during mouse development by injecting one cell with dye and tracking coloured cells during a limited time-frame[2,3,4,5]. Later, these dye injections have been replaced by fluorescent reporters. Systems such as Zebrabow and Brainbow make use of this strategy by combining multiple fluorophores, thereby increasing the number of cells that can be uniquely labelled[6,7]. Imaging techniques have also vastly improved, allowing the tracking of mouse embryo development in real time and building partial lineage trees through reverse fate mapping without the need for fluorescent reporters[8]. However, despite the precise tracking of cell positions and cell divisions, the major drawback of imaging approaches is the lack of detailed cell typing.

The combination of genetic barcoding and single-cell sequencing has made it possible to overcome this limitation. Genetic barcodes can be established through integration of a complex pool of lentiviral barcodes[9,10], or by using dynamic genome editing of a defined locus with CRISPR-Cas9. The latter introduces double-stranded breaks, which during non-perfect DNA repair introduces variable mutations functioning as barcodes. Using this technology, the embryonic development of zebrafish[11,12,13,14,15] and later of mouse embryos[16,17] has been traced from single-cell to whole-organism level. Most lineage tracing approaches applied to zebrafish have relied on injections with the Cas9 protein and a gRNA at the one-cell stage, thereby initiating scar build-up from the zygote stage onwards. Zebrafish contain 1000 cells already 3 hours post fertilization (hpf), go through gastrulation at 5 hpf and contain all cell types within 24 hours of development. Because of this fast development, most scars are introduced during and after cell fate decisions are being made. In contrast, mammalian organisms such as mice generally develop much slower, and only complete one cell division in 24 hours. Therefore, a similar lineage tracing approach would only allow lineage tracing of a limited window of development, since scar target sites would be saturated before the completion of development[16]. In an attempt to circumvent this, a technique has been developed to target the gRNA sequences themselves, thereby maintaining a continuous cycle of mutations[17]. A draw-back of this method is that when mutations are produced, they are not preserved indefinitely but continuously re-mutated instead. Over time, this obscures the mutations that were first introduced and therefore complicates the

reconstruction of one lineage tree. Another possibility would be the development of an inducible CRISPR-Cas9 system, which allows introduction of scars at the time point of interest.

Some major challenges remain with *in vivo* lineage tracing in mammalian systems. Since embryos are encapsulated in the womb, it remains challenging to microscopically trace development or introduce genetic barcodes. In addition, the generation of the necessary transgenic mouse lines is laborious and time-consuming. In contrast, *in vitro* models are easier to genetically modify and can usually be grown in large quantities. Gastruloids have recently been developed as an *in vitro* model system that recapitulates aspects of embryogenesis[18, 19, 20]. These are three-dimensional aggregates of mouse embryonic stem cells (mESCs) that over the course of five days grow out into structures that display axial organisation, symmetry breaking and develop cell types derived from all three germ layers, resembling the mouse embryo at the E8.5 stage[19]. Recent advances have shown that this system is able to faithfully reproduce aspects of heart development[21], somitogenesis[22], neural tube development[23], and the development of anterior neural structures[24].

Here, we have adapted the ScarTrace technology to create an inducible lineage tracing method and applied it to gastruloids. The inducible lineage tracing system is encoded in one plasmid and can be used after introduction into any line expressing tamoxifen-inducible Cre-recombinase (Cre-ERT2). Since the system also introduces a fluorescent colour switch, cells with the (in)activated lineage tracer can be easily identified. This allows for a versatile application of the system, including to *in vivo* settings. We show that upon induction of our lineage tracing system, scars build up over time. In addition, the design allows us to separate the precise genomic and allelic scar locations, allowing us to track a maximum of 22 unique scar target sites. Furthermore, we show that upon differentiation into gastruloids, cell types derived from all three germ layers are detected, allowing for complex lineage tracing of cell differentiation from pluripotency until the formation of a body plan. However, once the lineage tracer is implemented in this differentiation system, it is silenced. We have also observed leaky activation before induction of the construct, resulting in pre-scars and uncontrolled activation of scarring. We conclude that the system in its current form is unusable and propose several alternatives that should produce improved results.

Results

Generation of inducible lineage tracing mESC lines with colour switch

Previous CRISPR-Cas9-based lineage tracing designs targeted artificially integrated constructs, and were not inducible[11, 12, 13, 14, 15]. To develop an inducible construct targeting endogenous loci without generating toxicity, we decided to make use of a Cre-ERT2 system and a custom lineage tracing construct (Figure 1A). This construct can be integrated using the Tol2 transposase system, making it applicable to *in vitro* as well as *in vivo* systems.

In short, the lineage tracing construct is driven by an SV40 promoter, resulting in expression of Neongreen and reading into a polyA termination site. Upon activation with tamoxifen, Cre recombinase is translocated to the nucleus,

where it removes the Neongreen and polyA cassettes, resulting in expression of the Cas9-Scarlet cassette. This colour switch allows the selection of cells which contain the construct in its non-induced (Neongreen$^+$) and induced (Scarlet$^+$) form using fluorescence-activated cell sorting (FACS). The gRNA, driven by a U6 promoter, is constitutively expressed. Once induced, the gRNA and Cas9 form a complex which targets 13 endogenous sites in the IgH locus (Figure 1B). This locus is not expressed except in differentiated B-cells, therefore mutations in this locus are unlikely to cause detrimental secondary effects. These 13 loci can be distinguished by sequencing, since all of them have unique site-specific base-pair changes. In addition, since we make use of a mESC line with a mixed 129/Bl6 genetic background we can separate the two alleles, resulting in a maximum of 26 target sites and single-molecule scar resolution (Figure S1A). Because of the repetitive character of the IgH locus, we can amplify 11 out of these 13 target sites with just one primer pair.

To generate a mouse embryonic stem cell (mESC) line, the inducible lineage tracing cassette was integrated into 129/Bl6 mESCs already expressing a Cre-ERT2 cassette in the ROSA26 locus[25], using the Tol2 transposase system. mESCs were transfected with the lineage tracing cassette and Tol2 protein, after which the Neongreen$^+$ fraction was enriched by FACS every week, once weekly, for three weeks consecutively. Single Neongreen$^+$ cells were sorted in individual wells and grown out into clonal lines. The resulting clonal lines 3, 5, 7, and 9 were induced for 48 hours using tamoxifen, causing a reduction of Neongreen fluorescence and an increase in Scarlet fluorescence over time (Figures 1C, S1B). Clonal line 5 showed the most efficient induction response and was therefore selected for use in subsequent experiments. To see if this fluorescence switch also results in a build-up of scars, the Scarlet$^+$ cell population of clone 5 was sorted during the 9-day time course and sequenced (Figures 1, S1). In addition, to detect if there are consistent differences between cells with different fluorescent colours, the Neongreen$^+$ and Scarlet$^+$ fraction of clone 5 were FACS sorted and sequenced four days after induction (Figures 1DE, S1B). Indeed, Scarlet$^+$ cells show an increase of both the total percentage of scars and number of unique scar sequences detected (Figure 1D-E).

In addition, reads from both alleles and from different target sites could be distinguished, allowing for a single-molecule read-out (Figures 1F, S1E). At four days after induction, Scarlet$^+$ cells show a greater build-up of scars than Neongreen$^+$ cells, in total as well as in individual target sites (Figures 1D-E, S1C-F). The fraction of negative cells show a scar build-up similar to Scarlet$^+$ cells and most likely contain an activated construct, but the fluorescent signal is not strong enough to classify them to the Scarlet$^+$ fraction (Figures 1D-E, S1B). These cells are excluded from further analyses. Finally, different scar patterns could also be observed in individual target sites: some sites show a gradual accumulation of scars, whereas other sites appear to be fully scarred already one day after induction (Figure 1F).

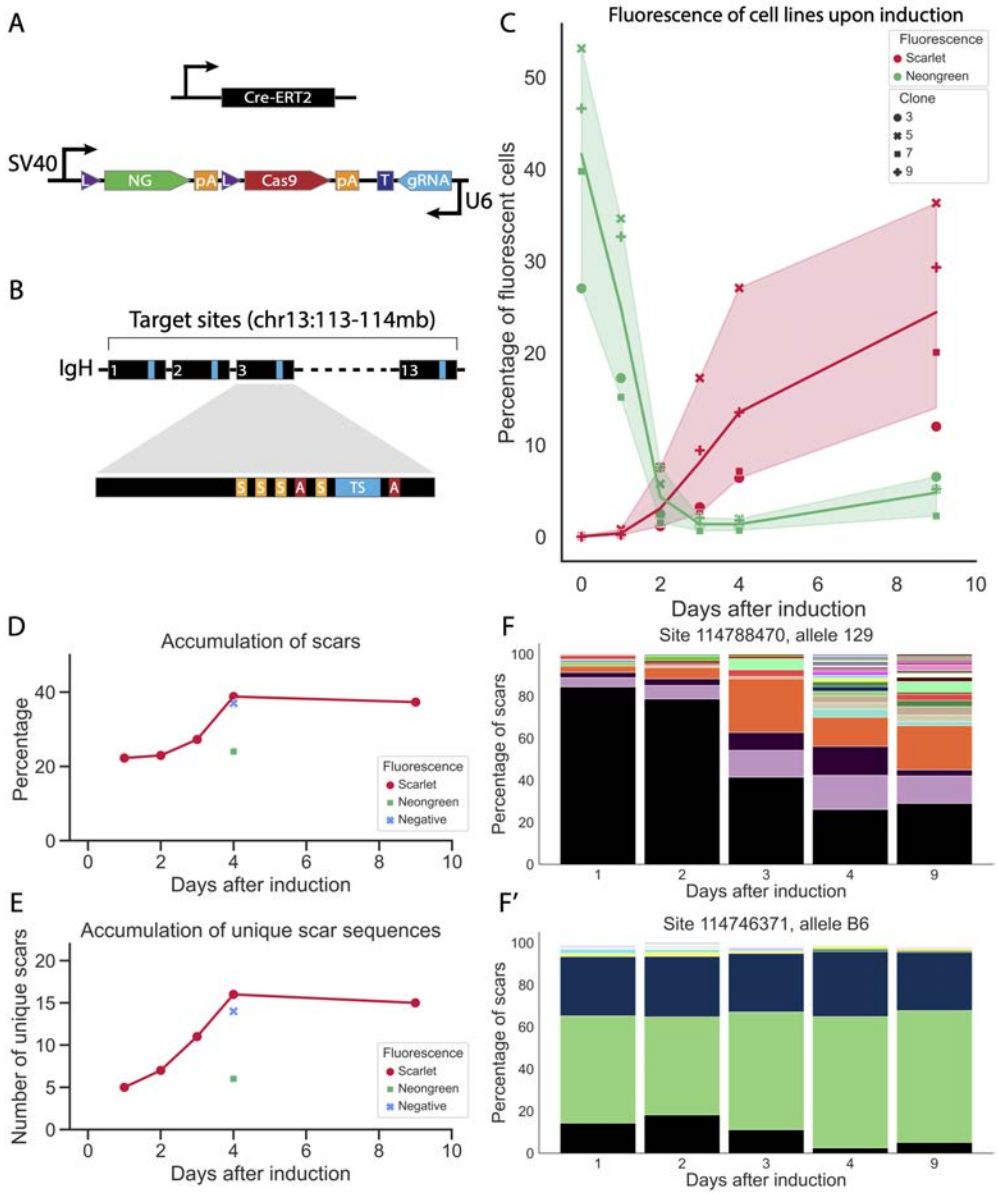

Figure 1 | (Continued on the following page.)

Figure 1 | Inducible ScarTrace set-up in clonal mESCs demonstrates fluorescent colour switch and build-up of scars. (**A**) The constructs cloned into mESCs. Top; Cells already contained a Cre-ERT2 construct cloned into the ROSA26 locus. Bottom; Lineage tracing construct that was transfected into the cells and integrated into the DNA using the Tol2 system. From the left: An SV40 promoter drives expression of Neongreen, followed by a PolyA stop signal, in between LoxP sites. Upon induction with tamoxifen, this cassette is cleaved out and a Cas9-Scarlet cassette is expressed instead. From the right: A U6 promoter drives expression of the gRNA sequence. (**B**) gRNA target sites located in the IgH locus. One gRNA targets 13 different genomic locations within this locus. Each site can be differentiated through site-specific SNPs. In addition, the two alleles can be distinguished by allele-specific SNPs, leading to a total of 26 unique target sites. (**C**) Colour switch in four clonal lines after induction with tamoxifen. At day 0, all four clones show Neongreen-only expression. Upon induction, Neongreen expression reduces while Scarlet expression increases. (**D**) DNA sequencing results of the induction time-course show an accumulation in Scarlet$^+$ cells from day 1 until day 4. At day 4, Negative and Neongreen$^+$ cells were also sorted and sequenced. While fluorescent negative cells show a similar degree of scarring as compared to Scarlet$^+$ cells, the Neongreen$^+$ cells show a lower degree of scars recovered. This is consistent with the expectation that Neongreen fluorescence indicates an inactive construct, while Scarlet fluorescence is coupled to Cas9 activity. Cells sorted in the negative gate probably contain an activated construct but do not show strong enough fluorescence to group them with Scarlet$^+$ cells. These cells are removed from further analyses. (**E**) DNA sequencing results of the induction time-course show an accumulation in the number of unique scar sequences recovered over time. The number of unique scars recovered increases from day 1 until day 4, when it levels off. In fluorescent negative cells, the number of unique scars recovered at day 4 is comparable to those of the Scarlet$^+$ fraction. The number of unique scar sequences recovered from Neongreen$^+$ cells at day 4 is lower. (**F**) Scar accumulation after induction in two example sites. The top site (**F**) shows a gradual accumulation of different scars, the bottom site (**F'**) shows the presence of non-WT sequences already from day 1 onwards. WT is black; each scar is a different colour.

Applying the ScarTrace system to gastruloids recovers distinct cell types but no lineage segregation

After confirming functionality of inducible ScarTrace, we applied the system to a complex *in vitro* model and differentiated clone 5 mESCs into gastruloids. We induced the formation of scars in mESCs for two days prior to gastruloid formation, and profiled the transcriptome and scar build-up of 18 individual gastruloids (Figures S2A-B) at 120 h after aggregation (AA). When looking at the transcriptome, we observed a mix of embryonic cell types derived from all three germ layers (Figure 2A), supported by specific expression of marker genes (Figures 2B, S2C). After filtering cells for standard parameters (Figure S2D), we counted the total number of recovered cells per gastruloid (Figure S2E), selected gastruloids with >150 cells and looked at the cell type distribution for these gastruloids (Figure 2C). While on average we sampled all cell types in all gastruloids, their distributions varied, especially when looking at the endoderm and neuronal fractions (Figure 2C).

Next, we looked at the recovered scar information. We first converted read counts to percentages and allele counts (Figures S2F-G) and looked at the number of detected scars per cell (Figure S2H). On average, we recover 9 sites per cell, of which 5 are mutated to scars. Next, we looked at which target sites we recover in the dataset as a whole. Recovered reads map to 9 out of the 13 target sites (Figure S2I). Specifically, we recover reads mapping to 8 target sites on the 129 allele and 6 target sites on the Bl6 allele (Figures S2J-K). When mapping all recovered target sites for recovered single cells, we noticed that several sites are sparsely covered (Figure S2L). After filtering out target sites with less than 80 % recovery across cells, six informative target sites remain (Figure 2D).

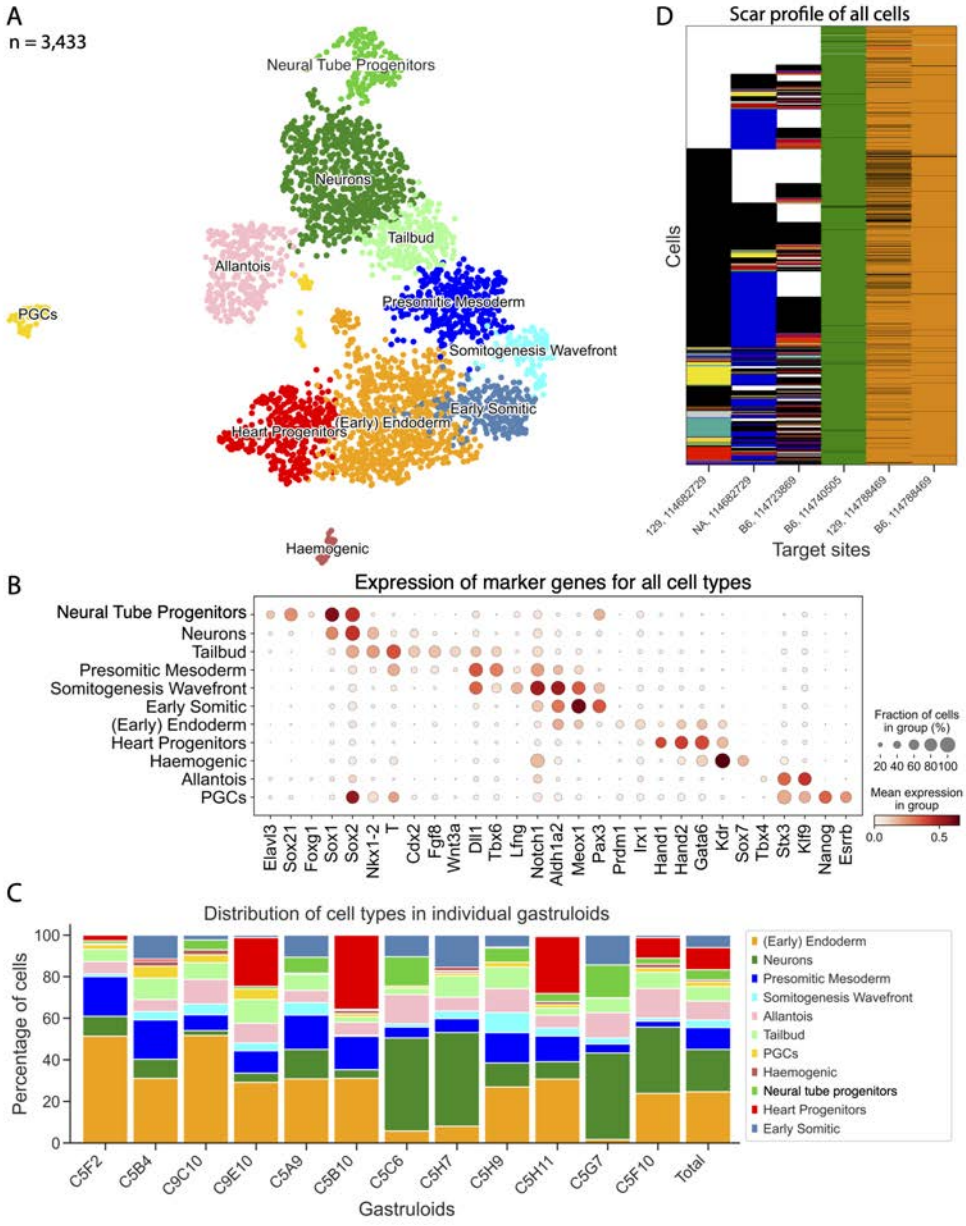

Figure 2 | (Continued on the following page.)

Figure 2 | ScarTrace in mouse gastruloids generated from clone 5 shows the recovery of distinct cell types from all three germ layers. (**A**) UMAP showing clustering of single-cell transcriptome data of all cells from all gastruloids. As reported previously, cell types from all three germ layers - endoderm, mesoderm, and ectoderm - are recovered[19]. Cell type annotation is based on localised expression of marker genes (Figures **B**, S2C). (**B**) Dot plot showing expression of marker genes used for cell type annotation. Dot colour represents the mean expression of marker genes in specific groups, and dot size represents the fraction of cells within the group that show expression of the gene. Expression of the genes included in this dot plot can be viewed in detail on the UMAP in Figure S2C. (**C**) Distribution of cell types in individual gastruloids. Cells per cell type in each gastruloid are divided by the total number of cells from that gastruloid, and colours correspond to the UMAP colours in (**A**). (**A**) Heatmap showing all scars for all gastruloids in filtered target sites. Target sites with > 80 % missing data are filtered out. A heatmap with all unfiltered target sites can be seen in Figure S2L. WT is black; missing data is white; each scar is assigned a unique colour, consistent with heatmaps and trees throughout this chapter.

Because each gastruloid develops isolated in a separate well, we classified individual gastruloids as replicates. By identifying the cell types of cells that share unique scar profiles within one gastruloid, we can study which lineages arise from a shared ancestor, and observe whether lineage compensation occurs. Therefore, we looked at the four largest sampled gastruloids individually. While in gastruloid C5F10 and C5H11 we sample most cell types (Figures 2C, 3A, S3A), C5G7 and C5H9 show a more skewed distribution, mostly lacking cells contributing to the heart progenitor cluster (Figure S3B-C). After clustering the cells derived from one gastruloid based on their scar profile, we observed the emergence of a few clear sub-clusters (Figure 3B). When looking at individual scars occurring in one specific target site, we noticed two different patterns. Some scar sequences are recovered in cells from all different cell types and show an un-localised distribution across the UMAP (Figure 3C). Others concentrate in one or few close cell types and show a more localised distribution (Figure 3D). To identify patterns, we created a scar profile by clustering the unique scar sequences of all recovered target sites for gastruloid C5F10 together (Figure 3E).

To answer the question whether we see co-emergence of related cell types and a shared scar profile between these cells, we next related scar profiles and cell type annotations of individual cells per gastruloid (Figures 3F, S3). We grouped the cells with identical scar profiles into nodes and calculated pseudotime based on the similarity of each node's scar profile to generate one lineage tree per gastruloid. We then added the transcriptome information and number of cells to each branch in the shape of dots coloured according to their respective cell type. The location of cells on the x-axis represents their localisation in UMAP space, so dots placed closely together in the lineage trees resemble cells with a highly similar transcriptome profile. It should be noted that since all gastruloids are generated from the same pool of induced mESCs, their scar profile also looks very similar (Figures 3E, S3).

In the resulting lineage trees, we observed some repeating patterns. For instance, PGCs (yellow) tend to segregate in small branches either by themselves or at most share with mesendodermal cell types such as allantois (pink), heart (red) and endoderm (orange). This could indicate a late emergence of PGCs, or a low division rate of these cells during gastruloid development. In contrast, neurons (green) most often populate larger branches containing a large amount of cells. Possibly, these cells are dividing fast during gastruloid development, leading to clonal expansion of its scar profile in the same lineage. For most cell types however, such clear patterns could not be identified (Figures 3F, S3). To conclude statistical significance, we calculated the enrichment of a certain cell type in one branch compared to the absolute ratio of cell types per gastruloid using the Fisher test.

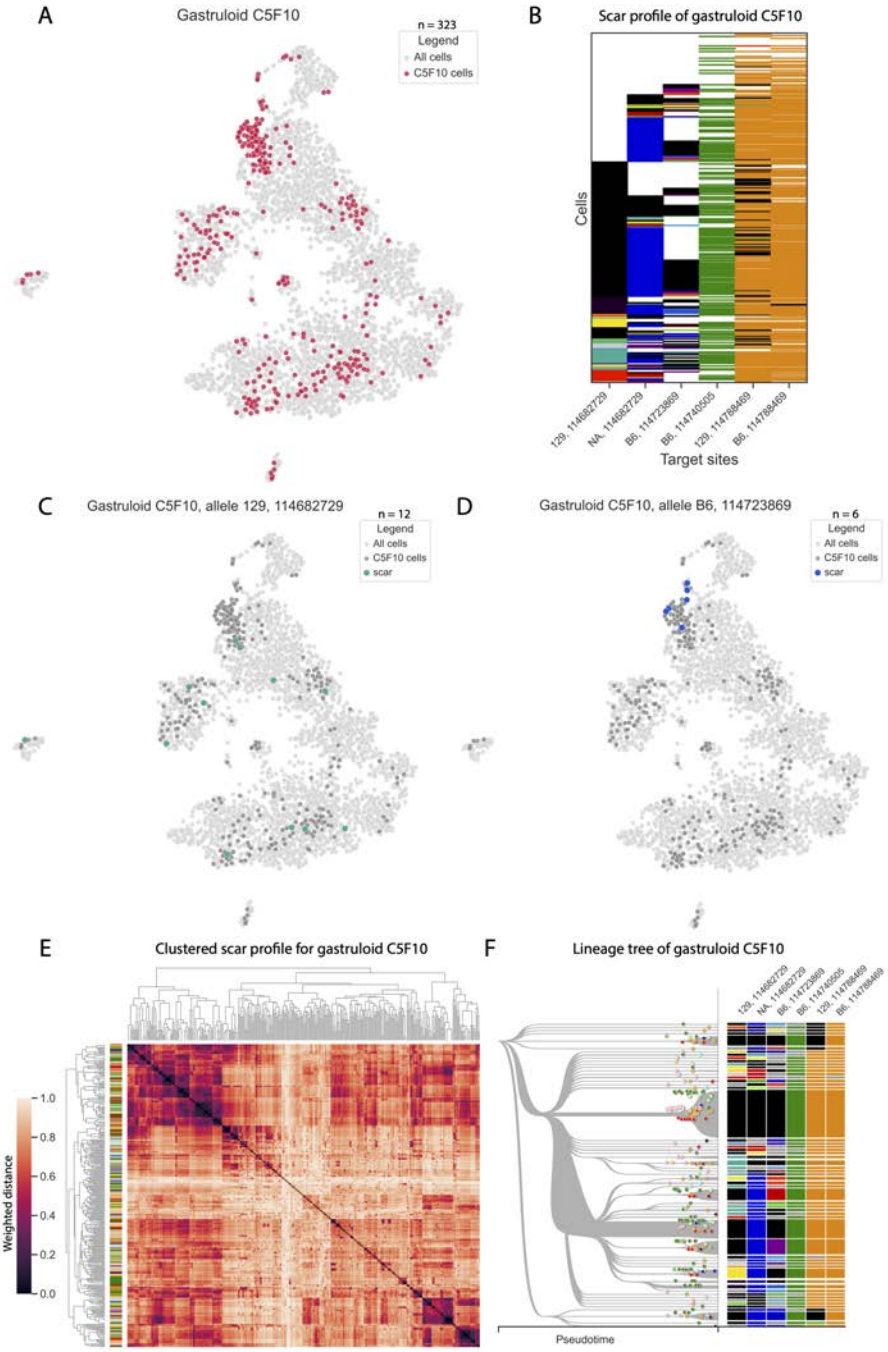

Figure 3 | (Continued on the following page.)

Figure 3 | Lineage tracing results from individual gastruloid C5F10 show an accumulation of scars but no lineage segregation. (**A**) Distribution of cells from gastruloid C5F10 across the transcriptome UMAP, coloured in red. (**B**) Heatmap showing all scars for gastruloid C5F10 in filtered target sites. Target sites with > 80 % missing data are filtered out. (**C**, **D**) Example of an un-localised scar (**C**) and a localised scar (**D**) in gastruloid C5F10. Localisation of the scar in **D** could indicate enrichment but could also be due to a higher density of C5F10 cells in this fraction of the UMAP. All cells are coloured in light grey, all cells from gastruloid C5F10 are coloured in dark grey, cells with a specific scar are coloured uniquely, consistent with heatmaps and trees throughout this chapter. (**E**) Heatmap showing all cells of gastruloid C5F10 clustered based on weighted distance [26]. Here, a distance of 0, indicated in black, indicates a shared identical scar pattern between two cells. (**F**) Lineage tree of gastruloid C5F10. Scar information is collapsed from cells into nodes, where all cells with identical scar patterns are assigned to one node. Nodes are then clustered according to their scar pattern, where similar nodes are close to each other in pseudotime and distant nodes separate early in pseudotime. Scar patterns per node can be seen on the right. Individual cells per node are shown as dots and coloured according to their cell type annotation as seen in Figure 2A. The location of individual cells on the x-axis corresponds to their location in the transcriptome UMAP, where cells similar in transcriptome space are placed close together on the x-axis. Notably, almost all cells/nodes show the same scar in target sites B6 - 114740505, 129 - 114788469, and B6 - 114788469: these sites are pre-scarred. WT is black; missing data is white in heatmaps and grey in lineage trees; each scar is assigned a unique colour, consistent with heatmaps and trees throughout this chapter.

For all gastruloids, there was no statistically significant enrichment of a cell type within a branch with unique scar profile. We hypothesize that this is due to the relatively low number of sampled cells per replicate in relation to the complex mix of cell types formed and the variable ratios of cell types per replicate.

In addition, we observed low complexity of scar profiles, because 3 of the 6 informative scar target sites show only one scar sequence. When we examined the sequence of these target sites in non-induced clone 5 mESCs, we saw that these non-WT sequences were already present and labelled them as pre-existing scars (pre-scars). The presence of pre-scars prevents these sites from accumulating more complex mixtures of additional scars and thereby reduces the overall barcode complexity. We hypothesized that these pre-scars were introduced during integration of the lineage tracing construct in the cell line. Since the Cre-ERT2 cassette is already integrated in the cells before the addition of the lineage tracing construct, its protein is continuously synthesised and present in the cytoplasm. It is therefore possible that multiple lineage tracing constructs enter the cells upon transfection, and part of these get activated in the cytoplasm leading to transient Cas9 expression, while a non-activated Neongreen^{+} copy is integrated.

Since the gRNA is continuously expressed, it can be coupled to Cas9 protein and introduce pre-scars during the integration period of the construct. We expected that the issues mentioned above could be resolved by sampling more cells per replicate and by creating a new clonal line without pre-scars, thereby increasing both the sample size and barcode complexity. We therefore optimised our set-up to increase the number of cells we can collect per gastruloid, and improved the method for generating clonal lines to avoid the introduction of pre-scarred target sites.

Generation and characterisation of clonal mESC lines without pre-existing scars

To generate higher complexity scar profiles, we planned to generate clonal lines without pre-scarred target sites and and included an extra screening to exclude these. We hypothesized that if transient cytoplasmic activation after transfection of the construct caused the introduction of pre-scars, we could FACS-sort out Scarlet-negative cells to counter-select for these cells. In addition, by selecting for

Neongreen-high positive cells, we were expecting to enrich for cells that would be less sensitive to silencing of the construct upon differentiation. We adapted the generation of clonal lines as follows: two days after transfection with Tol2 protein and the lineage tracing construct, we selected Neongreen-positive and Scarlet-negative cells, after which we sorted separately Neongreen-high and Neongreen-low fractions for three consecutive weeks (Figure S4A). We separated the cells into a high- and a low-fluorescent fraction, because we hypothesized that while the high fluorescent cells would show a more stable expression of the construct and would be less prone to silencing events during differentiation, the low fluorescent cells might retain WT scar target sites and a stable karyotype due to lower integration events. After three weeks, a total of 179 clonal lines were generated.

These clonal lines were screened for having an unaltered karyotype, indicative of the absence of unwanted side-effects introduced by CRISPR-Cas9, and WT scar target site sequences. While only a fraction showed karyotype abnormalities, most clonal lines showed variable degrees of pre-scars. Approximately 20 clonal lines remained that passed all thresholds, although not all scar target sites could be recovered and checked for its WT sequence in all lines. To detect if this improved strategy to generate clonal lines would prevent silencing events and allow efficient induction using tamoxifen during development, we opted for a simple differentiation system that can be tested on multiple cell lines simultaneously. To this end, four clonal lines, including example clone H13, were selected for differentiation into embryoid bodies (EBs). These four lines were selected for their clean karyotype and high recovery of scar target sites showing WT sequences. We observed no silencing upon differentiation (Figure S4B).

However, after tamoxifen induction, we observed a switch from Neongreen-positive to fully negative instead of to Scarlet-positive. We selected the next 15 clonal lines and checked by PCR amplification if we could detect the presence of the Neongreen cassette and the Cas9-Scarlet cassette of the lineage tracing construct. While we did detect the Neongreen cassette and observed it was cleaved out upon induction, we could only find the presence of the Cas9-Scarlet cassette in 2 out of 15 clonal lines (Figure S4C). We concluded that most clonal lines with an intact karyotype and WT scar target site sequences did not contain the intact lineage tracing construct. We noticed that only clonal line L18 showed a switch from Neongreen to Scarlet after tamoxifen induction (Figure S4D). When checking the karyotype (Figure 4A) and scar target sites (Figure 4B) of this clonal line, we concluded that this cell line would be our best candidate.

To verify that cell line L18 would efficiently induce scar formation, we performed an induction time-course in mESCs and EBs. Until 9 days after induction of L18 mESCs and EBs, we sampled cells, tracked their fluorescence by FACS and sequenced the scar target sites. In the mESC samples we noticed an efficient switch from Neongreen-positive to Scarlet-positive fluorescence (Figures 4C, S4D), and observed a gradual build-up of scars (Figure 4D-E, S4E). However, in the EBs we noticed simultaneous activation and silencing of the construct during differentiation, where cells initially were able to express Scarlet and generate scars, but over time reverted back to negative and did not continue scar build-up (Figures 4F-H, S4D, F).

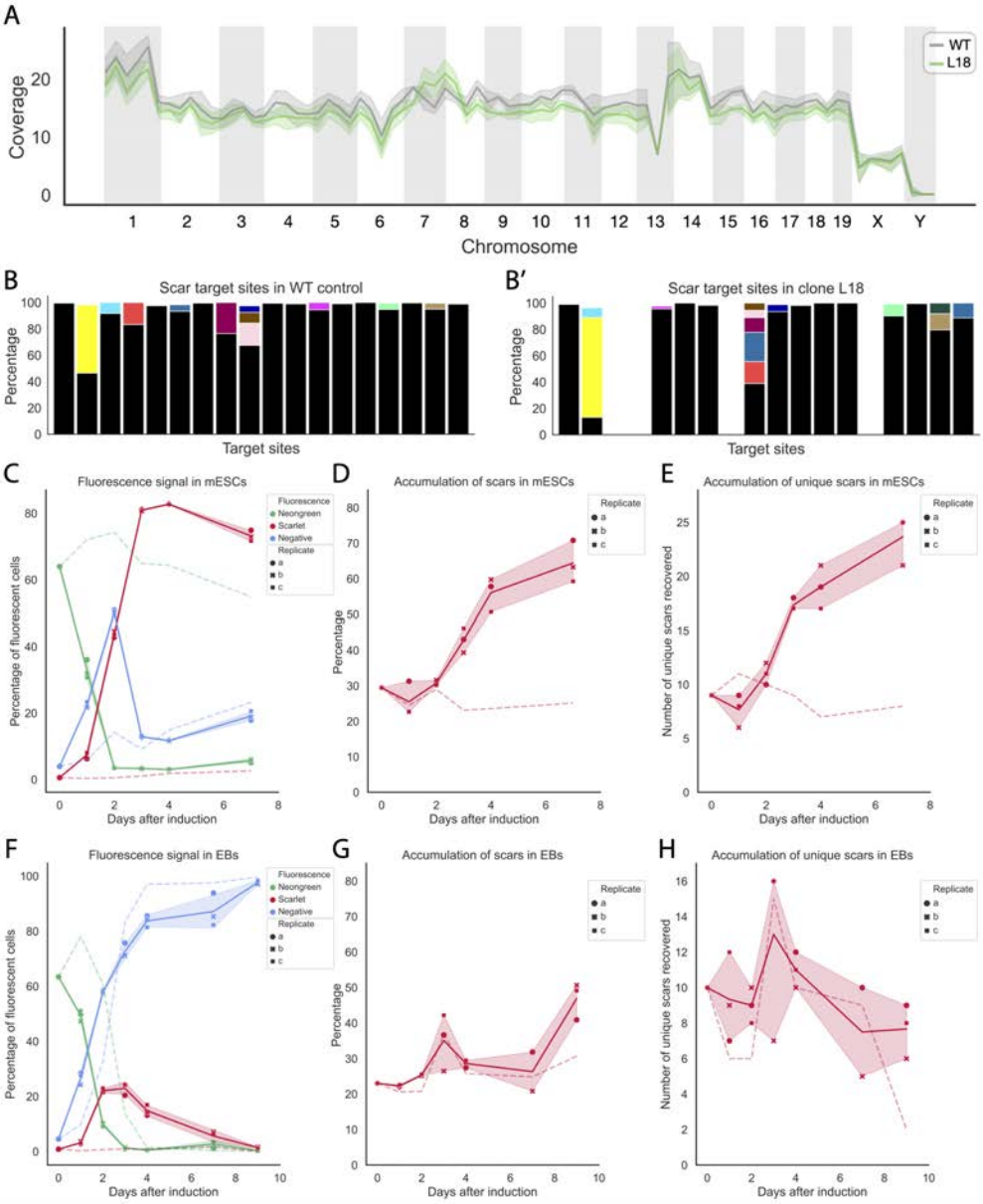

Figure 4 | (Continued on the following page.)

Figure 4 | Selection strategy of clonal lines reveals candidate line L18. (**A**) Karyotype tracks of WT control cells (grey) and clone L18 (green). Karyotypes were checked to exclude the introduction of chromosome aberrations in new clonal lines. After performing genome-wide DNA sequencing, reads per bin and per chromosome are counted. When a chromosome has doubled, it is expected that the total counts for that chromosome is doubled, whereas when a depletion has occurred, the total counts is expected to be halved. It can be seen that the chromosome coverage track of L18 closely follows that of the WT control, therefore it is unlikely that new chromosomal aberrations have been introduced. It can also be seen that this cell line contains an amplification in chromosome 14 and possibly in chromosome 1, and a possible deletion in chromosome 6. In addition, this cell line contains one X chromosome and no Y chromosome. (**B**, **B'**) Screening of gRNA target sites in WT control (**B**) and clonal line L18 (**B'**). Since clone 5 showed the introduction of pre-scars in 3 target sites, new clonal lines were screened for the presence of WT target site sequences. Black is WT, white is missing data, each scar is differently coloured. (**C** - **E**) Induction experiment of clonal line L18 in mESCs. Upon induction with tamoxifen, both the fluorescence profile of the cells (**C**) and scar build-up in the gRNA target sites (**D**, **E**) were monitored over time. Induction was done in triplicate, and one non-induced control was taken along. (**C**) Upon induction, Neongreen fluorescence (green) decreases, cells temporarily fall back to no fluorescence expression (blue), followed by an increase of Scarlet fluorescence (red) over time. Non-induced control (dashed line) remains stably Neongreen-positive. Original FACS plots can be seen in Figure S4D. (**D**) Accumulation of scars and (**E**) accumulation of number of unique scars in mESCs over time. Both the total number of scars versus WT sequences detected and the total number of unique scars detected build up over time from day 3 onwards, whereas the non-induced control remains stable. Bar plot of scar accumulation can be seen in Figure S4E. (**F** - **H**) Induction experiment of clonal line L18 in EBs. Upon induction with tamoxifen, both the fluorescence profile of the cells (**F**) and scar build-up in the gRNA target sites (**G**, **H**) were monitored over time. Induction was done in triplicate, and one non-induced control was taken along. (**F**) Upon induction, Neongreen fluorescence (green) decreases, and cells fall back to no fluorescence expression (blue) over time. As the cells activate the Cas9-Scarlet construct (red), they also differentiate and silence the construct. Non-induced control (dashed line) also falls back from Neongreen-positive to negative at a similar rate. Original FACS plots can be seen in Figure S4D. (**G**) Accumulation of scars and (**F**) accumulation of number of unique scars in EBs over time. No clear build-up of the percentage of scars or the total number of unique scars detected can be seen. There is no clear enrichment of the scarring rate in induced replicates compared to the non-induced control (dotted line). Bar plot of scar accumulation can be seen in Figure S4F.

Generation of gastruloids from new clonal line shows leaky induction of the ScarTrace construct

To determine when silencing occurs, we generated gastruloids from induced and non-induced L18 cells and tracked their fluorescence by imaging. We noticed that the construct remains expressed for the first three days of differentiation, similar to the effect in EBs (Figure 4F-H).

Since silencing of the construct occurs after 3 days of differentiation, we hypothesized that we could still observe scar build-ups over time during the first days of the protocol, and read out its effects at 120 h AA as before. To this end, we generated gastruloids and induced them at different time points: 2 days before aggregation in mESC culture, from 24-72 h AA, and from 48-96 h AA. We sampled in between 2,688 and 3,840 cells per gastruloid. We then selected one gastruloid per induction condition and a non-induced control, and sequenced the first 1000 cells to determine whether this strategy worked (Figures S5A-B). The transcriptome data showed a highly complex mixture of cell types from all lineages, supported by specific expression of marker genes (Figures 5A-B, S5C), similar to our previous findings. With a final number of >850 cells for each replicate after filtering (Figure S5D), we sampled each cell type in each replicate, though variation in the ratios between cell types in different replicates remained (Figure 5C). By changing the sequencing strategy (see Methods), we improved the scar detection to cover 10 target sites for both alleles (Figure S5E), resulting in an average recovery of 13 target sites per cell (Figure S5F). However, the mean number of scars per cell decreased to 1.6 (Figure S5F). Indeed, when looking at the recovered sequences for these target sites, we observed that most of them were still WT (Figure 5D).

To determine whether the few scars recovered would reveal interesting

patterns, we generated lineage trees for all four individual gastruloids (Figures 5E-H). Strikingly, the non-induced control generated the most complex scar pattern and lineage tree (Figure 5E). This indicates that, even without inducing the system, scars are generated over time. Due to this background accumulation of mutations, we conclude that this experiment cannot be used for controlled lineage tracing.

These results indicate a number of issues inherent to the system: constructs like these are prone to be silenced in differentiation systems such as EBs and gastruloids. Furthermore, induction can be leaky, both during introduction of the construct leading to pre-scars, and after the generation of clonal lines leading to induction and scar accumulation in the absence of tamoxifen.

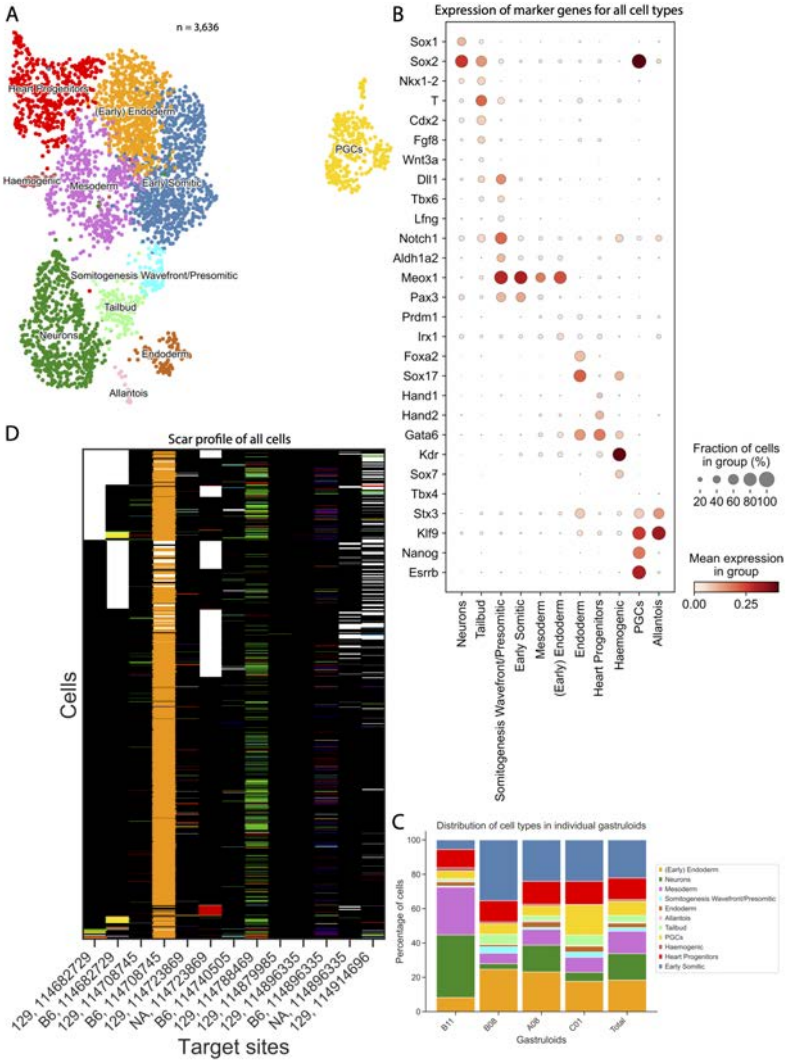

Figure 5 | (Continued on the following page.)

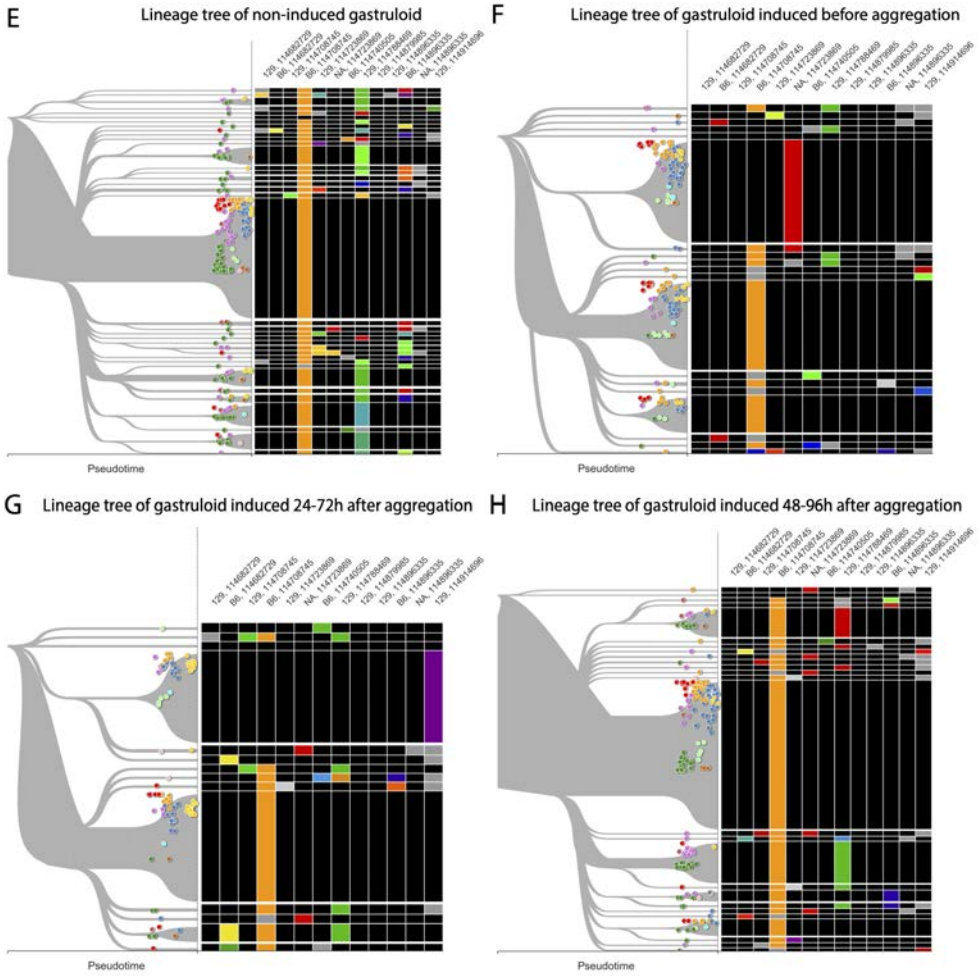

Figure 5 | **ScarTrace in mouse gastruloids generated from clone L18 shows the recovery of distinct cell types, but no lineage information due to scar bleed-through.** (**A**) UMAP showing clustering of single-cell transcriptome data of all cells from all gastruloids generated using clonal line L18. Cell types from all three germ layers are recovered. Cell type annotation is based on localised expression of marker genes (Figures **B**, S5C). (**B**) Dot plot showing expression of marker genes used for cell type annotation. Dot colour represents the mean expression of marker genes in specific groups, and dot size represents the fraction of cells within the group that show expression of the gene. Expression of the genes included in this dot plot can be viewed in detail on the UMAP in Figure S5C. (**C**) Distribution of cell types in individual gastruloids. Cells per cell type in each gastruloid are divided by the total number of cells from that gastruloid, and colours correspond to the UMAP colours in (**A**). (**D**) Heatmap showing all scars for all gastruloids in filtered target sites. Target sites with > 80 % missing data are filtered out. (**E** - **H**) Resulting lineage trees from gastruloids (**E**) B11 - non induced control, (**F**) B08 - induced 2 days before gastruloid aggregation, (**G**) C01 - induced 24-72 h AA, and (**H**) A08 - induced 48-96 h AA. Lineage trees are computed as described in Figure 3F. Most gRNA target sites are completely black, indicating their sequence has remained fully WT, i.e. no scarring has occurred. In addition, the scar pattern of non-induced control B11 seems more complex than that of induced samples. WT is black; missing data is white in heatmaps and grey in lineage trees; each scar is assigned a unique colour, consistent with heatmaps and trees throughout this chapter.

Alternatives to the inducible lineage tracing construct

While we concluded that this version of the inducible ScarTrace design is currently not working reliably, CRISPR-Cas9-based methods have shown to be a useful tool for lineage tracing[11, 12, 13, 14, 15, 16, 17]. One of the possible reasons why our system did not yield the expected results could be because of the construct design itself. To remove this variable, we tested whether a transfection of the IgH-targeting sgRNA with Cas9 protein would result in scar formation in differentiating cells. We transfected WT mESCs for 24 h two days and one day before gastruloid formation, and sampled mESCs at the time-point of aggregation (day 2) and 5 days after (day 7) and gastruloids at 120 h AA (day 7; Figure 6A). While transfection control samples showed mostly WT target sites, we saw an accumulation of total percentage of scarred sequences and number of target sites containing scars, both in the gastruloid and mESC samples (Figures 6B-C, S6). We conclude that the strategy of targeting an endogenous locus could be used for lineage tracing during differentiation, however a different design is needed to make this strategy inducible without encountering the issues we described above. In addition, considering the ratio of WT sequences retrieved, the system should be optimised for more efficient scar build-up.

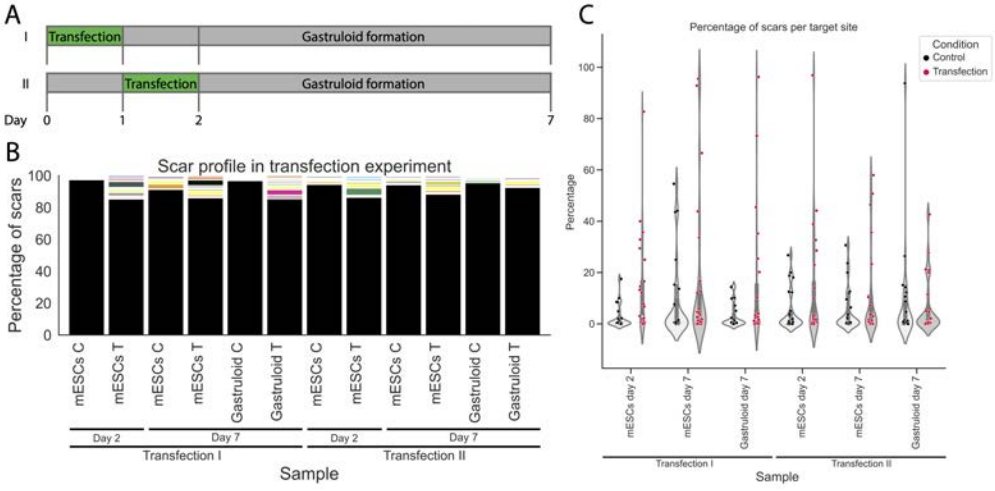

Figure 6 | Transfection of WT mESCs with gRNA and Cas9 protein introduces scars in gRNA target sites in both mESCs and gastruloids. (**A**) Schematic overview of the experiment. Transfection I was performed from 48 until 24 hours before gastruloid aggregation. Transfection II was performed from 24 hours until 0 hours before gastruloid aggregation. For both transfection strategies, bulk control and transfected mESC samples were taken at day 2 (time point of aggregation) and day 7 (end point gastruloid protocol). For both transfection strategies, bulk control and transfected gastruloid samples were taken at day 7. (**B**) Resulting DNA sequence read-out of gRNA target sites. All target sites are shown together. In control samples, the gRNA target sites remain mostly WT. In transfected samples, a small build-up of scars is seen and persists in both mESCs and gastruloids during the full experiment. Transfection I results in a larger fraction of scars than transfection II. WT is black; scars are uniquely coloured; C is control; T is transfected. (**C**) Resulting DNA sequence read-out of individual gRNA target sites. The percentage of scars recovered for each individual gRNA target site is shown per sample. The majority of gRNA target sites in control samples do not accumulate scar sequences and remain fully WT, whereas target sites in transfected conditions show a bigger spread and higher scar content. Control samples are shown in black; transfected samples are shown in red.

Discussion

CRISPR-Cas9 combined with single-cell sequencing is one of the most rapidly emerging tools for lineage tracing in complex organisms. Here, we show a novel and inducible design targeting 13 endogenous loci with single molecule resolution due to site- and allele-specific variations. This results in a build-up of mutations, or scars, in a maximum of 26 unique sequences. We applied this design to mESCs, as well as two differentiation systems - EBs and gastruloids. Upon induction of the system in mESCs, we see a gradual build-up of scars. In differentiating cells however, this build-up is less clear due to simultaneous activation and silencing of the system. Moreover, we observed that the system in its current form can already be transiently expressed during the integration of all necessary constructs in clonal lines. This causes target sites to already be mutated, or pre-scarred, in the cell line before activation, thereby reducing final barcode complexity. Other issues that were encountered during the integration of the lineage tracing construct into a cell line included karyotype abnormalities, spontaneous differentiation, excision of (part of) the construct, and loss of the ability to induce complex scar profiles. Therefore, we conclude that the system in its current form is unsuitable for its purpose of lineage tracing in a dynamic and differentiating developmental system.

However, after transfection with Cas9 protein and the IgH-targeting gRNA, we did observe a build-up of scars while control samples remained mostly unscarred. This indicates that the main issues that we encountered are caused by the design of the lineage tracing construct combined with activation through Cre-ERT2, and not the CRISPR-Cas9 genome editing technique itself. It should also be noted that spontaneous mutations, mismatches introduced during PCR amplification, sequencing errors, and mismapping cannot be excluded as additional complications. Here, we propose several changes to the existing design that should resolve the majority of the difficulties encountered.

One of the major challenges when studying embryonic development is the silencing of transgenes during differentiation. Many transgenes are integrated into the genome using lentiviral or transposon-based methods, and integrate into the genome in seemingly random positions. However, different integration methods show distinct preferences in terms of integration location[27]. Tol2-based systems preferentially integrate into weakly expressed or epigenetically silenced regions of the genome, whereas PiggyBac-based constructs typically integrate in highly expressed regions[27]. An alternative to this would be the integration of the current ScarTrace construct in a known stable and open locus, such as the MYH9 locus[28].

To avoid unintentional activation of the lineage tracing construct itself, multiple measures can be taken. Because the ScarTrace construct itself is already Cas9-based and the associated gRNA is constantly expressed, any Cas9-based integration method could result in unwanted side-effects such as the formation of pre-scars. When choosing to insert the lineage tracing construct in a stable locus, its insertion should therefore not be Cas9-mediated but rather Cas12a- or TALENs-mediated, for example. In addition, integrating the lineage tracing construct first followed by the Cre-ERT2 cassette should avoid the possibility of cytoplasmic Cre recombination and thereby activation of non-integrated copies of the construct.

When applying this design *in vivo*, one could cross-breed Cre-ERT2 mice with lineage-tracing mice. One could also consider to switch from the tamoxifen-inducible Cre-LoxP system to the doxycycline-inducible tetracycline (Tet) system, and test whether this activation method would be less prone to background scar accumulation before construct activation.

Another aspect to take into consideration is the biological system itself. In this study gastruloids were used. Similar to embryos, this *in vitro* system is rapidly developing, complex and dynamic. In a model with considerable biological variability, it is crucial to include numerous replicates and to achieve a high sampling rate for each replicate. In the current set-up there is considerable sampling variability of cell types in individual replicates, where not all cell types were recovered for lowly sampled replicates (Figures 2C, S2E). In order to identify common ancestors for certain lineages or take the effects of lineage compensation into account, it is crucial to treat each gastruloid as a replicate.

In later experiments, the number of recovered cells per replicate was increased (Figure S5D; only ~30 % of cells per replicate were processed) and while variable cell type ratios are still recovered, the effect is less extreme (Figure 5C). This indicates that the observed variability can in part be explained by a low sampling rate in earlier experiments. The average cell number in gastruloids at 120 h AA was determined to be ~50.000 for non-induced and ~28.000 for induced replicates (results not shown). In later experiments, on average 3.264 cells were sampled, which is about 8.5 % of the expected in total. Since the gastruloids in this study developed from aggregates of 600 cells each, one can expect each cell to grow out to a cluster of ~40 - 80 cells over 120 hours. With the current sampling rate, we would recover ~5 of these. It is therefore expected that the achieved sampling rate is high enough for a study like this, but it should be emphasized that it is crucial to include a reasonable number of replicates. Especially in a system as dynamic and variable as mammalian embryonic development, the same pattern needs to be observed in multiple independent instances before one can draw appropriate conclusions.

With this in mind, we propose the following improvements to the system. We would opt for stable integration of the lineage tracing construct in an open locus. In addition, we propose to switch from Cre-LoxP to the dox-inducible Tet system, to avoid leaky expression. We consider mESCs an ideal cell line for setting up this system, because of its varied *in vitro* usages, as well as the possibility to expand to *in vivo* models through generation of chimera mouse embryos. A hybrid genetic background such as 129/B6, or Cast/Bl6 would be ideal to retain single-molecule scar resolution.

When setting up this system, we propose the following quality controls that were established previously (see Figures 4, S4): (1) shortly after integration of the construct, check for the presence of Scarlet$^+$ cells, (2) select Neongreen$^+$ cells and grow clonal lines to ensure an identical genetic background, (3) verify the correct integration of the lineage tracing construct, (4) assess karyotype to avoid the selection of a cell line with additional abnormalities, (5) ensure all scar target sites have remained wild-type, (6) perform a differentiation assay to assess if the system is silenced, (7) perform an induction time-course to assess whether the system is reliably induced and builds up mutations over time. Once these improvements have been implemented and the

system has been verified to function reliably, we envision this could provide a useful new approach to perform lineage tracing in a dynamic differentiation model.

Methods

Generation of inducible ScarTrace construct

The lineage tracing construct was a kind gift from the Geijsen lab (LUMC, The Netherlands). The ScarTrace gRNA sequence (GGAAAGAGCCTTGAGTGGATTGG) was cloned as follows: the gRNA sequence was ordered as forward and reverse oligo with sticky overhangs (forward: 5′ ACCG– 3′, reverse: 5′ AAAC– 3′). 10 µM forward and 10 µM reverse primer were annealed by heating up to 98 ℃ for 2 mins, followed by a ramp down from 65 ℃ until 45 ℃ in 5 ℃ steps of 5 minutes each. The lineage tracing construct was digested with BsmBI at 55 ℃ followed by purification with the PCR purification kit (Qiagen). Double-stranded gRNA with BsmBI overhangs was annealed into the linearised construct with T4 ligase (NEB) in 10x buffer at 16 ℃ for two hours, followed by heat inactivation at 65 ℃ for 10 mins. The construct was transformed in DH5α E. coli by incubation on ice for 20 mins followed by a 40 second 42 ℃ heat shock, plating on an agar plate with ampicillin and incubating overnight at 37 ℃. Colonies were picked, grown in LB with ampicillin and the construct was isolated using the PureLink miniprep kit (Invitrogen).

Cell culture

129S1/SvImJ / C57BL/6 Cre-ERT2 mESCs[25] were cultured in a humidified incubator (5 % CO2, 37 ℃) on 0.1 % gelatin-coated cell culture dishes in ESLIF medium (GMEM (Gibco) containing 10 % fetal bovine serum (FBS), 1 % Sodium Pyruvate, 1 % non- essential amino acids, 1 % GlutaMAX, 1 % penicillin-streptomycin, 0.2 % β-mercaptoethanol, and 1000 units/mL mouse leukemia inhibitory factor (mLIF, ESGRO)). Cells were split every second day by washing with PBS0, dissociating with TrypLE for 5 mins, centrifuging at 300 g for 3 mins, and replating the pellet 1:5 on gelatinised plates. Clonal line L18 was grown on mouse embryonic fibroblasts (MEFs) in ESLIF or in 2i medium (48 % DMEM/F12 and 48 % Neurobasal supplemented with 0.5 % N-2 supplement, 1 % B-27 supplement, 1 % GlutaMAX, 1.1 % penicillin-streptomycin, 0.2 % β-mercaptoethanol, 1000 units/mL mouse leukemia inhibitory factor (mLIF, ESGRO), 3 µM Chiron (CHIR99021, Sigma) and 1 µM PD0325901 (Sigma)).

Transient transfection of mESCs was done by adding 0.5 µg IgH-targeting sgRNA (IDT) and 2.5 µg Cas9 protein in Lipofectamine CRISPRMAX (Invitrogen) and opti-MEM to ~10×10^4 cells. Transfection controls contained all components except the Cas9 protein.

Generation of clonal lines

1 µg ScarTrace construct and 1 µg Tol2 protein were co-transfected into 129S1/SvImJ / C57BL/6 Cre-ERT2 mESCs[25] with Lipofectamine2000 (Invitrogen) in opti-MEM

(Gibco). Every seven days for three consecutive weeks, Neongreen$^+$ cells were enriched by FACS sorting and replating. The third week, single Neongreen$^+$ cells were sorted into individual wells of a 96-well flat bottom plate pre-coated with 0.1 % gelatin. After a week, cells that have grown out into colonies are dissociated for 5 mins at 37 ℃ with TrypLE, re-plated on a new 96-well plate and subsequently expanded. This yielded clonal lines 3, 5, 7, and 9 (Figures 1-3). For the second generation of clonal lines, an extra FACS enrichment step for Neongreen$^+$ and Scarlet$^-$ cells was included 2 days after transfection, to reduce the pool of cells with cytoplasmic activation of the construct possibly resulting in the introduction of pre-scars. This yielded clonal line L18 (Figures 4-5) and all clonal lines in Figure 4.

Karyotyping

10.000 cells per clonal line were lysed with proteinase K (Ambion) and NP40 for 6 h at 55 ℃ followed by 20 mins at 80 ℃. DNA was digested with NLAIII (NEB) for 2 h at 37 ℃ followed by 20 mins at 65 ℃, and adapters were ligated with T4 ligase (NEB, 400.000 U/mL) overnight at 16 ℃. Samples were cleaned with 0.8x AMPure XP beads (Beckman Coulter), pooled, and *in vitro* transcribed (IVT; Invitrogen) for 13 h at 37 ℃. DNA contamination was removed with Turbo DNase (Invitrogen) for 15 mins at 37 ℃, RNA was fragmented for 1.5 mins at 94 ℃ and the reaction was stopped by adding 0.5 M EDTA on ice. aRNA was cleaned with AMPure RNAclean XP beads (Beckman Coulter), followed by reverse transcription (RT) with RNaseOUT (Invitrogen), SuperScriptII (Invitrogen) and random hexamers for 10 mins at 25 ℃, followed by 1 h at 42 ℃ and 15 mins at 70 ℃. Library PCR was done using NEBNext High-Fidelity 2x PCR MasterMix (NEB) and Illumina TruSeq adapters with an annealing temperature of 60 ℃ and 8 cycles. Samples were cleaned twice with 0.8x AMPure beads and sequenced on the Illumina NextSeq500 platform at 2 x 150 bp.

Differentiation assays

Gastruloids were generated as described before[20,22] with the following adaptations: N2B27 medium (Takara) was passed through a .22 µm filter before use and 600 cells were used for aggregation instead of 300. Before gastruloid aggregation, clone 5 mESCs were cultured in ESLIF medium, and L18 mESCs were cultured in 2i medium for 2 days followed by ESLIF medium for two days. **Embryoid bodies** (EBs) were made by dissociating mESCs as described before and seeding 2×10^6 cells on low-attachment dishes in 15 mL ESLIF without mLIF. After 24 h, EBs were collected and centrifuged at 300 g for 5 mins and the pellet was re-plated on 10 low-attachment dishes in 15 mL ESLIF without mLIF. Medium was refreshed every other day.

Tamoxifen induction and DNA extraction

Clone 5 mESCs and gastruloids were induced with 0.1 µM tamoxifen (Merck) for 48 h, clone L18 mESCs, EBs and gastruloids were induced with 0.5 µM tamoxifen. DNA extractions for bulk scar profiling were done using the DNeasy Blood & Tissue Kit

(Qiagen). To check the presence of the Neongreen and Cas9-Scarlet cassette (Figure S4C), a PCR was done on extracted DNA with 30 cycles and an annealing temperature of 62 °C. Samples were loaded on a 1 % agarose gel in TAE with EtBr.

Dissociation and FACS sorting of mESCs, single gastruloids and single EBs

Single gastruloids were washed with PBS0, followed by incubation with 50 µL trypsin without phenol red (Gibco) for ~5 mins at 37 °C. Gastruloids were mechanically broken up into a single-cell suspension by pipetting with a P200 pipette, diluted to 1 mL with PBS0 supplemented with 5 % FBS, and filtered through a 35 µm filter (Falcon, 352235). Prior to fluorescence-activated cell sorting (FACS), DAPI was added to the cell suspension to assess cell viability. Single EBs were transferred to separate wells of a 96-well plate, after which dissociation was done as described above. mESCs were dissociated as described before, the cell pellet was re-suspended in PBS0 with 5 % FBS, filtered through a 35 µm filter (Falcon, 352235), and DAPI was added to the cell suspension to assess cell viability.

Detection of single-cell scars and transcriptome

The ScarTrace protocol was followed as previously described[11], with the following alteration of scar amplification by PCR: instead of amplifying GFP, unique barcoded primers with the following sequence were added to amplify the IgH locus:
forward primer
`5′-TTCCTTGGCACCCGAGAATTCCANNNNNNNNTCTTTCTCTTCACAGGTG-3′`,
reverse primer
`5′-GTTCAGAGTTCTACAGTCCGACGATCCCTTGAACTTCTGGTTGTAG-3′`
with NNNNNNNN as cell barcode position. For the PCR, 15 cycles were done with an annealing temperature of 62 °C.

Sequencing

Scars were paired-end sequenced on the Illumina Miseq with 2 x 300 bp reads (Figures 1-3) or Illumina NextSeq 500 with 2 x 150 bp reads (Figures 4-6). Transcriptome was sequenced on the Illumina NextSeq 500 with 2 x 75 bp reads or 1 x 75 bp reads (26 bp at read 1 and 61 bp at read 2).

Transcriptome analysis

Raw fastq files were processed into count tables using the transcriptome Snakemake[29] pipeline of the SingleCellMultiOmics package (see `https://github.com/BuysDB/SingleCellMultiOmics/wiki/Transcriptome-data-processing`, version 0.1.30). The pipeline works as follows: reads were demultiplexed for CelSeq2 barcodes with a hamming distance of 0 using SingleCellMultiOmics' `demux.py` and trimmed with `cutadapt`[30] (version 4.1), then mapped with STAR[31] (version 2.5.3a) to the 129/B6 SNP-masked GRCm38 mouse genome (Ensembl 97). Feature counting was performed with `featureCounts`[32]

(version 1.6.2) and reads were tagged and deduplicated, then transformed to a .csv file. Downstream analysis was performed with Scanpy[33] (version 1.8.1). Cells with less than 1000 reads and 100 genes detected and genes detected in less than 2 cells were filtered out. Cells with more than 15% mitochondrial reads were also excluded. Mitochondrial reads, Malat1, and ribosomal reads were excluded from further analysis. Counts were normalized to 10.000 transcripts per cell and logarithmized. The number of counts and percentage of mitochondrial reads were regressed out and each gene was scaled to unit variance with a maximum of 10. Principal component analysis was performed and the 40 highest principal components were used to generate the UMAP. The data was clustered using the Leiden algorithm (scanpy.tl.leiden, resolution set to 1).

Scar analysis

Raw fastq files were processed into count tables using the ScarTrace Snakemake[29] pipeline of the SingleCellMultiOmics package (see `https://github.com/BuysDB/SingleCellMultiOmics/wiki/scartrace-data-processing`, version 0.1.30). Reads were demultiplexed for scartrace barcodes with a hamming distance of 0 using SingleCellMultiOmics' `demux.py` and trimmed with `cutadapt`[30] (version 4.1), then mapped with BWA mem[32] (version 0.7.16a) to the 129/B6 SNP-masked GRCm38 mouse genome (Ensembl 97) with the following parameters: `-M -I 220`. Read tagging and scar and allele annotation were performed with `bamtagmultiome.py` with the following parameters: `-method scartrace -every_fragment_as_molecule -allele_samples "C57BL_6NJ,129S1_SvImJ"` using a custom allele vcf file. The custom vcf file was generated using SingleCellMultiOmics tool `heterozygousSNPedit.py` and a .bam file with WT sample data from either clone 5 or clone L18. This script finds all heterozygous 129S1/SvImJ SNPs in the vcf file, checks which of the two alleles is present in the clonal line and overwrites this to a homozygous SNP, to improve allele calling. After tagging, reads are filtered to have a mean phred score of > 98 % using `bamFilter.py` with the following parameter: `'r.has_tag("SQ") and r.get_tag("SQ") > 0.98'` and reads are filtered to have a maximum insert size of < 1000 using bamFilter.py with the following parameter: `'r.has_tag("fS") and (r.get_tag("fS") < 1000)'`. Filtered bam files are transformed into count tables with chromosome, allele, site and scar counts for each cell using `bamToCountTable.py` with the following parameters: `-joinedFeatureTags chrom,DA,DS,SD -sampleTags SM`. For each site+allele with a minimum of 3 counts, raw counts are transformed to percentages. Since each site+allele should be derived from a single molecule, scars with a minimum percentage of > 55 % for each site+allele are transformed into allele counts. Each unique scar is assigned a unique colour for plotting on heatmaps. Lineage trees are generated using the ToverBoom package (see `https://github.com/BuysDB/ToverBoom`, version 1.0.0). In short, all allele+sites with > 80 % missing data are removed before all cells with identical scar patterns are collapsed into nodes. Nodes with <2 cells and >1 site with missing data are removed. Distances across pseudotime are calculated as follows: it is possible to go from WT to scar and from WT or scar to missing data (distance +1),

and from WT to WT or from scar to the identical scar (same distance), but not from scar to WT or from missing data to WT or scar. After this, a graph is made and the most efficient path through all nodes is calculated.

Code and data availability

All data generated for this project, code used and annotated notebooks for performing the analysis and generating figures can be found on the mouse-scartrace github (`https://github.com/marloes3105/mouse-scartrace`) and SingleCellMultiOmics github (`https://github.com/BuysDB/SingleCellMultiOmics`). Unprocessed data is available upon request.

Acknowledgements

The authors thank Max Wellenstein for critical reading of the manuscript and all AvO lab members for discussions and input. We thank Alfonso Martinez Arias, Naomi Moris and Susanne van den Brink for helpful discussions and input, in particular on cell typing and interpreting of the biological systems. We thank the Hubrecht Sorting Facility (HSP) and Utrecht Sequencing Facility (USEQ) for cell sorting and sequencing. The 129/Bl6 Cre-ERT2 cell line was a gift from Matyas Flemr and Marc Bühler and the ScarTrace inducible construct was a gift from Peng Shang and Niels Geijsen (LUMC).

Author contributions

MB, PZ and AA contributed to initial experimental design. MB and BAB performed data analysis, MB designed and performed experiments, generated cell lines and wrote the manuscript. RvdL and MB performed cell sorting for single-cell experiments and clonal expansion of cell lines. PZ and AvO supervised the study.

References

1. J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental Biology*, 100(1):64–119, 1983. ISSN 00121606. doi: 10.1016/0012-1606(83)90201-4.

2. Kirstie A. Lawson, Juanito J. Meneses, and Roger A. Pedersen. Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer. *Developmental Biology*, 115(2):325–339, 1986. ISSN 00121606. doi: 10.1016/0012-1606(86)90253-8.

3. K. A. Lawson, R. A. Pedersen, and S. Van De Geer. Cell fate, morphogenetic movement and population kinetics of embryonic endoderm at the time of germ layer formation in the mouse. *Development*, 101 (3):627–652, 1987. ISSN 09591991. doi: 10.1242/dev.101.3.627.

4. K. A. Lawson, J. J. Meneses, and R. A. Pedersen. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development*, 113(3):891–911, 1991. ISSN 09501991. doi: 10.1242/dev.113.3.891.

5. Kirstie A. Lawson. Fate mapping the mouse embryo. *International Journal of Developmental Biology*, 43(7):773–775, 1999. ISSN 02146282. doi: 10.1387/ijdb.10668985.

6. Jean Livet, Tamily A. Weissman, Hyuno Kang, Ryan W. Draft, Ju Lu, Robyn A. Bennis, Joshua R. Sanes, and Jeff W. Lichtman. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450(7166):56–62, 2007. ISSN 14764687. doi: 10.1038/nature06293.

7. Y. Albert Pan, Tom Freundlich, Tamily A. Weissman, David Schoppik, X. Cindy Wang, Steve Zimmerman, Brian Ciruna, Joshua R. Sanes, Jeff W. Lichtman, and Alexander F. Schier. Zebrabow: multispectral cell labeling for cell tracing and lineage analysis in zebrafish. *Development*, 140(13): 2835–2846, jul 2013. ISSN 1477-9129. doi: 10.1242/dev.094631. URL https://journals.biologists.com/dev/article/140/13/2835/76622/Zebrabow-multispectral-cell-labeling-for-cell.

8. Katie McDole, Léo Guignard, Fernando Amat, Andrew Berger, Grégoire Malandain, Loïc A Royer, Srinivas C Turaga, Kristin Branson, and Philipp J Keller. In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell*, 175(3):859–876.e33, oct 2018. ISSN 1097-4172. doi: 10.1016/j.cell.2018.09.031. URL http://www.ncbi.nlm.nih.gov/pubmed/30318151.

9. Justus M. Kebschull and Anthony M. Zador. Cellular barcoding: lineage tracing, screening and beyond. *Nature Methods*, 15(11):871–879, 2018. ISSN 15487105. doi: 10.1038/s41592-018-0185-x. URL http://dx.doi.org/10.1038/s41592-018-0185-x.

10. Brent A. Biddy, Wenjun Kong, Kenji Kamimoto, Chuner Guo, Sarah E. Waye, Tao Sun, and Samantha A. Morris. Single-cell mapping of lineage and identity in direct reprogramming. *Nature*, 564(7735): 219–224, 2018. ISSN 14764687. doi: 10.1038/s41586-018-0744-4. URL http://dx.doi.org/10.1038/s41586-018-0744-4.

11. Anna Alemany, Maria Florescu, Chloé S. Baron, Josi Peterson-Maduro, and Alexander Van Oudenaarden. Whole-organism clone tracing using single-cell sequencing. *Nature*, 556(7699): 108–112, 2018. ISSN 14764687. doi: 10.1038/nature25969. URL http://dx.doi.org/10.1038/nature25969.

12. Kirsten L. Frieda, James M. Linton, Sahand Hormoz, Joonhyuk Choi, Ke Huan K. Chow, Zakary S. Singer, Mark W. Budde, Michael B. Elowitz, and Long Cai. Synthetic recording and in situ readout of lineage information in single cells. *Nature*, 541(7635):107–111, 2017. ISSN 14764687. doi: 10.1038/nature20777.

13. A. McKenna, G. M. Findlay, J. A. Gagnon, M. S. Horwitz, A. F. Schier, and J. Shendure. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, 353 (6298):aaf7907–aaf7907, jul 2016. ISSN 0036-8075. doi: 10.1126/science.aaf7907. URL http://www.ncbi.nlm.nih.gov/pubmed/27229144http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4967023http://www.sciencemag.org/cgi/doi/10.1126/science.aaf7907.

14. Bushra Raj, Daniel E. Wagner, Aaron McKenna, Shristi Pandey, Allon M. Klein, Jay Shendure, James A. Gagnon, and Alexander F. Schier. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nature Biotechnology*, 36(5):442–450, 2018. ISSN 15461696. doi: 10.1038/nbt.4103.

15. Bastiaan Spanjaard, Bo Hu, Nina Mitic, Pedro Olivares-Chauvet, Sharan Janjuha, Nikolay Ninov, and Jan Philipp Junker. Simultaneous lineage tracing and cell-type identification using CrIsPr-Cas9-induced genetic scars. *Nature Biotechnology*, 36(5):469–473, 2018. ISSN 15461696. doi: 10.1038/nbt.4124.

16. Michelle M. Chan, Zachary D. Smith, Stefanie Grosswendt, Helene Kretzmer, Thomas M. Norman, Britt Adamson, Marco Jost, Jeffrey J. Quinn, Dian Yang, Matthew G. Jones, Alex Khodaverdian, Nir Yosef, Alexander Meissner, and Jonathan S. Weissman. Molecular recording of mammalian embryogenesis. *Nature*, 570(7759):77–82, 2019. ISSN 14764687. doi: 10.1038/s41586-019-1184-5. URL http://dx.doi.org/10.1038/s41586-019-1184-5.

17. Reza Kalhor, Kian Kalhor, Leo Mejia, Kathleen Leeper, Amanda Graveline, Prashant Mali, and George M. Church. Developmental barcoding of whole mouse via homing CRISPR. *Science*, 361 (6405), 2018. ISSN 10959203. doi: 10.1126/science.aat9804.

18. Peter Baillie-Johnson, Susanne Carina van den Brink, Tina Balayo, David Andrew Turner, and Alfonso Martinez Arias. Generation of Aggregates of Mouse Embryonic Stem Cells that Show Symmetry Breaking, Polarization and Emergent Collective Behaviour In Vitro. *Journal of Visualized Experiments*, e53252(105):1–10, nov 2015. ISSN 1940-087X. doi: 10.3791/53252. URL http://www.jove.com/video/53252/generation-aggregates-mouse-embryonic-stem-cells-that-show-symmetryhttps://www.jove.com/t/53252/generation-aggregates-mouse-embryonic-stem-cells-that-show-symmetry.

19. Leonardo Beccari, Naomi Moris, Mehmet Girgin, David A. Turner, Peter Baillie-Johnson, Anne Catherine Cossy, Matthias P. Lutolf, Denis Duboule, and Alfonso Martinez Arias. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature*, 562(7726): 272–276, 2018. ISSN 14764687. doi: 10.1038/s41586-018-0578-0. URL http://dx.doi.org/10.1038/s41586-018-0578-0.

20. Susanne C. Van Den Brink, Peter Baillie-Johnson, Tina Balayo, Anna Katerina Hadjantonakis, Sonja Nowotschin, David A. Turner, and Alfonso Martinez Arias. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development (Cambridge)*, 141 (22):4231–4242, 2014. ISSN 14779129. doi: 10.1242/dev.113001.

21. Giuliana Rossi, Nicolas Broguiere, Matthew Miyamoto, Andrea Boni, Romain Guiet, Mehmet Girgin, Robert G. Kelly, Chulan Kwon, and Matthias P. Lutolf. Capturing Cardiogenesis in Gastruloids. *Cell Stem Cell*, 28(2):230–240.e6, 2021. ISSN 18759777. doi: 10.1016/j.stem.2020.10.013. URL https://doi.org/10.1016/j.stem.2020.10.013.

22. Susanne C. van den Brink, Anna Alemany, Vincent van Batenburg, Naomi Moris, Marloes Blotenburg, Judith Vivié, Peter Baillie-Johnson, Jennifer Nichols, Katharina F. Sonnen, Alfonso Martinez Arias, and Alexander van Oudenaarden. Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature*, 582(7812):405–409, jun 2020. ISSN 0028-0836. doi: 10.1038/s41586-020-2024-3. URL http://www.nature.com/articles/s41586-020-2024-3.

23. Jesse V. Veenvliet, Adriano Bolondi, Helene Kretzmer, Leah Haut, Manuela Scholze-Wittler, Dennis Schifferl, Frederic Koch, Léo Guignard, Abhishek Sampath Kumar, Milena Pustet, Simon Heimann, René Buschow, Lars Wittler, Bernd Timmermann, Alexander Meissner, and Bernhard G. Herrmann. Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science*, 370(6522), 2020. ISSN 10959203. doi: 10.1126/science.aba4937.

24. Mehmet U. Girgin, Nicolas Broguiere, Lorenzo Mattolini, and Matthias P. Lutolf. Gastruloids generated without exogenous Wnt activation develop anterior neural tissues. *Stem Cell Reports*, 16(5): 1143–1155, 2021. ISSN 22136711. doi: 10.1016/j.stemcr.2021.03.017. URL https://doi.org/10.1016/j.stemcr.2021.03.017.

25. Matyas Flemr and Marc Bühler. Single-step generation of conditional knockout mouse embryonic stem cells. *Cell Reports*, 12(4):709–716, 2015. ISSN 22111247. doi: 10.1016/j.celrep.2015.06.051.

26. Andrea Attardi, Timothy Fulton, Maria Florescu, Gopi Shah, Leila Muresan, Martin O. Lenz, Courtney Lancaster, Jan Huisken, Alexander van Oudenaarden, and Benjamin Steventon. Neuromesodermal progenitors are a conserved source of spinal cord with divergent growth dynamics. *Development (Cambridge)*, 146(2), 2019. ISSN 14779129. doi: 10.1242/dev.175620.

27. Junko Yoshida, Keiko Akagi, Ryo Misawa, Chikara Kokubu, Junji Takeda, and Kyoji Horie. Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. *Scientific Reports*, 7(March):1–18, 2017. ISSN 20452322. doi: 10.1038/srep43613.

28. Tanbin Liu, Yi Hu, Shiyin Guo, Lei Tan, Yang Zhan, Lingchen Yang, Wei Liu, Naidong Wang, Yalan Li, Yingfan Zhang, Chengyu Liu, Yi Yang, Robert S. Adelstein, and Aibing Wang. Identification and characterization of MYH9 locus for high efficient gene knock-in and stable expression in mouse embryonic stem cells. *PLoS ONE*, 13(2):1–17, 2018. ISSN 19326203. doi: 10.1371/journal.pone.0192641.

29. Johannes Köster and Sven Rahmann. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*, 28(19):2520–2522, 2012. ISSN 14602059. doi: 10.1093/bioinformatics/bts480.

30. Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, may 2011. ISSN 2226-6089. doi: 10.14806/ej.17.1.200. URL http://journal.embnet.org/index.php/embnetjournal/article/view/200.

31. Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013. ISSN 13674803. doi: 10.1093/bioinformatics/bts635.

32. Yang Liao, Gordon K. Smyth, and Wei Shi. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btt656.

33. F. Alexander Wolf, Philipp Angerer, and Fabian J. Theis. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, 19(1):15, dec 2018. ISSN 1474-760X. doi: 10.1186/s13059-017-1382-0. URL https://genomebiology.biomedcentral.com/articles/10.1186/s13059-017-1382-0.

Supplemental figures

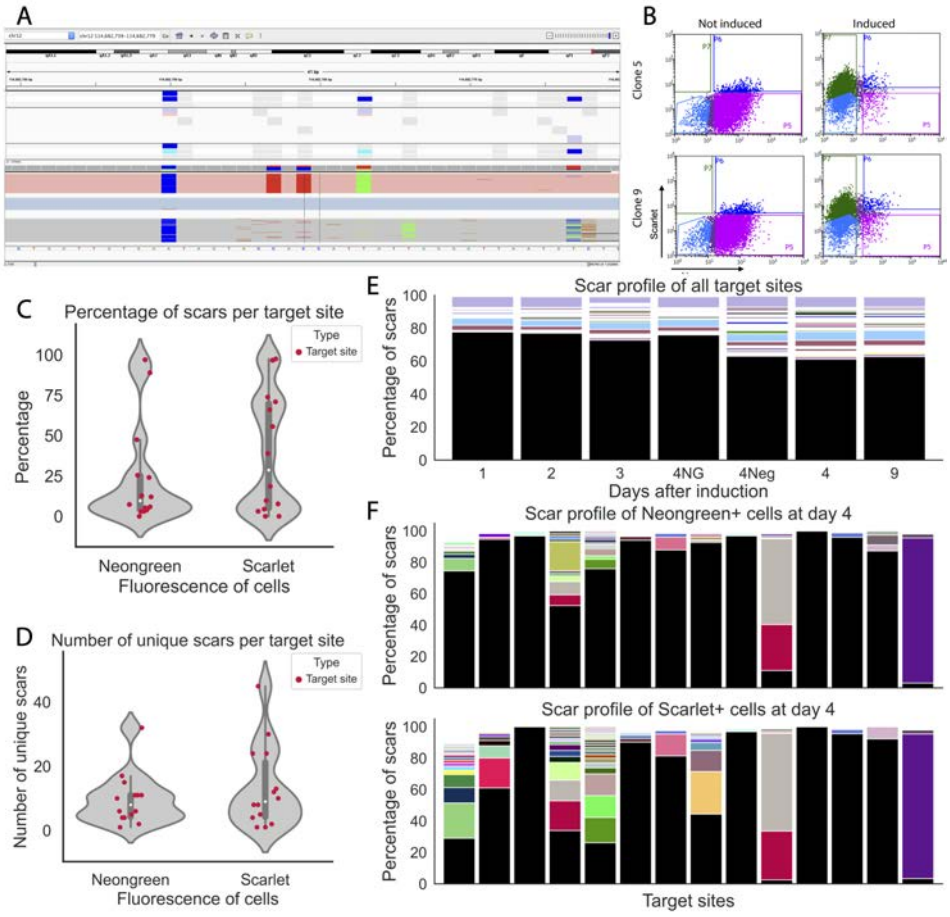

Figure S1 | Inducible ScarTrace set-up in clonal mESCs demonstrates fluorescent colour switch and build-up of scars. (**A**) Example target site showing allele calling. Top track: location of allele-specific SNPs, pink track: 129 allele, blue track: Black6 allele. The 129 allele has four distinguishing SNPs in this location to separate its reads from the Black6 allele. (**B**) Tamoxifen induction of clone 5 (top) and clone 9 (bottom) mESCs. Non-induced cells are mainly Neongreen-positive, whereas induced cells are mainly Scarlet-positive. (**C**, **D**) Percentage of scars(**C**) and number of unique scars (**D**) per target site for Neongreen$^+$ and Scarlet$^+$ fluorescent cells 4 days after induction. Both the percentage of scars and number of unique scars recovered per target site is generally lower in Neongreen$^+$ cells than in Scarlet$^+$ cells, and Scarlet$^+$ cells show a wider distribution of scarring percentages and number of unique scars recovered per target site. This indicates that a higher percentage of scars and more unique sequences are recovered in Scarlet$^+$ cells. (**E**) Scar profile of all target sites together across the time-course corresponding to Figures 1C-D. The percentage of scars increases most notably between day 2 and 4. For day 4, cells from three different gates were sorted (see **S1A**): The Neongreen$^+$ (NG), Negative (Neg), and Scarlet$^+$ gate. While Neongreen$^+$ cells show a lower percentage of accumulated scars, there appears to be no clear difference between fluorescent negative and Scarlet$^+$ cells. Samples for all time points have been sorted for Scarlet$^+$ fluorescence, unless otherwise stated. WT is black; each scar is a unique colour. (**F**) DNA sequencing results corresponding to Figures 1C-D, where each bar indicates a recovered target site. Top: scar profile of Neongreen-positive cells four days after tamoxifen induction, bottom: scar profile of Scarlet-positive cells four days after tamoxifen induction. Scarlet-positive cells show a higher degree of scarring and more target sites with scars than Neongreen-positive cells. WT is black; each scar is a unique colour.

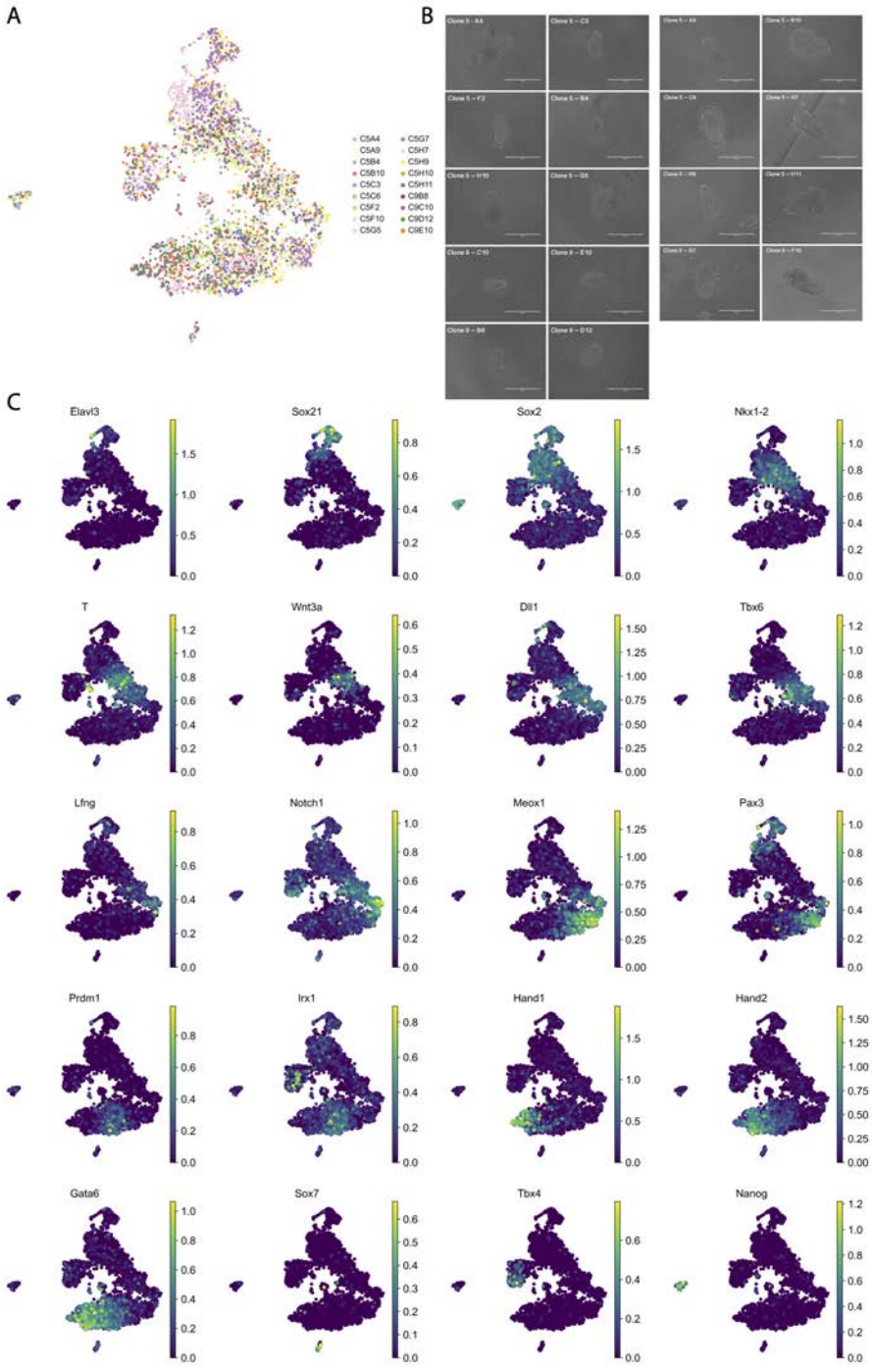

Figure S2 | (Continued on the following page.)

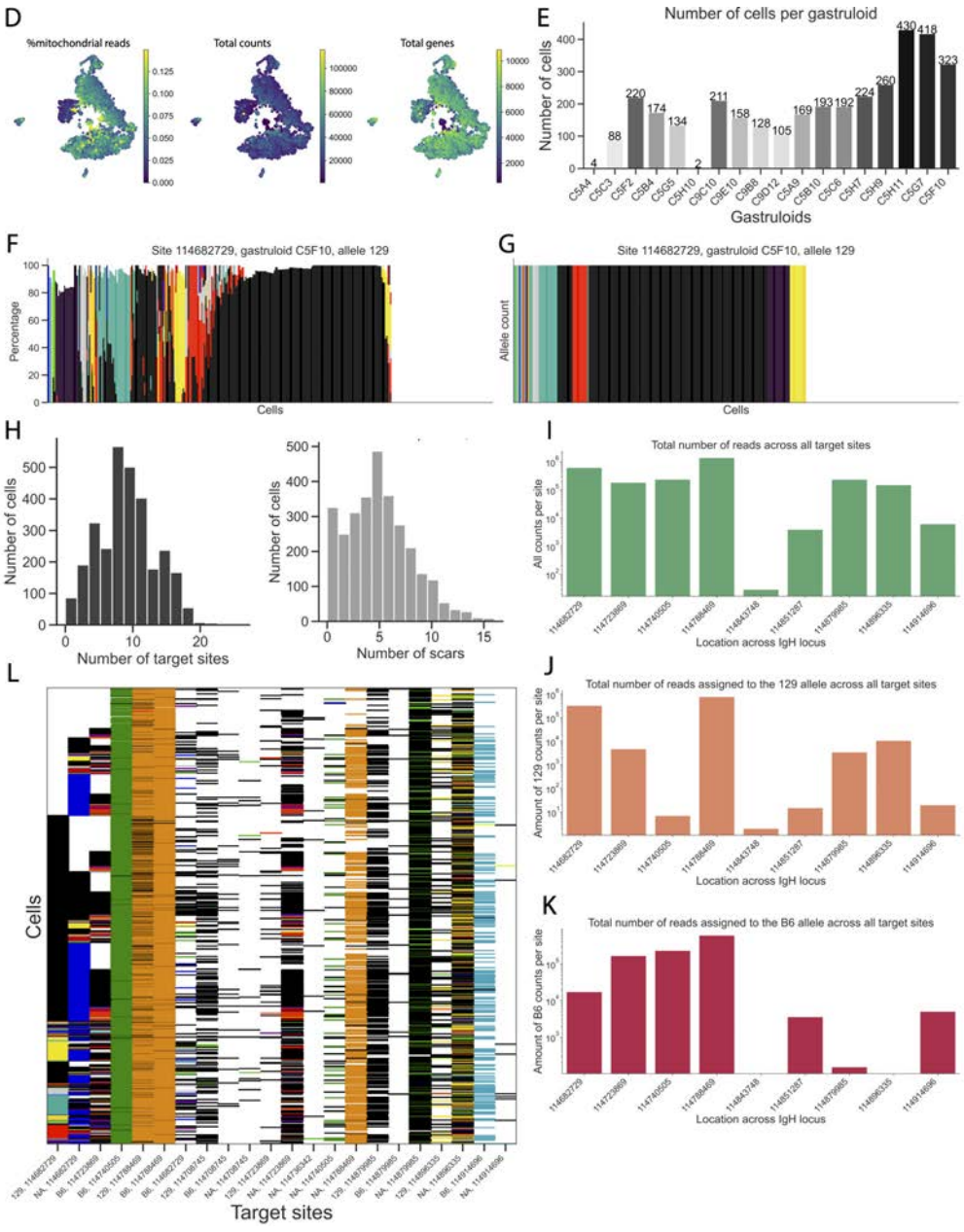

Figure S2 | (Continued on the following page.)

Figure S2 | ScarTrace in mouse gastruloids generated from clone 5 shows the recovery of distinct cell types and an accumulation of scars. (**A**) Transcriptome UMAP showing distribution of cells from each individual gastruloid. (**B)** Bright-field microscopy showing morphology of gastruloids that were selected for transcriptome and scar sequencing. (**C**) Expression of marker genes selected in dot plot (Figure 2B) plotted on the UMAP. *Elavl3* and *Sox21*, Neural Tube Progenitors; *Sox2* and *Nkx1-2*, Neurons; *T* and *Wnt3a*, Tailbud; *Dll1* and *Tbx6*, Presomitic Mesoderm; *Lfng* and *Notch1*, Somitogenesis Wavefront; *Meox1* and *Pax3*, Early Somitic; *Prdm1* and *Irx1*, (Early) Endoderm; *Hand1*, *Hand2*, and *Gata6*, Heart Progenitors; *Sox7*, Haemogenic; *Tbx4*, Allantois; *Nanog*, PGCs. (**D**) Distribution of mitochondrial gene expression (left), total count distribution (middle), and total gene count distribution (right) per cell. (**E**) Total number of cells per gastruloid. Only gastruloids with > 150 cells were selected for follow-up analyses. (**F** - **G**) Scar pattern per cell for one example site, allele, and gastruloid. (**F**) Percentage of scars per cell. Percentages < 5 % are not plotted. In principle, one site and one allele should only contain one molecule with 100 % of the same sequence. However, many cells also show bleed-through of signal. Only if a sequence occurs for > 55 % within one cell, site and allele, is it converted to an allele count (**G**). This removes noise and bleed-through. (**H**) Total number of recovered gRNA target sites per cell (left) and number of recovered gRNA target sites excluding WT sequences per cell (right). (**I** - **K**) Number of reads across gRNA target sites, in total (**I**), for the 129 allele only (**J**), and for the Black6 allele only (**K**). (**L**) Heatmap of Figure 2D before filtering of cells and target sites. Generally, target sites that are filtered out only show coverage in a small percentage of cells. WT is black; missing data is white; each scar is assigned a unique colour, consistent with heatmaps and trees throughout this chapter.

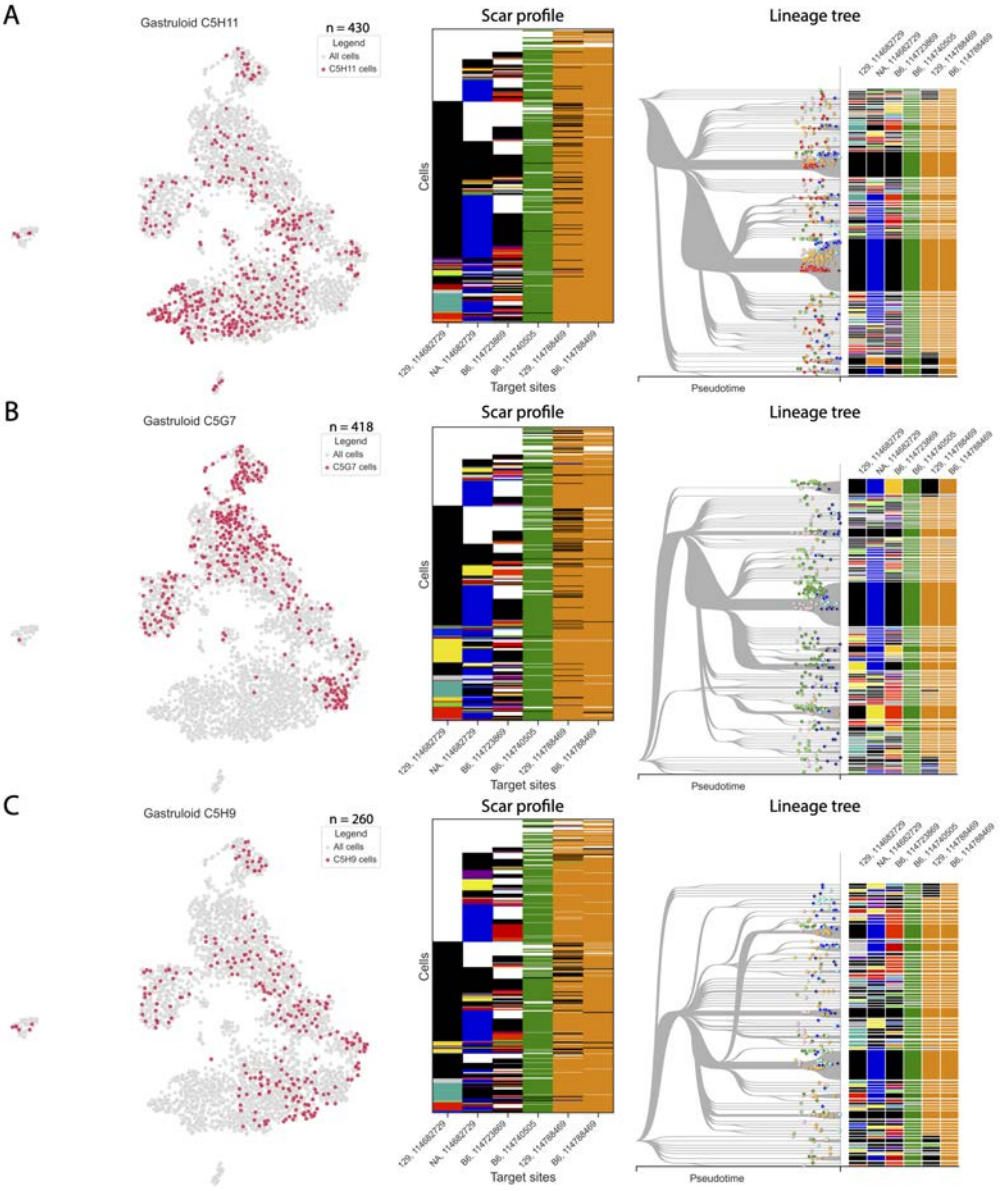

Figure S3 | (Continued on the following page.)

Figure S3 | Lineage trees of the three highest sampled gastruloids does not resolve clear lineage decisions. Distribution of cells from individual gastruloid on transcriptome UMAP (left), heatmaps (middle), and lineage trees (right) of gastruloids C5H11 (**A**), C5G7 (**B**), and C5H9 (**C**). Heatmaps are showing all scars for individual gastruloids in filtered target sites. Target sites with > 80 % missing data are filtered out. Lineage trees are based on scar information, which is collapsed from cells into nodes, where all cells with identical scar patterns are assigned to one node. Nodes are then clustered according to their scar pattern, where similar nodes are close to each other in pseudotime and distant nodes separate early in pseudotime. Scar patterns per node can be seen on the right: WT is black; missing data is grey; each individual scar is uniquely coloured, consistent with heatmaps and trees throughout this chapter. Individual cells per node are shown as dots and coloured according to their cell type annotation as seen in Figure 2B. The location of individual cells on the x-axis corresponds to their location in the transcriptome UMAP, where cells similar in transcriptome space are placed close together on the x-axis. Notably, almost all cells/nodes show the same scar in target sites B6 - 114740505, 129 - 114788469, and B6 - 114788469: these sites are most likely pre-scarred. Fisher test reveals no significant enrichment of cell types for specific branches of the trees. WT is black; missing data is white; each scar is assigned a unique colour, consistent with heatmaps and trees throughout this chapter.

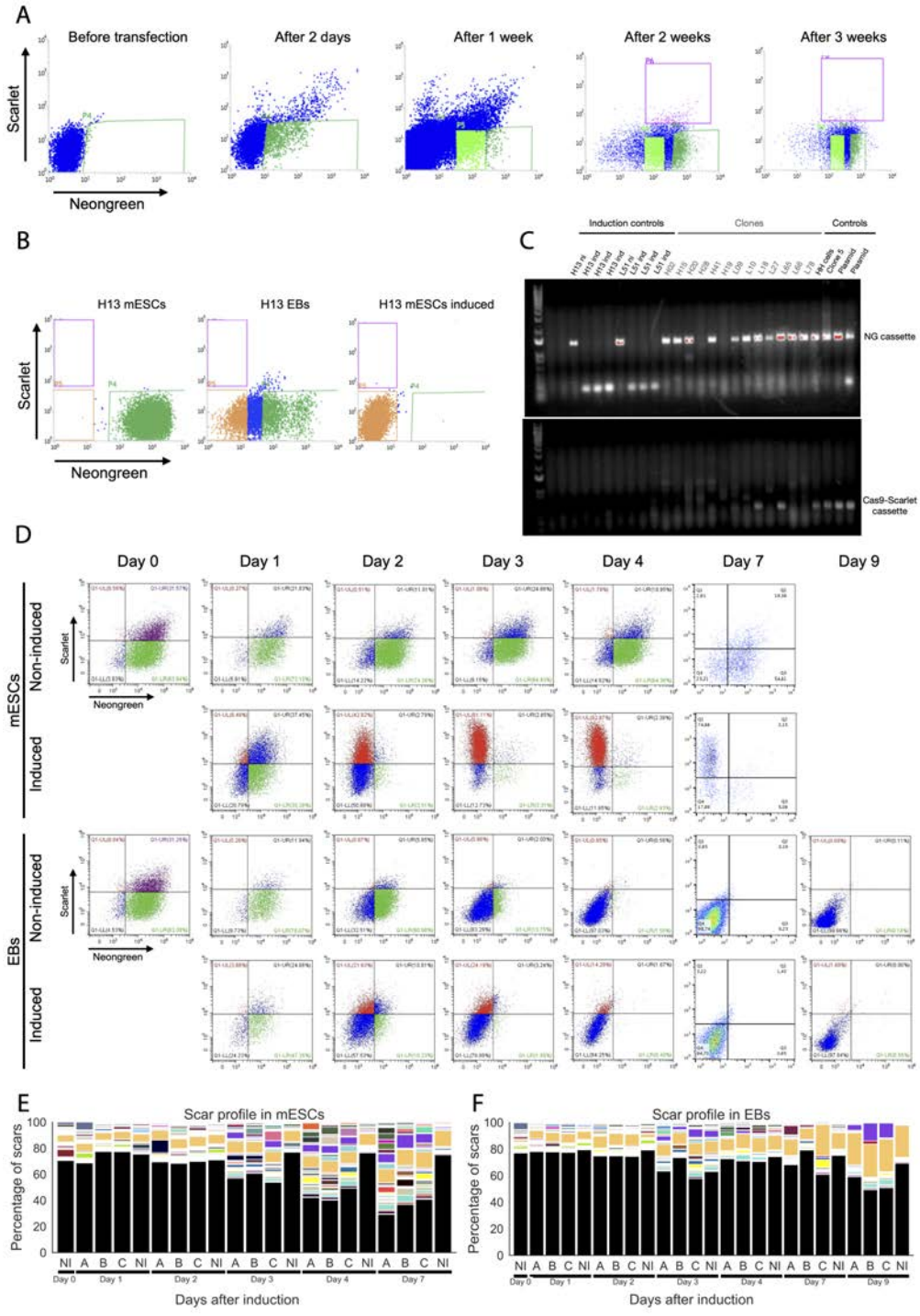

Figure S4 | (Continued on the following page.)

Figure S4 | Selection strategy of new clonal lines reveals candidate line L18. (**A**) FACS sorting schedule to generate new clonal lines. Cells are transfected with the lineage tracing construct (Figure 1A).Two days after transfection, cells that are Neongreen-positive and Scarlet-negative are selected and grown out, to select for cells which contain the construct (NG^+) and counter-select for cells with active Cas9 ($Scarlet^+$), possibly indicative of the introduction of pre-scars. After one, two and three weeks, Neongreen-positive cells are selected and grown out, to ensure the construct is not only present but also integrated in the cells. The last two sorts select a Neongreen low and high fraction, in light and dark green respectively. During the last sort, single cells are sorted into individual wells of a 96-well plate and grown out, resulting in the generation of clonal lines. Clonal lines are derived from the high-high (H) and high-low (L) fractions. (**B**) Example clone H13, non-induced mESCs (left), differentiated into EBs (middle), and induced mESCs with tamoxifen (right). Clonal line H13 does not silence the construct upon differentiation, but retains its fluorescence at least partially. However, upon activation, there is a switch from $Neongreen^+$ to negative, instead of to $Scarlet^+$, indicating that the Cas9-Scarlet cassette might be impaired. (**C**) PCR control showing induction controls of clonal lines H13 and L51, a selection of clonal lines, and control samples consisting of high-high mixed cells, clone 5, and two plasmid-only controls. A band in the top gel is indicative of the presence of the Neongreen cassette. Upon activation of H13 and L51, the Neongreen cassette is cleaved out. Most clonal lines also contain this cassette. A band in the bottom gel is indicative of the presence of the Cas9-Scarlet cassette. Whereas the control does show a band, most clonal lines do not, indicating that these do not contain the Cas9-Scarlet cassette. Only clones L18 and L65 remain as possible candidates with a functional lineage tracing cassette. NG is Neongreen; ni is non-induced; ind is induced. (**D**) FACS plots of induction time-course in clonal line L18 used to generate line plots in Figures 4C-D. Top two rows: mESC samples, non-induced control and one induced replicate, resp.; bottom two rows: EB samples, non-induced control and one induced replicate, resp. Whereas mESCs show a decrease of Neongreen and increase of Scarlet fluorescence over time, EBs show a decrease of Neongreen but fall back to negative over time. The mESC non-induced control remains stable, whereas the EB non-induced control falls back to negative too. This indicates that whereas mESCs reliably switch on the lineage tracing construct upon activation, differentiation causes silencing of the construct after 4 days. (**E** - **F**) Scar profiles of all target sites during induction time-course in clonal line L18 in mESCs (**E**) and EBs (**F**). Induction was done in triplicate (A-C), and one non-induced (NI) control was taken along. In mESCs, scars in induced replicates build up over time from day 3 onwards, whereas non-induced control remains stable. In EBs, no clear build-up of scars can be seen, the scarring rate in induced replicates is similar to the non-induced control. Light orange fraction is possibly a pre-scar in one of the gRNA target sites. This data was also used to generate line plots in Figures 4D-E and 4G-H. A, B, C, replicates; NI, non-induced control; WT is black, each scar is a unique colour.

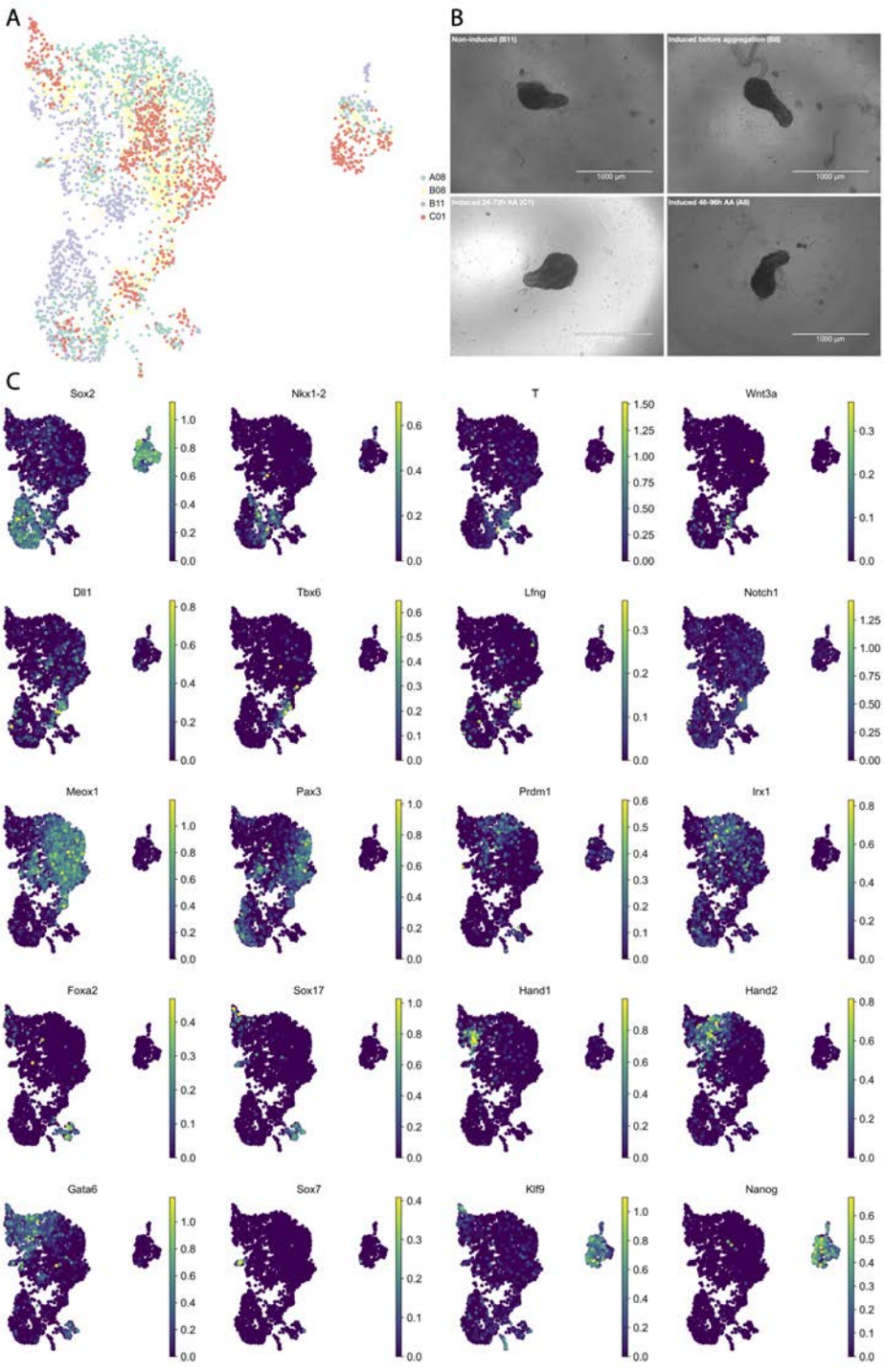

Figure S5 | (Continued on the following page.)

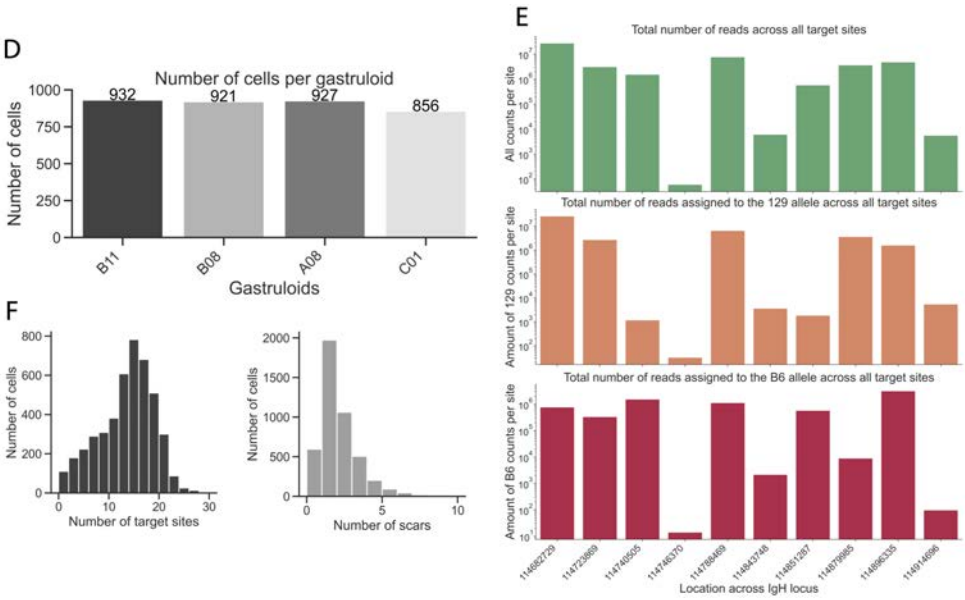

Figure S5 | ScarTrace in mouse gastruloids generated from clone L18 shows the recovery of distinct cell types and scar information. (**A**) Transcriptome UMAP showing distribution of cells from each individual gastruloid. (**B**) Bright-field microscopy showing morphology of the four gastruloids that were selected for transcriptome and scar sequencing. (**C**) Expression of marker genes shown in dot plot (Figure 5B) plotted on the UMAP. *Sox2* and *Nkx1-2*, Neurons; *T* and *Wnt3a*, Tailbud; *Dll1* and *Tbx6*, Presomitic Mesoderm; *Lfng* and *Notch1*, Somitogenesis Wavefront; *Meox1* and *Pax3*, Early Somitic; *Prdm1* and *Irx1*, (Early) Endoderm; *Foxa2* and *Sox17*, Endoderm; *Hand1*, *Hand2*, and *Gata6*, Heart Progenitors; *Sox7*, Haemogenic; *Klf9*, Allantois; *Nanog*, PGCs. (**D**) Total number of cells per gastruloid. (**E**) Total number of recovered gRNA target sites per cell (left) and number of recovered gRNA target sites excluding WT sequences per cell (right). (**F**) Number of reads across gRNA target sites, in total (top), for the 129 allele only (middle), and for the Black6 allele only (bottom).

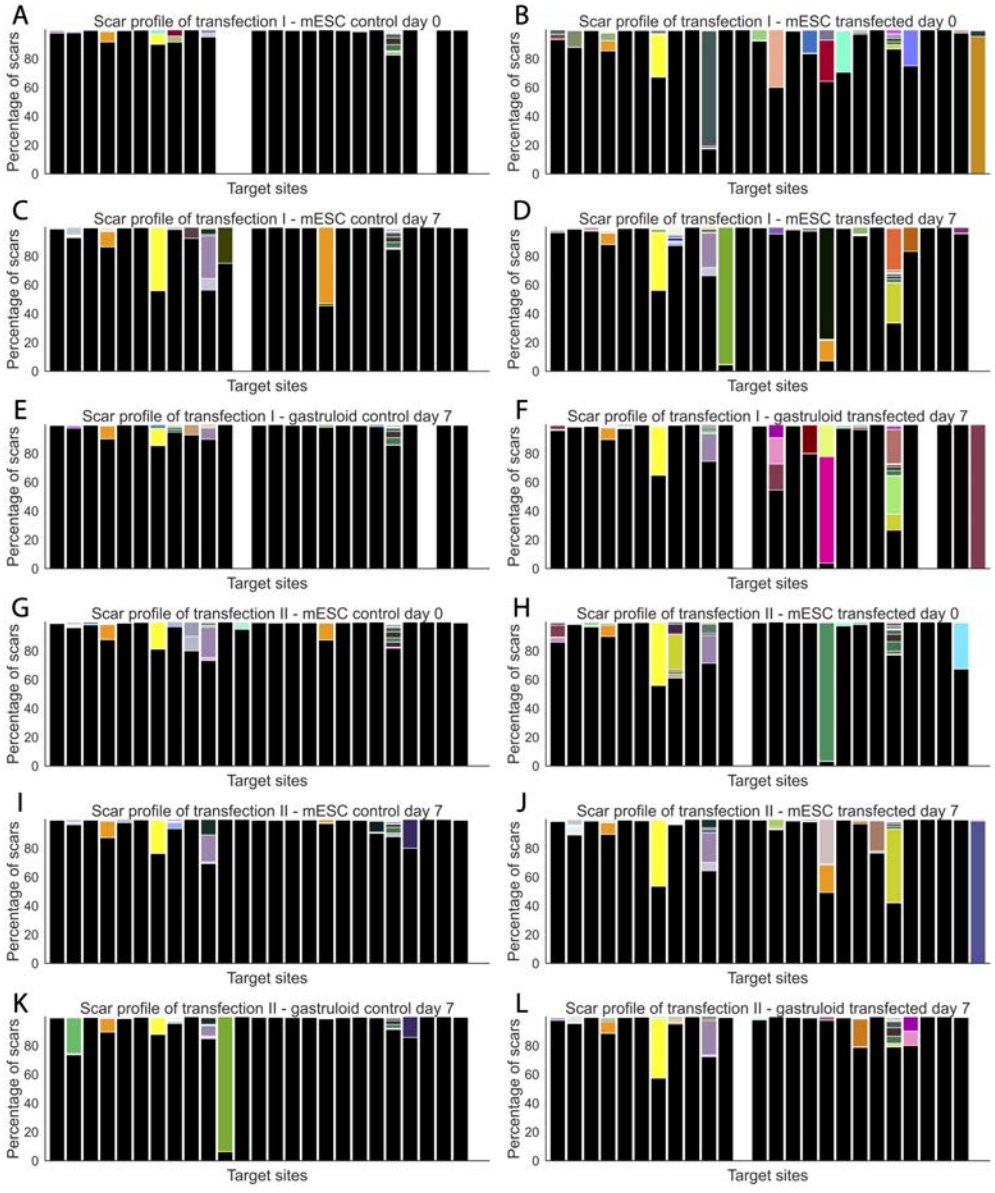

Figure S6 | Transfection of WT cell line with gRNA and Cas9 protein introduces scars in gRNA target sites. Shown here are all individual target sites for all samples included: (**A**, **B**) Transfection I, mESC control and transfected samples at day 0, (**C**, **D**) Transfection I, mESC control and transfected samples at day 7, (**E**, **F**) Transfection I, gastruloid control and transfected samples at day 7, (**G**, **H**) Transfection II, mESC control and transfected samples at day 0, (**I**, **J**) Transfection II, mESC control and transfected samples at day 7, (**K**, **L**) Transfection II, gastruloid control and transfected samples at day 7.