

**SUPPLEMENTARY MATERIAL****mESC Culture**

Rx-GFP K/I EB5 mESCs (Wataya et al., 2008) obtained from the Riken BRC Cell Bank (cell number AES0145) were maintained under feeder-free conditions. Specifically, cells were cultured on 0.1% gelatin (Sigma, G2500) coated tissue culture flasks (Corning) in GMEM media (Gibco, 21710025) supplemented with 1X non-essential amino acids (Sigma, M7145), 1 mM Sodium pyruvate (Sigma, S8636), 10  $\mu$ M 2-Mercaptoethanol (Gibco, 21985-023), 10% knock out serum replacement (KOSR) (Gibco, 10828-028), 1% Foetal Calf Serum (FCS) (Sigma, 12103C) and 500 U/ml leukaemia inhibitory factor (LIF) (produced in house). Care was taken to avoid over-confluency or undesirable differentiation of cells.

**Optic Vesicle Organoid Culture**

For the culture of OV organoids using a modified SFEBq (serum-free floating culture of embryoid body-like aggregates with quick reaggregation) technique, mouse ES cells were dissociated to single cells in TrypLE and reaggregated in a differentiation media at a concentration of 4500 cells per 100  $\mu$ l per well of a Nunclon Sphera 96-well U-bottomed low cell adhesion plate (Thermo Fisher Scientific, 174929). Two differentiation medias were used for this work due to the differentiation in CDM media becoming unstable with cell lines generated from CRISPR genome editing. The organoids used for RNA and ATAC-seq were grown in CDM media and the organoids with mutations introduced in potential CREs were grown in KSR media. Both differentiation medias produced organoids that are very similar in terms of size, structure and GFP expression (organoids grown in KSR are shown in Fig.S1A that are comparable to the CDM organoids in Fig.1A). KSR differentiation media consisted of GMEM supplemented with 0.1 mM non-essential amino acids, 1 mM Sodium pyruvate, 10  $\mu$ M 2-Mercaptoethanol and 1.5% KOSR. CDM differentiation media was made up of Iscove's Modified Dulbecco's Medium (IMDM) GlutaMAX<sup>TM</sup> Supplement (Thermo, 31980022) and Ham's F12 Nutrient Mix, GlutaMAX<sup>TM</sup> Supplement (Thermo, 31765027) mixed in a one-to-one ratio and supplemented with 1X Chemically Defined Lipid Concentrate (Thermo, 11905031), 5 mg/ml Bovine Serum Albumin (BSA) (Sigma, A3156-5G), 15 mg/ml bovine Apo-transferrin (Sigma, T1428) and 450  $\mu$ M 1-Thioglycerol (Sigma, M6145). The day on which cells are aggregated and differentiation started was defined as day0. For day0

samples, cells were seeded in the low cell adhesion plates in stem cell maintenance media and collected after 24 hours (Fig.S1B). These cells were grown up to day8 to ensure there was no differentiation or GFP expression (Fig.S1C). On day1, growth factor reduced Matrigel basement membrane matrix (Corning, 354230) was added to a final concentration of 2% (v/v). Cells were then incubated at 37°C with 5% CO<sub>2</sub> and the aggregated cells differentiated to form optic vesicle like structures expressing Rx-GFP by day5. It has been previously noted that addition of the Wnt agonist CHIR99021 at day4 is required when growing these organoids in CDM media, to promote differentiation toward a retinal cell fate. We did not see any effect on the organoids upon increasing Wnt signalling. Rather organoids grown without the addition of CHIR99021 exhibited the 3D structure and GFP expression patterns typical of the retinal organoids previously published (Sakakura et al., 2016).

**RNA Sequencing**

**RNA Extraction, Quantification and Sequencing.** 24 organoids were pooled, washed with PBS 3 times or until all Matrigel and media was removed, and then dissociated to single cells in TrypLE. Once dissociated, cells were resuspended in FACS buffer (PBS supplemented with 5% (vol/vol) FCS). Cells were sorted on the BD FACS Aria cell sorter, with cells gated manually into GFP positive and negative cell populations and collected in media. A non GFP expressing cell line was used to position the gate for the GFP negative sample, but was such that some cells with low GFP signal were included in the GFP negative sample. The gating for GFP positive cells was broad to capture cells with both high and moderate levels of GFP expression. Samples from days 0-3 were sorted for live single cells based on size, but were not separated into GFP and non GFP samples as there were too few GFP expressing cells. Day3 samples contained around 100 GFP expressing cells which was not enough to extract sufficient RNA from. Day2 samples had fewer than 10 GFP positive cells and the earlier time points had none. Cells were pelleted at 1200 RPM for 4 minutes at 4°C the supernatant was removed without disturbing the cell pellet and then cells were resuspended in 150  $\mu$ l trizol. The Zymo Direct-zol RNA MicroPrep kit was used to prepare RNA samples according to the manufacturer's instructions including the optional DNaseI treatment. Following RNA extraction, samples were sent to the Wellcome Trust Clinical Research Facility at the Western General Hospital for quality and integrity analysis on the Agilent 2100 Bioanalyser using the RNA 6000 Nano chip. Samples were required to have an RNA Integrity Number of greater than 8. Concentration was quantified using the Qubit RNA broad range assay kit according to instructions. Illumina mRNA-seq libraries were prepared from 200 ng of total RNA using the TruSeq library prep

kit. Libraries were pooled and sent for 75bp paired-end sequencing on the NextSeq 550 platform to generate around 40M reads per sample.

### ATAC-seq

**Cell Lysis and Transposition Reaction.** Cells from 48 organoids were pooled and prepared for FACS as described above, with the live cells counted and sorted at each day and the GFP expressing and non-expressing populations separated at days 4 and 5. ATAC-seq sample preparation was performed as described in [Buenrostro et al. 2015](#), with minor modifications. Cells were pelleted at 1200 RPM for 4 minutes at 4°C and then resuspended in ice cold PBS at a concentration of 100,000 cells per ml. 500  $\mu$ l of cell suspension was pelleted, resuspended in 100  $\mu$ l and centrifuged again before the supernatant was removed. Cells were then resuspended in 100  $\mu$ l of cold ATAC lysis buffer (10 mM TrisCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% IGEPAL CA-630) and kept on ice for 15 minutes, with occasional gentle pipetting. Samples were pelleted at 1200 RPM for 5 minutes at 4°C, supernatant discarded and resuspended in transposition mix consisting of 50  $\mu$ l ChIPmentation buffer (10 mM Tris pH8, 5 mM MgCl<sub>2</sub>, 10% Dimethylformamide) and 2.5  $\mu$ l Tn5 transposase enzyme (Illumina, 15027866) per sample. The transposition reaction was carried out at 37°C for 30 minutes. Samples were then purified using the Qiagen MinElute Reaction Cleanup Kit as per manufacturer's instructions, eluting in 10  $\mu$ l elution buffer.

**PCR Amplification.** PCR amplification was conducted in a 32  $\mu$ l reaction volume for each sample, made up of 10  $\mu$ l transposed DNA, 4.7  $\mu$ l of 10  $\mu$ M custom barcoded nextera primer mix consisting of universal forward primer and one of the barcoded reverse primers, 1.3  $\mu$ l 50X SYBR Green (Invitrogen, S7563) and 16  $\mu$ l NEB-Next 2x PCR master mix (NEB, M0541S). The barcoded primers are detailed in [Table S2](#). Samples were cycled as below:

- 1 72°C for 5 min.
- 2 98°C for 30 sec.
- 3 98°C for 10 sec.
- 4 63°C for 30 sec.
- 5 72°C for 1 min.
- 6 Repeat steps 3-5 for 12 cycles.

Following PCR amplification, 30  $\mu$ l of PCR reaction was made up to 50  $\mu$ l with nuclease free H<sub>2</sub>O.

**Size Selection.** To select against DNA over 1 Kb and remove primers, samples were purified using AMPure XP (Beckman Coulter, A63880) bead size selection. SPRI beads were brought to room temperature and vortexed and 50  $\mu$ l added to each sample at a 1X ratio. Samples were incubated at room temperature for 5 minutes to allow fragment binding. Reactions were then placed in a magnetic stand and allowed to separate for 5 minutes, the supernatant discarded, and the beads washed twice with 200  $\mu$ l 80% ethanol for 30 seconds. The samples were left to air dry for 10 minutes. 20  $\mu$ l 1X TE buffer was added for 30 seconds to elute DNA from the beads. Samples were returned to the magnetic stand to separate the beads from the samples.

**Quantification and Sequencing.** Before sequencing, the quality and quantity of samples was checked. The total amount of DNA was measured using the Qubit dsDNS high-sensitivity assay as per manufacturer's instructions. For quality assessment ATAC-seq

samples underwent high-sensitivity DNA bioanalysis at Edinburgh Genomics. Bioanalysis showed that the libraries contained the expected fragment size distribution, containing peaks corresponding to mononucleosomal, dinucleosomal and trinucleosomal fragments as described in the published protocol ([Buenrostro et al., 2013](#)). ATAC-seq samples were sent for 75bp paired-end sequencing on the Illumina Hi-seq platform at Edinburgh Genomics. Day4 samples were sequenced on a separate run to the other time points resulting in the batch effects that led to this time point being excluded from much of the analysis.

### Genome Editing in mESCs

**CRISPR-Cas9 Mutant Cell Line Generation.** CRE null mESCs and their WT counterparts were generated, using CRISPR-Cas9, from the Rx-GFP cell line. CRISPR single guide RNAs were designed targeting the site of TF binding sequences for the chosen *Rax* and *Six6* peaks. Sequences of the guide RNAs used are detailed in [Table S3](#).

The pSpCas9(BB)-2A-GFP (PX458) (Addgene, 48138) plasmid vector ([Ran et al., 2013](#)) was linearized by digestion with BbsI, gel purified and ligated with an annealed pair of guide oligos. The resulting plasmid DNA was transformed into chemically competent DH5 $\alpha$  cells and purified from liquid culture using the QIAprep Spin Miniprep Kit as per manufacturer's instructions.

Rx-GFP mESCs were transfected with 2  $\mu$ g of plasmid DNA diluted in Opti-MEM (ThermoFisher, 31985062) and 6  $\mu$ l Lipofectamine 2000 transfection reagent (ThermoFisher, 11668019). Media was changed after 6 hours. After 48 hours transfected cells were sorted using FACS based on GFP expression. Cells were plated at a density of 1000 cells per 10 cm dish and grown for around 10 days or until colonies began to appear. Colonies were picked and plated in duplicate in a 96-well plate. One well was used to extract genomic DNA from the cells. The region targeted by the guide RNA was amplified and Sanger sequenced using primers as detailed in [Table S4](#). Initially, cell lines were genotyped using the primers closest to the guide cut site.

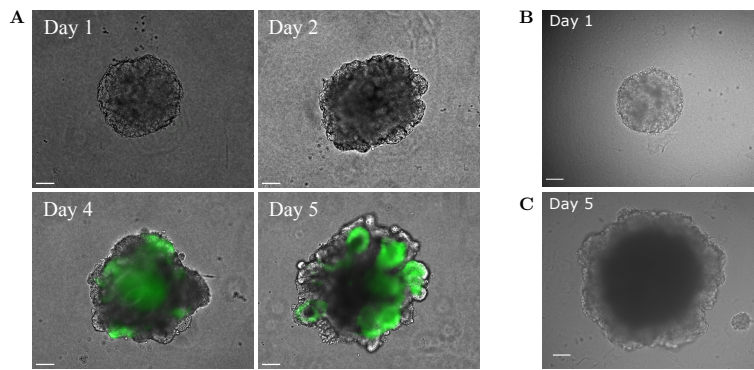
Cell lines with deletions encompassing the targeted TFBSs and control clones, that had no detectable mutations introduced, were expanded from the duplicate 96-well plate. Once expanded, the appropriate distal primers were used to amplify a larger region around the guide cut site and Sanger sequenced ([Fig.S7](#)), both to confirm the deletions are present in the expanded cell line and to check for any larger heterozygous deletions that may have been missed by using the initial primers.

**Nanostring nCounter RNA Quantification.** A custom NanoString CodeSet of 200 genes was designed to include genes that were differentially expressed for each sequential timepoint comparison, including genes that were up and down regulated and 20 housekeeping genes. 24 day5 organoids were pooled, and RNA was extracted as described for the RNA-seq assay. The NanoString nCounter analysis was performed by the HTPU within the IGC. For the hybridisation reactions, 70  $\mu$ l of Hybridisation Buffer was added to each vial of the Reporter CodeSet, 8  $\mu$ l of this was added to each of the hybridisation tubes. 5  $\mu$ l of 20 ng/ $\mu$ l RNA (100 ng total) was added to the appropriate hybridisation tube, followed by 2  $\mu$ l of the Capture Probeset. Tubes were incubated at 65°C for 18 hours. Following hybridisation, the samples were processed using the nCounter Prep station within 24 hours of hybridisation. The hybridised RNA samples and all components of the nCounter masterkit were loaded in the prep station and processed using

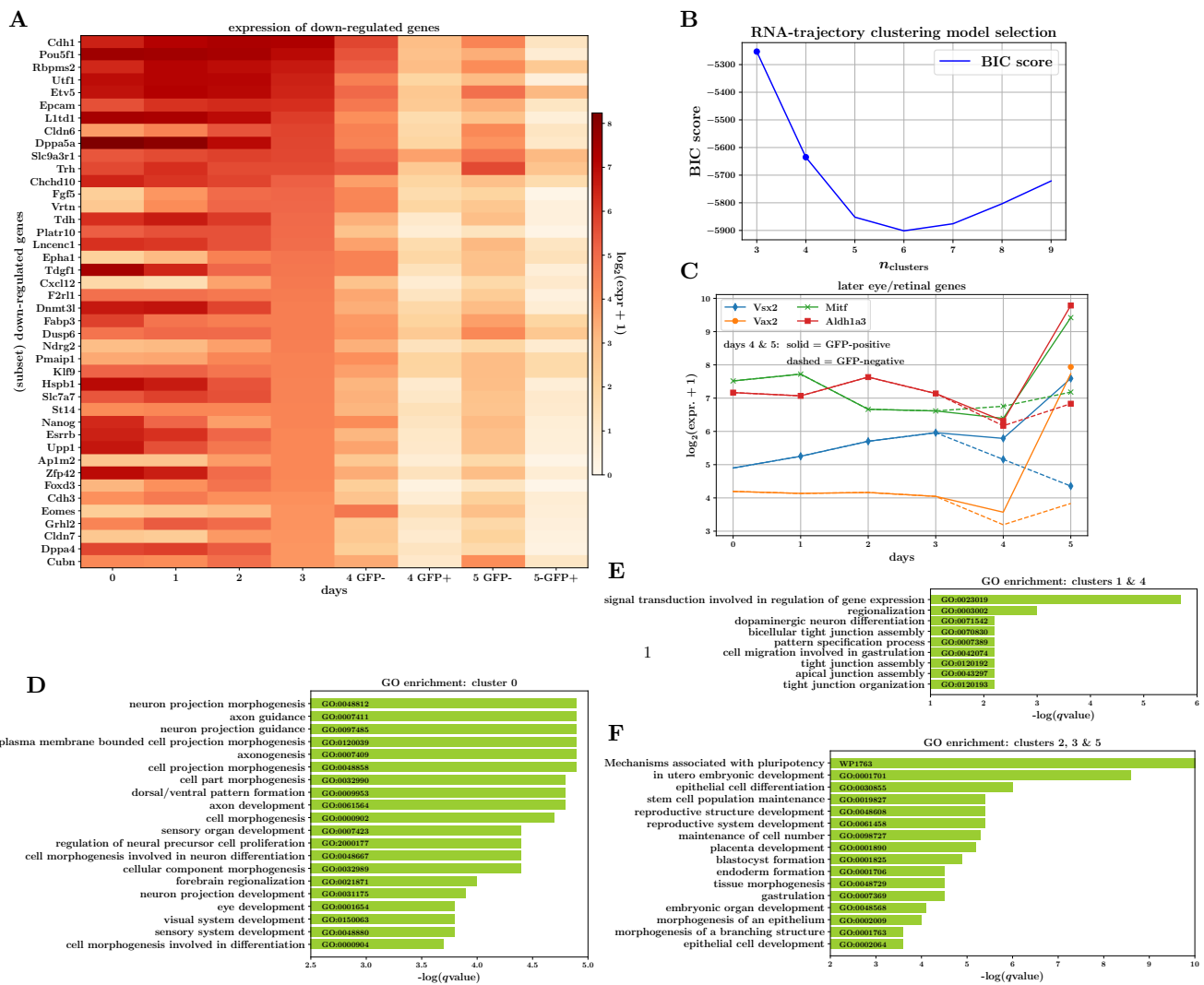
the high sensitivity protocol. The prep station was used to purify the hybridised samples by removing excess probes, then binding, immobilising and aligning them in a sample cartridge for analysis. At this point, each colour-coded barcode is attached to a single target-specific probe corresponding to an analyte of interest. The cartridges were sealed and read in the digital analyser using the max setting to count 555 FOV (Field of View). The Reporter code counts for each sample, as produced by the Digital Analyser, were QCed and normalised using a combination of positive control targets and CodeSet content normalisation, which uses housekeeping genes, to apply a sample specific correction factor to all target probes within that sample lane.

## REFERENCES

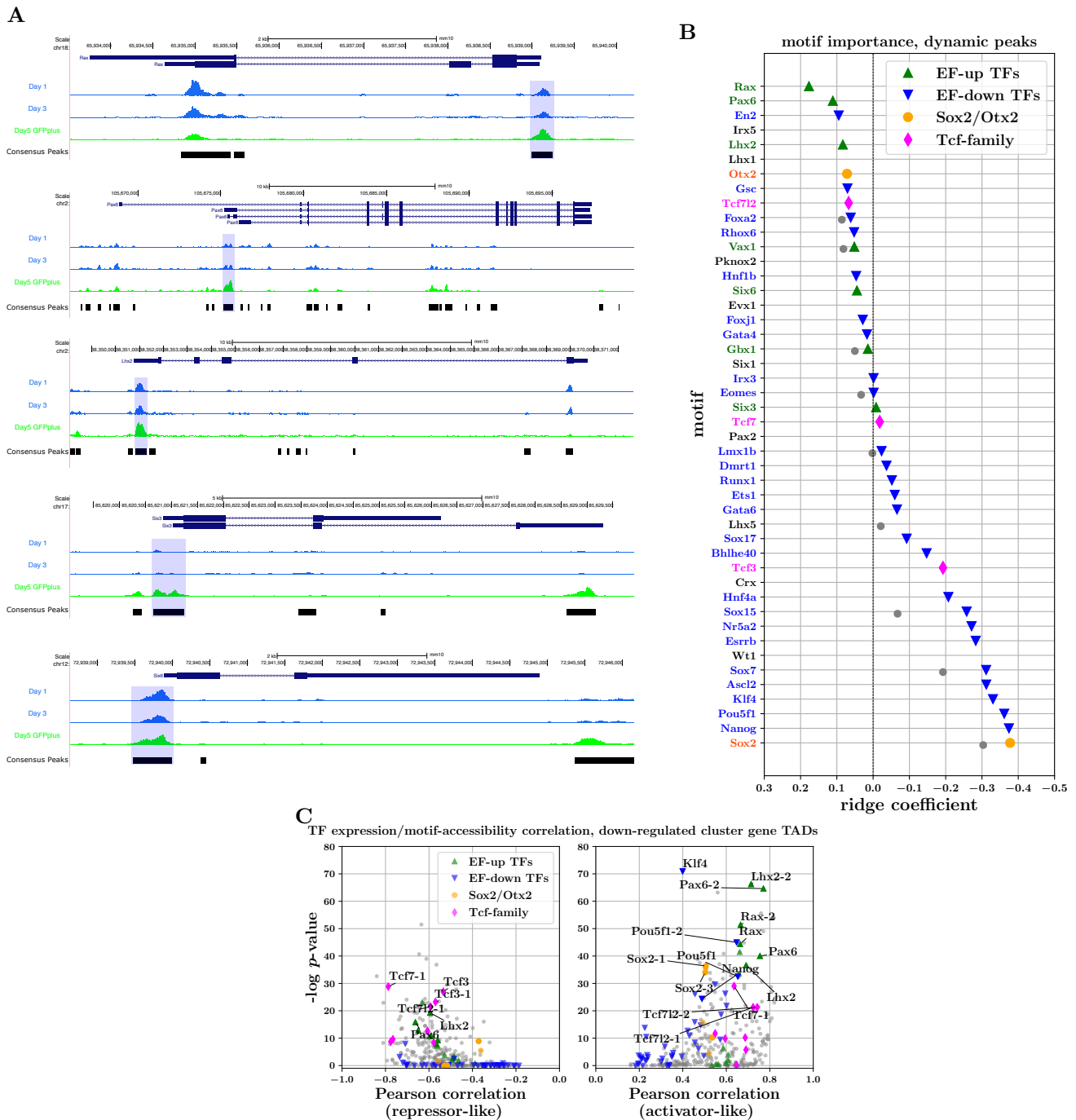
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J.** (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*. 10, 1213–1218
- Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J.** (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current protocols in molecular biology*. 109, 21.29.1–21.29.9.
- Danno H., Michiue T., Hitachi K., Yukita A., Ishiura S. and Asashima M.** (2008). Molecular links among the causative genes for ocular malformation: Otx2 and Sox2 coregulate Rax expression. *Proceedings of the National Academy of Sciences*. 105, 5408–5413
- Puig, R. R., Boddie, P., Khan, A., Castro-Mondragon, J. A. and Mathelier, A.** (2021). UniBind: maps of high-confidence direct TF-DNA interactions across nine species. *BMC Genomics*. 22, 482
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F.** (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*. 8(11), 2281–2308
- Sakakura, E., Eiraku, M. and Takata, N.** (2016). Specification of embryonic stem cell-derived tissues into eye fields by Wnt signaling using rostral diencephalic tissue-inducing culture. *Mechanisms of Development*. 141, 90–99
- Wataya, T., Ando, S., Muguruma, K., Ikeda, H., Watanabe, K., Eiraku, M., Kawada, M., Takahashi, J., Hashimoto, N. and Sasai, Y.** (2008). Minimization of exogenous signals in ES cell culture induces rostral hypothalamic differentiation. *Proceedings of the National Academy of Sciences*. 105(33), 11796–11801



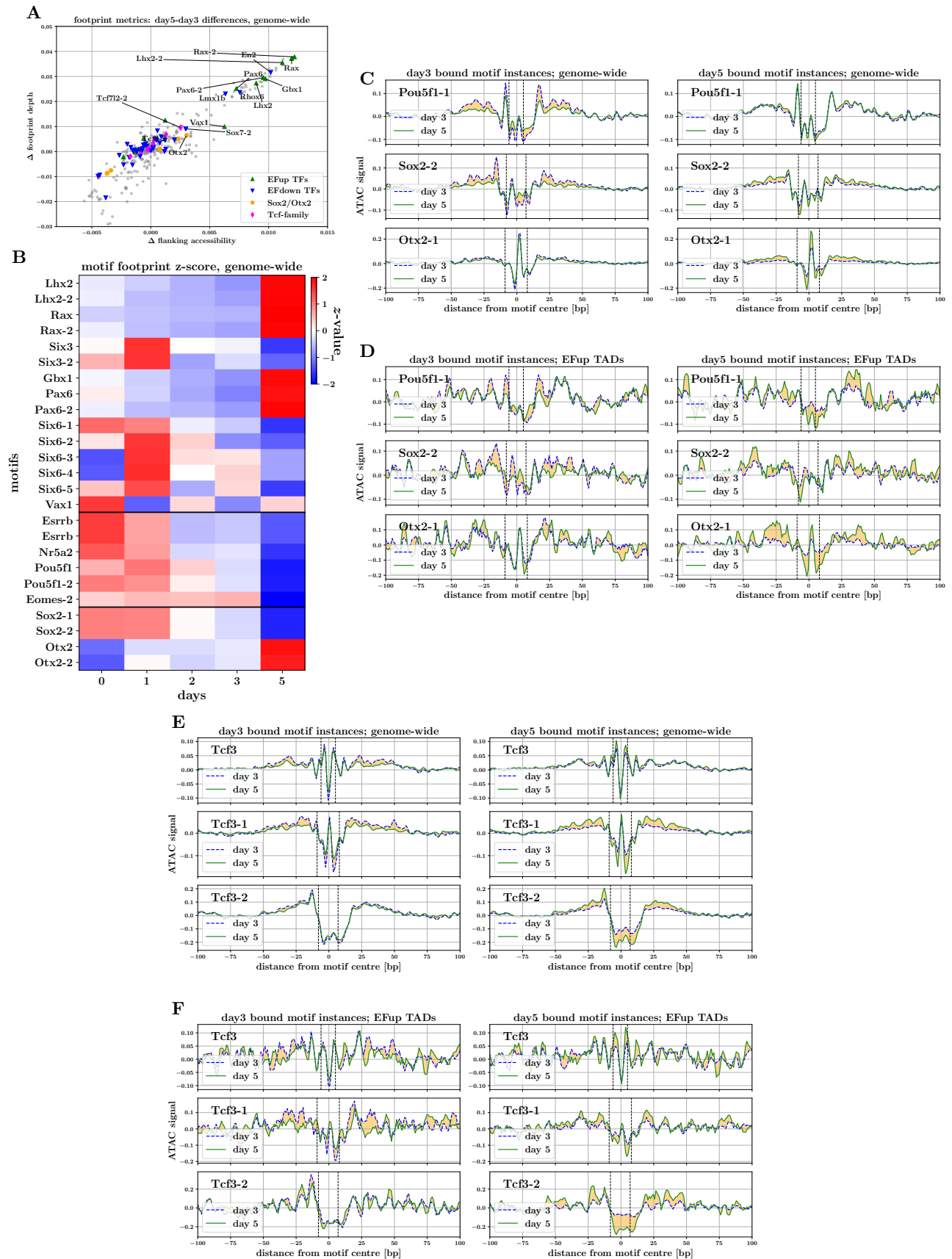
**Fig. S1.** Organoid culture controls for OV timecourse. A. Representative images of organoids cultured in KSR media at days 1, 2, 4 and 5 of growth. The same cell line was used to generate these organoids as was used to grow the organoids in Fig.1A. B. Representative image of organoid cultured in maintenance media at day1. C. Representative image of organoid cultured in maintenance media at day5 showing no GFP expression or OV like structures. Scale bars: 100  $\mu$ m.



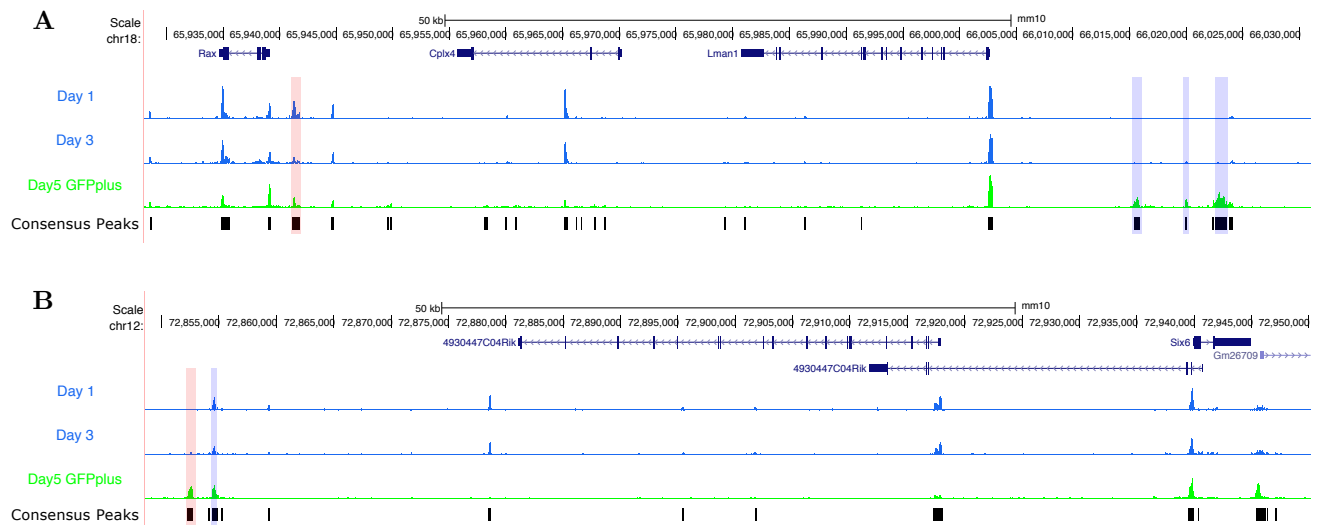
**Fig. S2.** Down-regulated genes indicate pluripotency and other gene-expression programs are suppressed during eye field specification. A. Expression heatmap displaying expression changes across the OV timecourse, for a selection of genes down-regulated across day3-day4/5 transition. B. BIC model selection plot for  $n_{clusters}$  parameter in GMM-clustering. C. Expression of several later eye markers, including *Vsx2*, indicating strong up-regulation in GFP-positive versus GFP-negative cells on day5 of the organoid timecourse. D–F. Gene ontology enrichment analysis of sets of genes within the three trajectory patterns revealed by clustering.



**Fig. S3.** Promoters of canonical eye field TFs are accessible from early in OV timecourse. Motif importance for genome-wide dynamic peaks mirrors importance in EF TADs. A. Bigwig tracks displaying changes in accessibility for regions around canonical EFTFs across the OV development timecourse. Promoters regions of these TFs have been highlighted in blue. B. Coefficients of logistic-regression model trained to predict opening-vs-closing behaviour of genome-wide dynamic ATAC-seq peaks, using presence of TF-motifs as input covariates. Magnitude of coefficients is indicative of importance of respective TF-motifs for these predictions. C. TF-expression/TF-motif-accessibility correlations for EF-down TAD peaks. RHS and LHS plots illustrate the median across positive and negative correlations respectively, indicative of activator-like and repressor-like behaviour of the respective TFs.

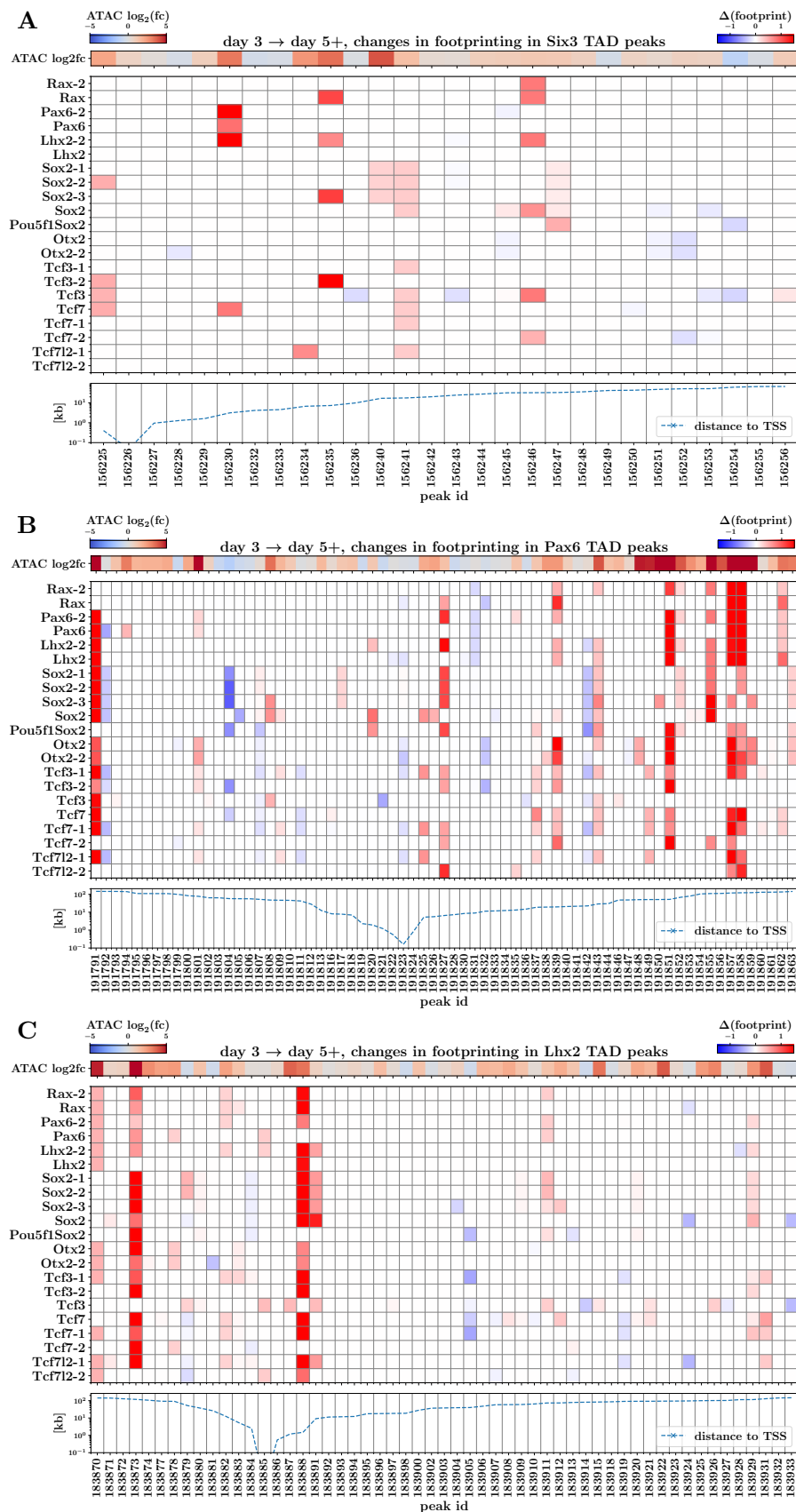


**Fig. S4.** Footprinting patterns genome-wide. **A.** Scatterplot of day3-day5 differences in TF-motif footprint depth versus flanking accessibility, for peaks genome-wide. **B.** Heatmap of TF-motif footprint scores in peaks genome-wide, z-transformed across time-course (median across peaks). **C. & D.** Aggregate footprint signals for Pou5f1, Sox2 and Otx2 binding motifs. Plots illustrate aggregate signal on day3 and day5 for motif-instances predicted to be bound on day3 (left) and day5 (right). **C.** Aggregate footprint signals for genome-wide motif instances. **D.** Aggregate footprint signals for EF-up TAD motif instances. **E. & F.** Aggregate footprint signals for the Tcf3 binding motifs in the JASPAR database. Plots illustrate aggregate signal on day3 and day5 for motif-instances predicted to be bound on day3 (left) and day5 (right). **E.** Aggregate footprint signals for genome-wide motif instances. **F.** Aggregate footprint signals for EF-up TAD motif instances.

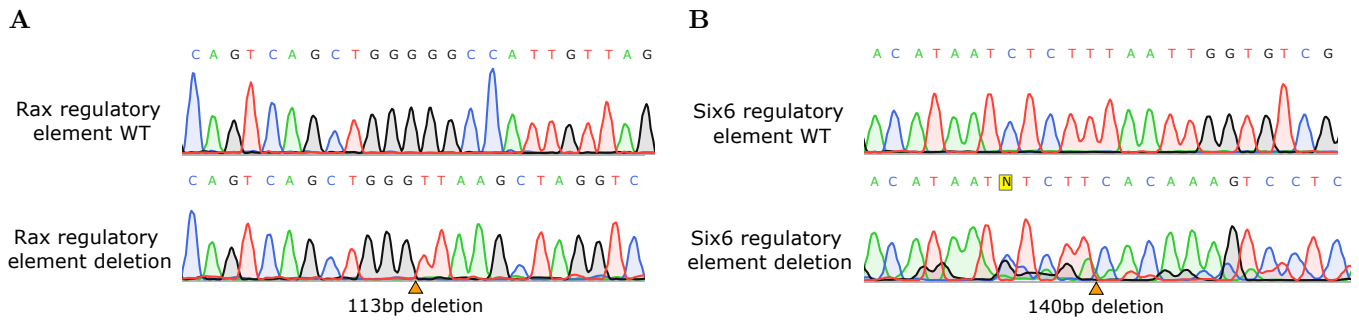


**Fig. S5.** Accessibility changes around candidate *Rax* and *Six6* enhancer elements. A. & B. Changes in ATAC-seq signal around the *Rax* and *Six6* loci, with consensus peaks illustrated by black bars. Peaks showing interesting footprint score changes mentioned in text are highlighted in blue, and the peaks chosen for perturbation in red.

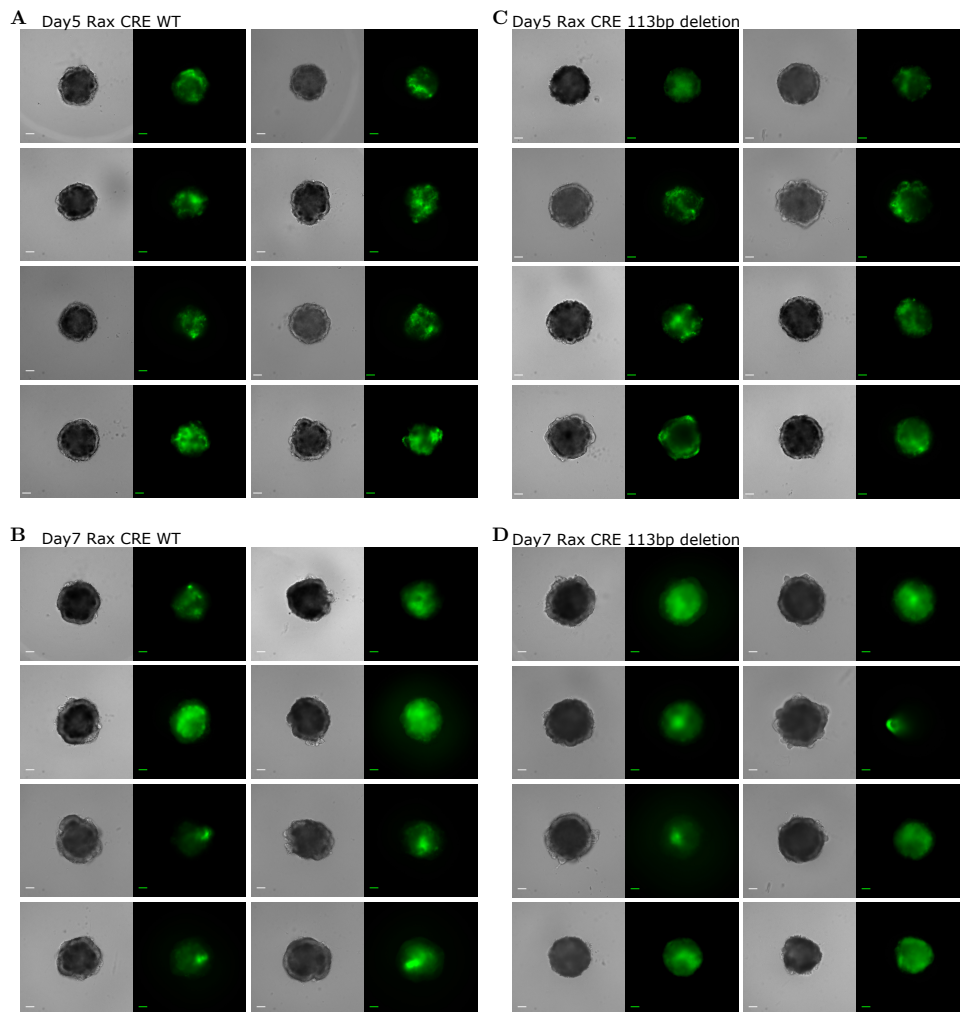




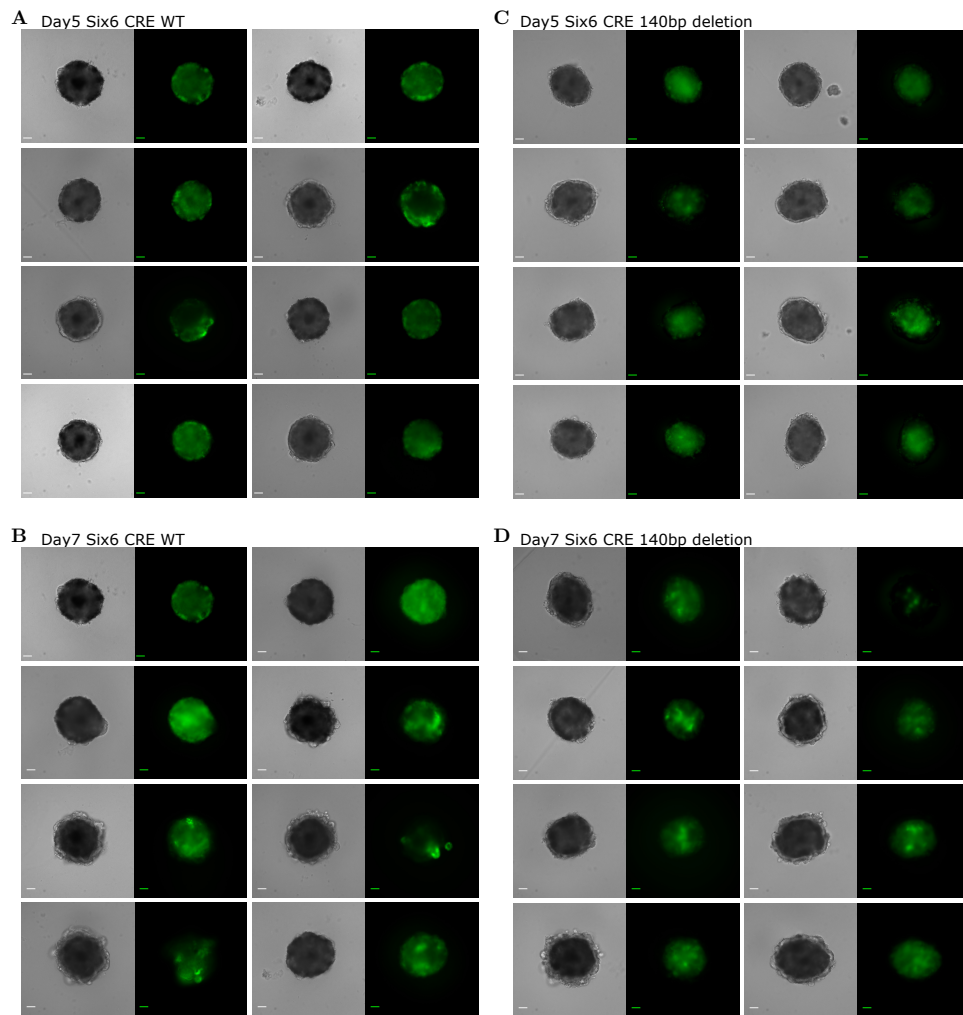
**Fig. S6. Changes in motif footprint scores of key TFs identify candidate *Pax6*, *Lhx2* and *Six3* cis-regulatory elements.** A. Heatmap of day3-to-day5 changes in footprint scores for motifs of key TFs at peaks contained in *Six3* TAD region. Also illustrated are associated changes in normalized ATAC-seq signal for each peak (upper panel), and distance of each peak from the *Six3* TSS (lower panel). B. & C.: same as A. but displaying changes for peaks within *Pax6* and *Lhx2* TAD regions.



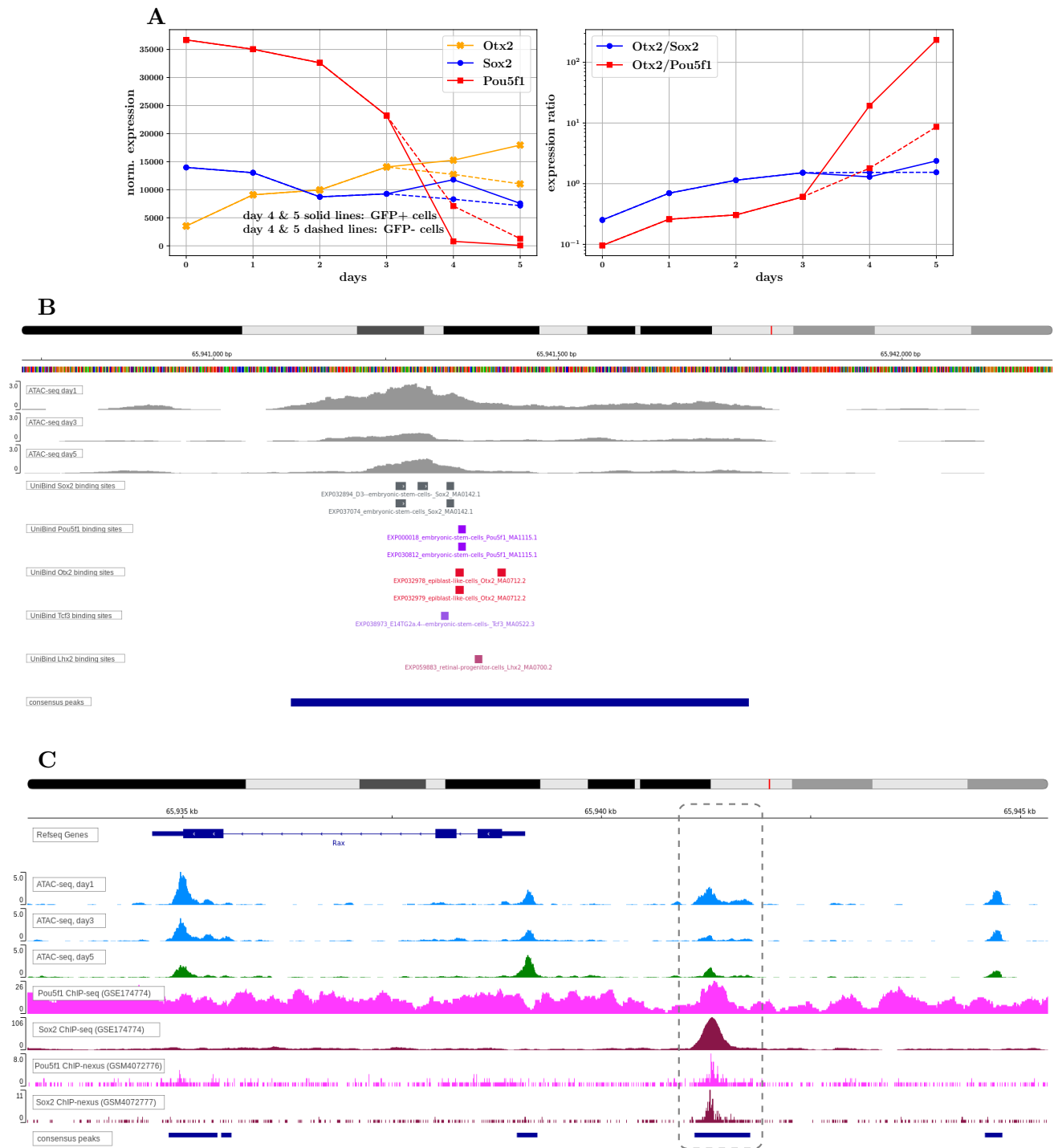
**Fig. S7. DNA sequencing analysis of CRISPR targeted regions.** Sanger sequencing traces of the region around the CRISPR-Cas9 guide target site in the **A. *Rax*** and **B. *Six6*** predicted enhancer regions. The sequence of the WT cell line used as the unedited control is shown in the top panel, and the sequence of the mutant cell line in the lower panel. Orange arrowheads indicate the site of the deletion.



**Fig. S8. Organoids derived from candidate *Rax* enhancer CRISPR-disrupted cell lines.** **A. & B.** Organoids derived from wildtype *Rax* enhancer CRISPR cell-line, at days 5 and 7. 3 replicates of 24 organoids were grown for each cell line, to minimise the stress the organoids were subjected to, 8 were imaged at random. **C. & D.** Organoids derived from mutant *Rax* enhancer CRISPR cell-line, at days 5 and 7. Scale bars: 100  $\mu$ m.



**Fig. S9. Organoids derived from candidate *Six6* enhancer CRISPR-disrupted cell lines. A. & B.** Organoids derived from wildtype *Six6* enhancer CRISPR cell-line, at days 5 and 7. 3 replicates of 24 organoids were grown for each cell line, to minimise the stress the organoids were subjected to 8 were imaged at random. **C. & D.** Organoids derived from mutant *Six6* enhancer CRISPR cell-line, at days 5 and 7. Scale bars: 100  $\mu\text{m}$ .



**Fig. S10. ATAC-seq, ChIP-seq and RNA-seq data lead to hypothesis of *Rax* element activation. A.** Expression (left) and expression-ratios (right) for key TFs hypothesized to play a role in the activation of the proximal *Rax* element identified by [Danno et al. 2008](#), and perturbed in the OV-organoid system using CRISPR. **B.** Evidence of direct TF-DNA interactions, at the locus of the proximal *Rax* element, extracted from the UniBind database ([Puig et al., 2021](#)). Data is shown for *Sox2*, *Pou5f1*, *Otx2*, *Tcf3* and *Lhx2* TFs across all UniBind-processed ChIP-seq experiments for which evidence of TF-binding overlaps with the genomic region of the *Rax* element. Bigwig tracks for day1, day3 & day5 ATAC-seq signal (our organoid data) are also displayed. **C.** Bigwig tracks for day1, day3 & day5 ATAC-seq signal (our organoid data) and E14 mESC *Pou5f1* and *Sox2* ChIP-seq and ChIP-nexus data around *Rax* locus. Candidate regulatory element highlighted in dashed box overlaps with putative *Rax* enhancer element identified by [Danno et al. 2008](#).

**Table S1. Counts of live cells from single dissociated organoids and associated percentages of GFP positive cells.**

Organoid development day	Total Live Cells	Total GFP+ cells	% GFP+
day5 (n=35)	12,731	9,404	74
day6 (n=39)	19,313	13,566	72
day7 (n=38)	19,013	12,002	65
day8 (n=15)	17,811	9,752	54

**Table S2. Nextera PCR primers used for ATAC-seq sample preparation. Ad1 was the forward primer common to all samples, whereas barcoded primers 2.1-2.8 were unique to each sample.**

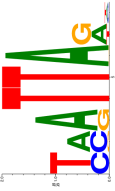
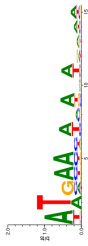
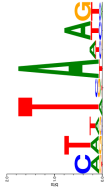

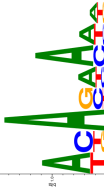

Ad1_noIndex	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATAACGAGATAGGAGTCCGTCTCGTGGGCTCG
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATAACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATAACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATAACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT

**Table S3. Guide RNA sequences for disruption of *Rax* and *Six6* candidate *cis*-regulatory elements.**

Guide Name	Guide RNA Sequence
Rax Forward	CACCGTGTAGATTAGCTCCTAACAA
Rax Reverse	AAACTTGTTAGGAGCTAATCTACAC
Six6 Forward	CACCGATAATCTCTTTAATTGGTGT
Six6 Reverse	AAACACACCAATTAAAGAGATTATC

**Table S4. Sequencing primers used for amplification of putative *Rax* and *Six6* *cis*-regulatory elements and Sangar sequencing for confirmation of mutations**

Primer Name	Primer Sequence
Rax TFBindingSite Fwd	TGGAGCCTGCCTTTGTGTAG
Rax TFBindingSite Rev	CAGGTTGGAGCTGGGAAGAG
Distal Rax Fwd	TCCAGCTGGCTAGGTAGAG
Distal Rax Rev	GAAGCAGTGCATGCTGGATA
Six6 TFBindingSite Fwd	CACAGTGCCAACATGCAAGT
Six6 TFBindingSite Rev	AGCAGGCTTTCCAAAGAGGT
Distal Six6 Fwd	ACAGGAGGGAGACTATGATTGG
Distal Six6 Rev	AGTCGCATAAGACACCGTGG

streme motif	visualization	e-value (streme)	similar motifs & p-values (tomt.com)
1-TMATTARH		$3.0 \times 10^{-34}$	Lhx2-2 ( $1.8 \times 10^{-5}$ ), Rax ( $2.1 \times 10^{-5}$ ), Ak1 ( $4.1 \times 10^{-5}$ ), Alx1-1 ( $4.1 \times 10^{-5}$ ), Esx1 ( $4.8 \times 10^{-5}$ ), Alx3 ( $4.8 \times 10^{-5}$ ), En2 ( $4.8 \times 10^{-5}$ ), En1 ( $6.2 \times 10^{-5}$ ), Phox2a ( $6.2 \times 10^{-5}$ ), Lhx2 ( $6.6 \times 10^{-5}$ ), Prrx1 ( $6.6 \times 10^{-5}$ ), Pax7 ( $6.6 \times 10^{-5}$ ), Pou3f2 ( $7.7 \times 10^{-5}$ ), Nkx1-2 ( $8.9 \times 10^{-5}$ ), Emx2 ( $8.9 \times 10^{-5}$ ), Vax2 ( $1.1 \times 10^{-4}$ ), Lhx6-1 ( $1.4 \times 10^{-4}$ ), Nkx6-1 ( $1.6 \times 10^{-4}$ ), Arx ( $1.6 \times 10^{-4}$ ), Arx ( $1.6 \times 10^{-4}$ ), Msx2 ( $1.8 \times 10^{-4}$ ), Alx1-2 ( $1.8 \times 10^{-4}$ ), Dlx3 ( $1.8 \times 10^{-4}$ ), Pou2f1 ( $1.9 \times 10^{-4}$ ), Evx1 ( $2.4 \times 10^{-4}$ ), Prrx2 ( $2.7 \times 10^{-4}$ ), Rhox6 ( $2.7 \times 10^{-4}$ ), Vax1 ( $2.9 \times 10^{-4}$ ), Rax-2 ( $3.1 \times 10^{-4}$ ), Lhx6-2 ( $3.4 \times 10^{-4}$ ), Msx1 ( $3.7 \times 10^{-4}$ ), Pax6 ( $3.7 \times 10^{-4}$ ), Prrx2 ( $3.8 \times 10^{-4}$ ), Gbx1 ( $3.9 \times 10^{-4}$ ), Nkx6-1-2 ( $4.1 \times 10^{-4}$ ), Pax6-2 ( $4.2 \times 10^{-4}$ ), Lhx1 ( $4.4 \times 10^{-4}$ ), Prrx2 ( $4.8 \times 10^{-4}$ ), Gbx2 ( $5.5 \times 10^{-4}$ ), Dlx3 ( $6.0 \times 10^{-4}$ ), Pou6f1-1 ( $7.7 \times 10^{-4}$ ), Lmx1b ( $8.6 \times 10^{-4}$ ), Nkx6-1-1 ( $1.1 \times 10^{-3}$ ), Isl1 ( $1.1 \times 10^{-3}$ ), Arid3b ( $1.1 \times 10^{-3}$ ), Pou3f4 ( $1.6 \times 10^{-3}$ ), Lhx5 ( $1.8 \times 10^{-3}$ ), Arid3a-1 ( $3.7 \times 10^{-3}$ ), Pou6f1-2 ( $3.7 \times 10^{-3}$ ), Pou3f1 ( $4.5 \times 10^{-3}$ ), Barhl2 ( $6.0 \times 10^{-3}$ ), Hmx1 ( $9.6 \times 10^{-3}$ )
2-ATGAACWMAVWDRAA		$2.4 \times 10^{-23}$	Mtf1-2 ( $2.8 \times 10^{-4}$ ), Foxj3 ( $5.0 \times 10^{-4}$ ), Foxj3-1 ( $5.0 \times 10^{-4}$ ), Lmx1b ( $7.8 \times 10^{-4}$ ), Bbx-1 ( $9.7 \times 10^{-4}$ ), Foxa2-2 ( $1.1 \times 10^{-3}$ ), Six4 ( $1.1 \times 10^{-3}$ ), Glis2-2 ( $3.8 \times 10^{-3}$ ), Sox7-1 ( $4.1 \times 10^{-3}$ ), Sox7-1 ( $4.1 \times 10^{-3}$ ), Tcf3-2 ( $4.5 \times 10^{-3}$ ), Arid3a-1 ( $6.2 \times 10^{-3}$ ), Srf-2 ( $7.0 \times 10^{-3}$ ), Sox4-2 ( $7.0 \times 10^{-3}$ ), Foxk1 ( $7.0 \times 10^{-3}$ ), Foxk1-1 ( $7.0 \times 10^{-3}$ ), Dmbx1 ( $7.3 \times 10^{-3}$ ), Dmbx1-1 ( $9.6 \times 10^{-3}$ ), Mybl1-1 ( $9.7 \times 10^{-3}$ )
3-CTWTSAWAG		$1.0 \times 10^{-8}$	No significant JASPAR motif matches.
4-TVTTAABA		$2.3 \times 10^{-5}$	No significant JASPAR motif matches.
5-ACASAAWW		$6.4 \times 10^{-4}$	No significant JASPAR motif matches.
6-ATNSATCA		$1.9 \times 10^{-2}$	Tcf3-1 ( $2.2 \times 10^{-4}$ ), Tcf7-1 ( $2.5 \times 10^{-4}$ ), Pbx1 ( $3.4 \times 10^{-4}$ )

**Table S5.** Denovo motif discovery results for dynamic peaks with increasing accessibility. Motifs were discovered using streme and matched to motifs in the JASPAR database using tomt.com.

streme motif	visualization	e-value (streme)	similar motifs & p-values (tontom)
1-CCTGAAGCAGGCT		$7.6 \times 10^{-4}$	No significant JASPAR motif matches.
2-CCATATCCTGCA		$1.5 \times 10^{-3}$	No significant JASPAR motif matches.
3-CTGCAGGTGACK		$1.5 \times 10^{-3}$	Arnt1 ( $1.8 \times 10^{-4}$ ), Bhlhe40 ( $9.3 \times 10^{-4}$ ), Arnt ( $1.1 \times 10^{-4}$ ), Tef3 ( $1.3 \times 10^{-3}$ ), Zic1::Zic2 ( $1.4 \times 10^{-3}$ ), Mycn ( $1.7 \times 10^{-3}$ ), Aif3 ( $3.4 \times 10^{-3}$ ), Mif1 ( $4.0 \times 10^{-3}$ ), Eomes-2 ( $4.6 \times 10^{-3}$ ), Zic1-2 ( $4.7 \times 10^{-3}$ ), Zic2-2 ( $4.7 \times 10^{-3}$ ), Zic2 ( $4.8 \times 10^{-3}$ ), Mixp ( $6.0 \times 10^{-3}$ ), Mycn ( $6.4 \times 10^{-3}$ ), Ase12 ( $6.5 \times 10^{-3}$ ), NR4A2 ( $6.9 \times 10^{-3}$ ), Rang(var.2) ( $7.7 \times 10^{-3}$ ), Creb5 ( $8.1 \times 10^{-3}$ ), FOS::JUN ( $8.1 \times 10^{-3}$ ), CREB1 ( $8.8 \times 10^{-3}$ ), Srebfl(var.2) ( $9.1 \times 10^{-3}$ ), Tef12 ( $9.7 \times 10^{-3}$ )
4-GAGGAAGGCTCCAC		$1.5 \times 10^{-3}$	No significant JASPAR motif matches.
5-AACAATAGAGTGA		$2.5 \times 10^{-3}$	Sox5 ( $2.1 \times 10^{-5}$ ), Sox15-1 ( $3.3 \times 10^{-4}$ ), Sox7-1 ( $4.0 \times 10^{-4}$ ), Sox5-1 ( $7.1 \times 10^{-4}$ ), FOXO3 ( $8.5 \times 10^{-4}$ ), Sox8-1 ( $8.7 \times 10^{-4}$ ), Foxj2 ( $1.0 \times 10^{-3}$ ), Sox12-1 ( $1.0 \times 10^{-3}$ ), Sox21-2 ( $1.6 \times 10^{-3}$ ), Sox13-1 ( $1.6 \times 10^{-3}$ ), Nanog ( $2.4 \times 10^{-3}$ ), Sox17-1 ( $3.4 \times 10^{-3}$ ), Sox2 ( $4.6 \times 10^{-3}$ ), Sox11-1 ( $4.6 \times 10^{-3}$ ), Sox4-1 ( $6.1 \times 10^{-3}$ )
6-GAGATTACYTA		$3.8 \times 10^{-3}$	No significant JASPAR motif matches.
7-ATTCAITCCW		$2.6 \times 10^{-2}$	No significant JASPAR motif matches.

**Table S6. Denovo motif discovery results for dynamic peaks with decreasing accessibility. Motifs were discovered using streme and matched to motifs in the JASPAR database using tontom.**

**Table S7. Table of differentially expressed genes across day3-to-day5 transition in organoid development timecourse.** For each gene the table documents whether the gene was up or down regulated across day3-day5 transition, and to which trajectory cluster it belongs (see Fig.2B).

[Click here to download Table S7](#)