



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Characterization of an eye field-like state during optic vesicle organoid development

Citation for published version:

Owen, L, Rainger, JK, Bengani, H, Kilanowski, FM, Fitzpatrick, D & Papanastasiou, A 2023, 'Characterization of an eye field-like state during optic vesicle organoid development', *Development*, vol. 150. <https://doi.org/10.1242/dev.201432>

Digital Object Identifier (DOI):

[10.1242/dev.201432](https://doi.org/10.1242/dev.201432)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Development

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Characterization of an eye field-like state during optic vesicle organoid development

Liusaidh J. Owen¹, Jacqueline Rainger¹, Hemant Bengani¹, Fiona Kilanowski¹, David R. FitzPatrick¹ and Andrew S. Papanastasiou^{1,2,*}

ABSTRACT

Specification of the eye field (EF) within the neural plate marks the earliest detectable stage of eye development. Experimental evidence, primarily from non-mammalian model systems, indicates that the stable formation of this group of cells requires the activation of a set of key transcription factors. This crucial event is challenging to probe in mammals and, quantitatively, little is known regarding the regulation of the transition of cells to this ocular fate. Using optic vesicle organoids to model the onset of the EF, we generate time-course transcriptomic data allowing us to identify dynamic gene expression programmes that characterize this cellular-state transition. Integrating this with chromatin accessibility data suggests a direct role of canonical EF transcription factors in regulating these gene expression changes, and highlights candidate *cis*-regulatory elements through which these transcription factors act. Finally, we begin to test a subset of these candidate enhancer elements, within the organoid system, by perturbing the underlying DNA sequence and measuring transcriptomic changes during EF activation.

KEY WORDS: Eye field specification, Transcription factors, Gene regulatory networks

INTRODUCTION

Vertebrate eye development begins with the formation of the eye field – a patch of morphologically indistinct neuroectodermal cells within the anterior neural plate defined by co-expression of a set of transcription factors (TFs). The eye field was first identified in *Xenopus laevis* embryos (Zuber et al., 2003; Zuber, 2010) and expresses a specific set of TFs, known as the eye-field transcription factors (EFTFs), with those being: *PAX6*, *RAX*, *SIX3*, *SIX6*, *LHX2*, *OTX2*, *NR2E1* and *TBX3* (gene symbols of human orthologues). Mutations in at least three of the genes encoding orthologues of the *Xenopus* EFTFs can cause severe bilateral eye malformations in humans (*OTX2*, *PAX6* and *RAX*) (Fitzpatrick and van Heyningen, 2005). It seems likely that these genes are involved in eye-field specification in humans. The EFTFs are thought to form part of a gene regulatory network (GRN) essential for initiating eye development and the goal of this work is to characterize this GRN in

mammals and gain a deeper understanding of how it is initiated and stably maintained.

The eye field has been difficult to study in mammals as it forms in early post-implantation embryos (~18 post-ovulatory days in humans, approximately embryonic day 7.5 in mouse). The Sasai lab developed a simple defined organoid culture system (Eiraku et al., 2011; Eiraku and Sasai, 2012; Nakano et al., 2012; Sakakura et al., 2016) that supports efficient differentiation of human and mouse embryonic stem cells (ESCs) into early eye structures; optic vesicles (OVs) evaginate from the main body of the organoid and then develop into optic cup structures with partially differentiated neural retina. This transformative technology established that early eye development is a self-organizing system and provides a window onto the hitherto cryptic processes driving the commitment to eye fate in mammals.

We have performed a functional genomic analysis of the eye field-like cells that appear in the very early stages of optic vesicle organoid differentiation. By performing computational analyses of the transcriptomic changes during OV differentiation, we find that establishment of the eye field in mouse ESC-derived OVs involves strong upregulation of a set of genes that includes the same key transcription factors identified in *X. laevis*. Furthermore, by integrating transcriptomic and chromatin-accessibility data, we show that these TFs likely regulate the stable transition of cells to an ocular fate by binding distal enhancers in genomic domains local to EF genes. By examining how chromatin accessibility, associated with key TFs, changes at individual genomic loci, we generate hypotheses of plausible *cis*-regulatory elements (CREs – enhancers and repressors) required for differentiation from pluripotency to eye-field cell states. Finally, we test a subset of these hypotheses by perturbing the underlying DNA sequences. In all, by exploiting the robust and reproducible model system offered by mESC-derived organoids, our study provides a first quantitative understanding of the regulatory mechanisms underlying mammalian eye-field specification.

RESULTS

Mammalian eye-field genes overlap significantly with *Xenopus* EFTFs

Mammalian eye-field specification – the crucial first stage of ocular development – is an event that occurs very early in development (during gastrulation), and as such is extremely challenging to study *in vivo*. We have exploited a reproducible, *in vitro* organoid model system enabling us to generate data from this cell-state transition and through computational analysis gain a quantitative understanding of the underlying regulatory mechanisms. We performed OV organoid culture using mouse embryonic stem cells (mESCs) that have a GFP fluorescent reporter cassette knocked into the *Rax* locus (Eiraku and Sasai, 2012). The expression pattern of *Rax* suggests that cells expressing this gene are either competent to become, or are already, retinal progenitors (Furukawa et al., 1997), although *Rax* is also

¹MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK. ²School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK.

*Author for correspondence (andrew.papanastasiou@ed.ac.uk)

 A.S.P., 0000-0002-8888-0633

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Handling Editor: Steve Wilson

Received 7 November 2022; Accepted 2 June 2023

expressed in the hypothalamus (Wataya et al., 2008). Rax-positive cells became detectable between days 3 and 4 in culture (Fig. 1A), and at days 5 and 6 $\geq 70\%$ of live cells were Rax positive (Table S1). By day 5, the cell aggregates had developed a 3D structure typical of retinal organoids, as previously described (Sakakura et al., 2016). We performed bulk RNA-sequencing (RNA-seq) analysis using biological triplicates of OV organoid cultures at days 0, 1, 2, 3, 4 and 5. To enrich for cells in an eye-field state, at days 4 and 5 the cells were sorted by fluorescence-activated cell sorting (FACS) into GFP-positive and GFP-negative populations (Fig. 1B).

To identify genes characterizing the transition to the eye field, we looked for differentially expressed genes between day 3 and the GFP-positive samples on days 4 and 5 [using DESeq2 (Love et al., 2014); false discovery rate (FDR) < 0.001 , absolute \log_2 -fold-change > 1.5]. This analysis revealed a small set of ‘EF-up’ genes (37) that were significantly upregulated between these time points, of which 11 correspond to known mouse TFs (Fig. 2A). Importantly, these included the canonical EFTF genes *Rax*, *Pax6*, *Lhx2*, *Six3* and *Six6*, indicating that drivers of the optic-fate in mammals include most of the key TFs also discovered in *Xenopus*. *Six6* was expressed slightly later, consistent with evidence that it is a downstream target of *Lhx2* and *Pax6* (T treault et al., 2009). Although not sufficiently upregulated to pass our threshold, *Nr2e1* displayed an expression pattern consistent with that of the other canonical EFTFs. The only canonical EFTF with an unexpected

expression pattern – one that is anti-correlated with the onset of the eye field – was *Tbx3*. This could be because *Tbx3* plays a different role in mouse than in *Xenopus* (*Tbx3* is a known component of the pluripotency network in mESCs; Russell et al., 2015) or because the *Xenopus Tbx3* orthologue in mouse may not be *Tbx3*. *Tbx2* is the likely candidate, being the only Tbx family member for which expression is highly correlated with onset of the eye field (Fig. 2A). Finally, as expected, when inspecting genes differentially upregulated specifically between the GFP-positive and -negative cells at day 5, we found markers of later eye development, including *Vsx2*, *Vax1*, *Mitf* and *Aldh1a3* (Fig. S2C). This provides transcriptomic evidence that Rax-expressing cells in the organoid system are transitioning to the ocular fate. We found low levels of Rax expression in the GFP-negative samples on days 4 and 5 (Fig. 2A). This is to be expected as our cell-sorting strategy groups cells with low levels of GFP, including those just beginning to express Rax, into the GFP-negative samples.

In parallel with the upregulation of this extended set of putative eye-field genes, the cell-state transition was also characterized by the downregulation of a larger set of ‘EF-down’ genes (448) across the day 3-to-day 4/5 transition (Fig. S2A). Many of these genes are known players in pluripotency, including *Nanog*, *Pou5f1* and *Dppa5a* [more generally, 35 EF-down genes were linked to the gene ontology (GO) terms ‘mechanisms associated with pluripotency’ or ‘stem cell population maintenance’], and although they were downregulated across the state transition, they were also consistently downregulated across the full time course of organoid differentiation. This suggests that successful specification of the eye field requires both the network of eye-field genes to be turned on as well as genes controlling pluripotency to be switched off. It is noteworthy that *Sox2* and *Otx2*, known to be crucial in eye development, are not differentially expressed across this crucial time point (Fig. 2A), consistent with these genes being more broadly expressed in the anterior neuroectoderm *in vivo*. Furthermore, although Tcf family genes are important in repressing β -catenin/Wnt signalling to allow forebrain specification (Kim et al., 2000; Dorsky et al., 2003) and play a role in eye-field specification in zebrafish (Young et al., 2019), we found that key Tcf family genes, including *Tcf3*, *Tcf7*, *Tcf7l1* and *Tcf7l2*, are expressed in the OV organoids but not differentially expressed across this time point (Fig. 2A).

Next, we probed our RNA-seq data further to see whether any of the EF-up or EF-down genes displayed significant changes between earlier time points. Performing similar differential expression analyses, we found that nine (of 165) genes upregulated between days 2 and 3 overlapped with the EF-up genes, including the TFs *Rax*, *Lhx2*, *Six3* and *Fezf1*. The fact that these key factors are significantly upregulated so early in OV development (prior to detection of a distinct GFP signal), may point to these TFs acting as very early eye-field determining factors. Interestingly, 12 genes upregulated between days 2 and 3 actually overlapped with EF-down genes, including *Wnt3*, *Wnt8a*, *Fgf8* and *Lmx1b*, pointing towards the eye-field state requiring repression of gene expression programmes that were turned on earlier in differentiation. This has close parallels with what is thought to happen *in vivo*, where, as well as promoting ocular fate, EFTFs are also involved in repressing gene expression programmes characterizing other neural cell states, such as diencephalic, mid- and hindbrain states, early in development (Roy et al., 2013; Fish et al., 2014; Liu et al., 2010; Takata et al., 2017).

Finally, to explore whether there are general patterns to the trajectories of differentially expressed genes, we performed a clustering

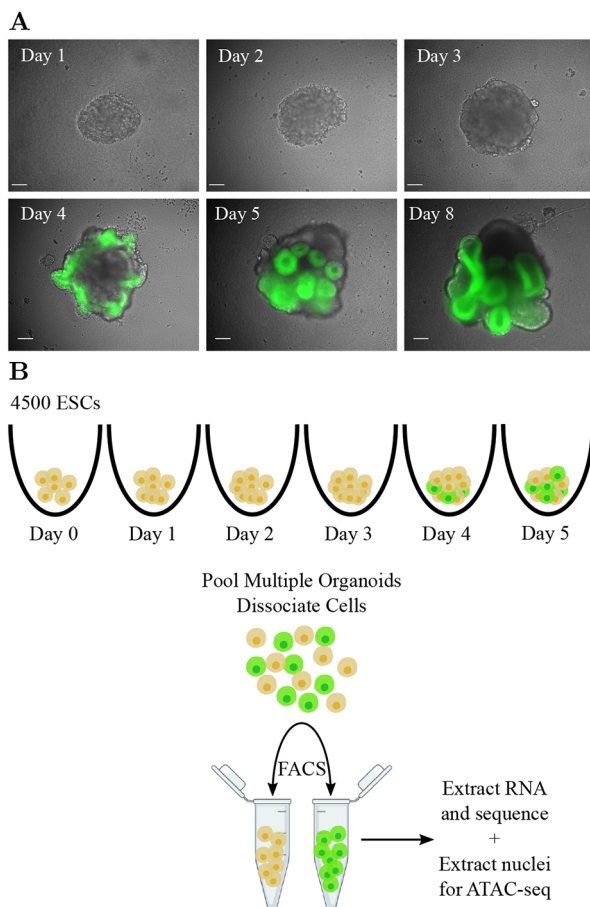


Fig. 1. Exploiting optic vesicle organoids to probe the molecular origins of eye-field specification. (A) Representative organoids from days 1-5, and day 8. Scale bars: 100 μm . (B) Schematic of the experimental setup for time-course RNA-seq and ATAC-seq data generation.

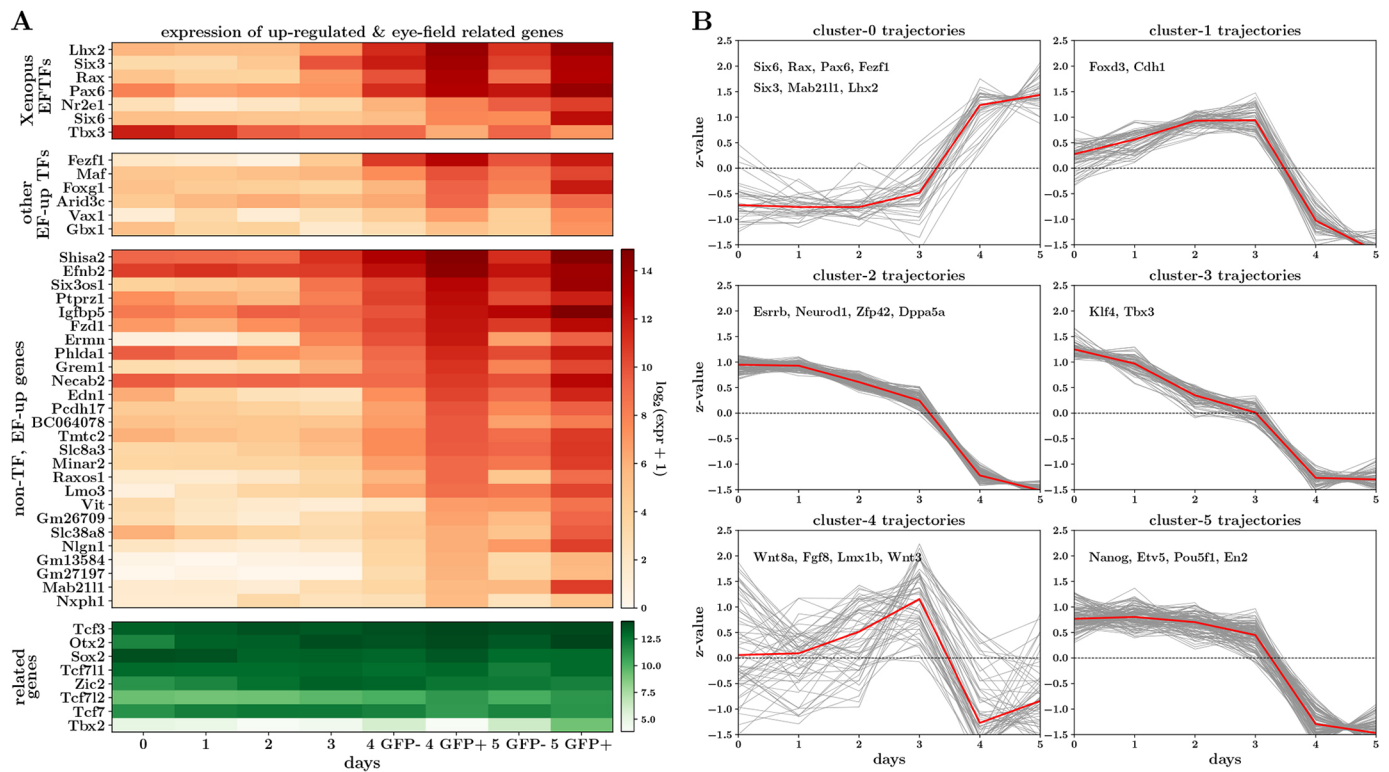


Fig. 2. RNA-seq analysis reveals coordinated upregulation of canonical EFTFs during optic vesicle development. (A) Heatmap displaying gene expression across the OV time course, for canonical eye-field genes, differentially upregulated (EF-up) genes and genes known to have an important role in early eye development. (B) Clustering of EF-up and EF-down gene trajectories reveals six clusters with three underlying patterns across the time course. At each time point, heatmap and trajectory plots indicate expression averaged across triplicates.

analysis on the trajectories of the EF-up and EF-down genes (Fig. 2B). This analysis resulted in six clusters, which fall into three trajectory patterns: genes in clusters 0 are steeply upregulated after day 3 (all the EF-up genes fell into this category); clusters 2, 3 and 5 contain genes with expression gradually downregulated with time; and clusters 1 and 4 contain genes with more transient expression, being broadly upregulated until day 3 and then sharply repressed after day 3. GO enrichment analysis (Zhou et al., 2019) on the gene sets displaying these three broad patterns provided an additional layer of evidence that the stable onset of the eye field is characterized by a balance of very different gene expression programmes: the upregulation of genes specifically related to eye and neuron development (Fig. S2D), and, in parallel, the downregulation of genes characterizing pluripotency, non-neuroectodermal differentiation (Fig. S2F) and possibly cell signalling and migration (Fig. S2E). In summary, RNA-seq data generated from the time course of OV organoid development show that this is a robust and relevant model system with which to study the gene dynamics underlying mammalian eye-field specification.

Dynamic ATAC-seq peaks local to eye-field genes suggest enhancer-driven logic

Changes in the expression of genes characterizing cell-state transitions, in development and disease, are thought to be largely controlled by changes in DNA accessibility and binding of TFs at CREs genomically local to those genes (Spitz and Furlong, 2012; Ong and Corces, 2012; Buecker and Wysocka, 2012; Long et al., 2016; Klemm et al., 2019; Ray-Jones and Spivakov, 2021; Minnoye et al., 2021). To gain insight into the regulation of the transcriptional changes important for the eye-field transition, and specifically to identify candidate TF-CRE pairs that control the expression of EF

genes, we performed bulk assay for transposase-accessible chromatin with sequencing (ATAC-seq; Buenrostro et al., 2013) experiments on single samples of organoids at days 0, 1, 2 and 3, as well as GFP-positive and GFP-negative cells on day 5. We identified a consensus set of 361,867 genomic regions displaying a significant accessibility peak on one or more days. Mirroring the strategy used in the RNA-seq analysis, we looked for peak regions that were dynamic (absolute log₂-fold-change >1.5) across days 3 and 5, and found 7782 and 53 such peaks with increased and decreased ATAC-seq signal, respectively. After excluding peaks overlapping exons (TFs typically bind non-exonic regions), we found that this set of dynamic peaks comprises largely (~97%) non-promoter peaks. This is consistent with previous work indicating that changes in the chromatin accessibility landscape during early development predominantly occur in non-promoter regions (Reddington et al., 2020), suggesting an enhancer-driven logic to cell-state changes.

We used GREAT analysis (McLean et al., 2010) to computationally assign putative links between these dynamic ATAC-seq peaks to genes in an unbiased manner. This revealed gene–peak associations for 22 of the 37 EF-up and 161 of the 448 EF-down genes, suggesting that the observed changes in chromatin accessibility may be functionally relevant in the regulation of genes characterizing the eye field.

Our key goal was to uncover changes in chromatin accessibility that play a direct role in regulating the changes in gene expression identified using RNA-seq data. To identify CREs relevant to the putative eye-field genes, we focused on peaks that lie within the topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012) containing those genes (using TADs defined from previously

published mESC Hi-C data; Bonev et al., 2017). This approach was motivated by the observation that TADs are remarkably stable across development and tissue types and that regulatory elements generally lie within the TAD containing the target gene (Symmons et al., 2016). We found that 98% and 93% of dynamic peaks in up- and downregulated eye-field TADs, respectively, correspond to non-promoter peaks. By contrast, peaks overlapping the promoters of many eye-field genes, including *Rax*, *Pax6* and *Lhx2*, were accessible from day 0, despite very low or zero corresponding gene expression (Fig. 3B; Fig. S3A). Indeed, whereas promoter regions of the putative eye-field genes displayed higher median ATAC-seq signals than the corresponding dynamic non-promoters (Fig. 3A; $P < 10^{-229}$, Mann–Whitney *U*-test), trajectories of the latter were more variable across the time course (Fig. 3A; $P < 10^{-243}$, Mann–Whitney *U*-test, and Fig. 3B). The above observations are suggestive of a model whereby gene-expression changes underlying the transition to ocular fate are orchestrated largely by changes in accessibility and TF binding at distal and intronic regulatory elements.

Motifs of upregulated eye-field genes are enriched in and predictive of regions of dynamic chromatin

As TFs bind DNA in a sequence-biased manner, analysing the sequence content of accessible genomic regions provides preliminary evidence of the TFs that may be involved in gene regulation. We performed motif enrichment analysis on the peaks identified as dynamic across the transition from day 3 to day 5 (Fornes et al., 2020; Grant et al., 2011). We found that peaks with increasing accessibility have a strong enrichment for the motifs of the canonical EFTFs *Rax*, *Lhx2* and *Pax6*, whereas peaks with decreasing accessibility appear enriched for pluripotency factors motifs such as *Pou5f1* and *Nanog* (Fig. 3C). Furthermore, a *de novo* motif discovery analysis on peaks with increasing accessibility revealed that the inferred enriched motifs are most similar to (amongst others) the consensus motifs for *Rax*, *Lhx2* and *Pax6* (see Table S5). This indicates that TFs characterizing the transcriptomic state transition may also play an important role in re-shaping and/or activating regulatory regions of chromatin. To assess whether the evidence from the motif-enrichment analysis of genome-wide dynamic peaks is consistent with accessibility dynamics in peaks related to the EF-up genes, we performed a multivariate determination of motif importance for peaks within EF-up TADs. Specifically, we fitted a logistic-regression model to predict whether the accessibility signal of a peak increases or decreases across the transition from day 3 to day 5, using as input the binary motif presence of differentially expressed TFs. We found that the motifs corresponding to upregulated TFs, in particular *Lhx2* and *Rax*, are most predictive of opening chromatin, whereas motifs corresponding to pluripotency factors are predictive of closing chromatin (Fig. 3E). The overall picture of EFTFs and TFs associated with pluripotency being predictive of opening and closing chromatin peaks, respectively, appears to be a genome-wide pattern as we found very similar coefficient patterns when we repeated the modelling exercise for genome-wide dynamic peaks instead of peaks in EF-up TADs (Fig. S3B).

This multivariate approach also reveals that there are a number of downregulated TFs that are predictive of increases in accessibility. This could be explained by considering that a subset of the highest-scoring downregulated TFs, including *En2*, *Gata6* and *Irx3* have been reported in other biological contexts to act as transcriptional repressors (Tolkunova et al., 1998; Whissell et al., 2014; Bilioni et al., 2005) and therefore their downregulation may explain some of the increases in accessibility. A second possible explanation is that

motifs for many TFs are similar and have overlapping occurrences across the genome (our motif discovery analysis showed that *En2* and *Lmx1b* motifs display significant similarities to the same *de novo* motif that strongly matches *Rax* and *Lhx2*; Table S5). This common issue with standard motif analyses makes it difficult to gain deeper insights into the molecular logic. One approach to move towards a more reliable set of regulatory TFs is to use ATAC-seq together with RNA-seq data. Assuming that a TF that alters the accessibility of regulatory elements must be expressed to do so, there ought to be a correlation between the accessibility of its DNA binding motif and its gene expression. Using an approach similar to that of Corces et al. (2016) and Berest et al. (2019), for each TF we computed the correlation between its expression and the accessibility of each peak containing the TF motif, and then compared the distributions of positive and negative correlations against relevant background distributions to determine significance. We found that for non-promoter peaks within EF-up gene TADs, the canonical EF genes *Rax*, *Pax6*, *Lhx2* and *Six3* as well as *Tcf3* and *Tcf7l2* displayed a significant positive expression-accessibility correlation (Fig. 3D; median corr > 0.5 , $P < 10^{-10}$ Wilcoxon rank-sum test), suggesting activator roles. The same analysis revealed potential repressive roles for the pluripotency factors *Pou5f1*, *Nanog* and *Sox2*, which displayed significant negative expression-accessibility correlation (Fig. 3D; median corr < -0.5 , $P < 10^{-10}$ Wilcoxon rank-sum test) for peaks within EF-up gene TADs. This is in line with evidence suggesting that, as well as being important for activating CREs maintaining pluripotency, these genes also act in a repressive manner to block activation of elements crucial for the upregulation of differentiation genes (Thomson et al., 2011; Wang et al., 2012). Similarly, we found evidence of a strong negative expression-accessibility correlation for *Tcf7* (Fig. 3D). Applying this analysis to peaks in EF-down TADs (Fig. S3C), we found that pluripotency factors there appear to have activator roles, in line with knowledge that these TFs bind enhancers that upregulate their expression. Surprisingly, we also found that peaks within EF-down TADs containing EFTF motifs tend to have an accessibility signal that correlates well with EFTF expression (Fig. S3C). This observation hints that, alongside being crucial for specifying the eye field, the EFTFs may also play a role at CREs repressing genes characteristic of alternative gene expression programmes (consistent with evidence that EFTFs repress alternative cell fates *in vivo*; Roy et al., 2013; Fish et al., 2014; Liu et al., 2010; Takata et al., 2017).

Taken together, the above TF-motif guided analyses and modelling of chromatin-accessibility changes point to the hypothesis that the stable upregulation of eye-field genes is enabled through the gain of binding of a small set of EFTFs at non-promoter CREs that are becoming increasingly accessible. Additionally, our analyses provide evidence that TFs downregulated across OV development act in a repressive manner. Therefore, loss of their binding at CREs with decreasing accessibility, perhaps through loss of ‘silencing’, may also contribute to the activation of the EF-up genes.

Footprinting analysis provides evidence of differential binding of TFs

Detectable ‘footprints’ or relative depletions in chromatin accessibility signal at regulatory elements, can result from DNA being protected from transposase cleavage owing to the presence of bound TFs, and can provide indirect evidence of TF binding. Footprinting analyses have the potential to distinguish subtle effects, such as changes in the distribution of TF binding across conditions, which can be particularly useful if regulatory elements do not display overall changes in accessibility. We applied the TOBIAS framework (Bentsen et al.,

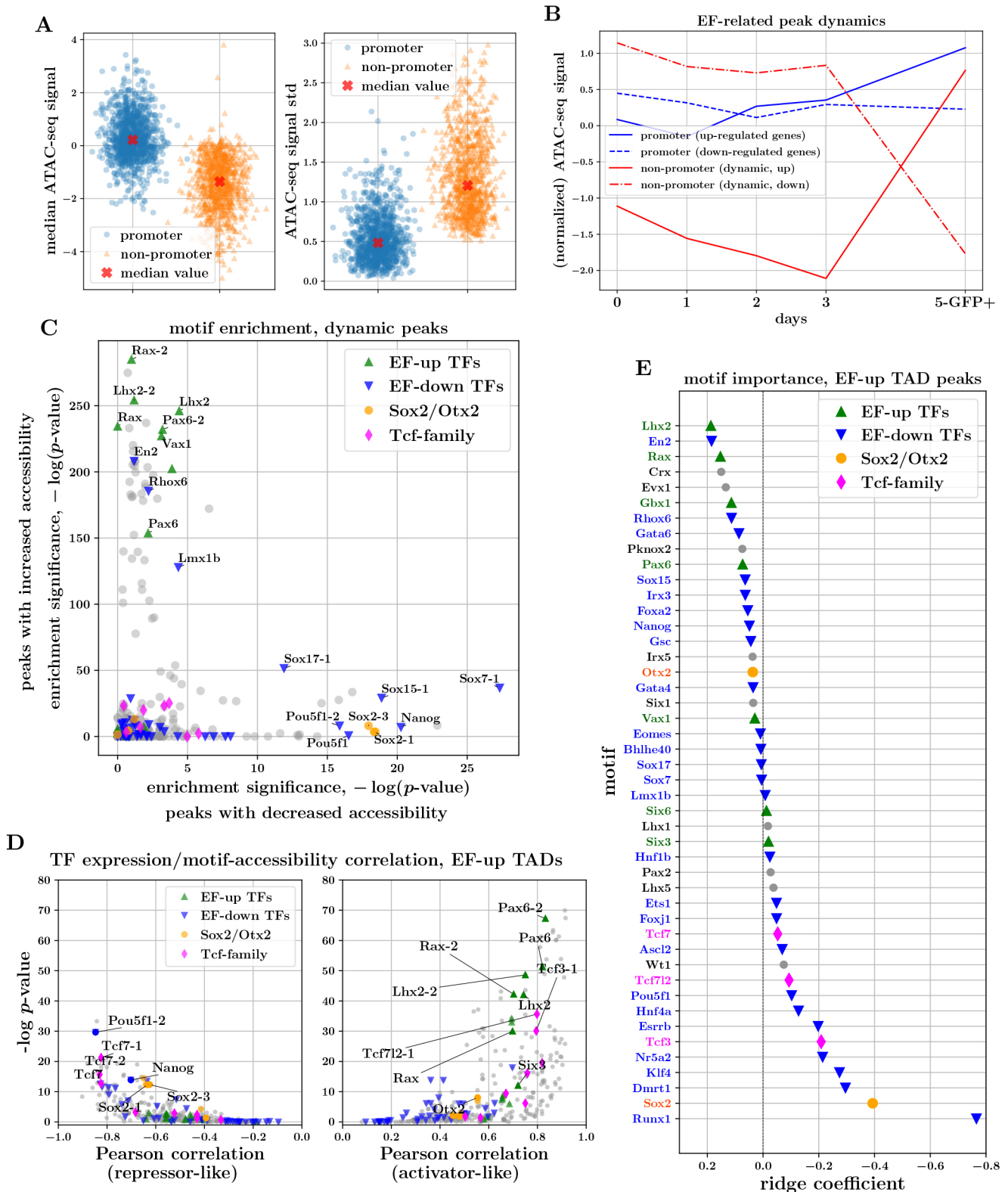


Fig. 3. Characteristics of ATAC-seq peaks associated with the transition to eye field point to enhancer-driven logic and regulatory roles for canonical EFTFs. (A) Mean and variability of ATAC-seq signal in promoter and non-promoter peaks associated with EF genes. (B) ATAC-seq trajectories of EF-up and EF-down gene promoters, and dynamic non-promoter peaks within EF TADs. (C) TF-motif enrichment analysis of peaks with increasing (y-axis) and decreasing (x-axis) accessibility across the transition from day 3 to day 5. (D) TF-expression/TF-motif-accessibility correlations for peaks within EF-up TADs. Plots on the right and left illustrate the median across positive and negative correlations, respectively, indicative of activator-like and repressor-like behaviour of the respective TFs. (E) Coefficients of the logistic-regression model trained to predict opening versus closing ATAC-seq peaks within EF-up TADs, using the presence of TF motifs as input covariates. The magnitude of coefficients is indicative of the importance of the respective TF motifs for these predictions.

2020) to our ATAC-seq data to identify putative changes in binding and footprinting of expressed TFs, both at a genome-wide level and in EF-TAD peaks. TOBIAS computes a 'footprint score' for each occurrence of a TF motif, which represents how well the underlying sequence matches the TF motif and the relative depletion in chromatin accessibility at the motif occurrence.

We performed differential binding analysis between days 3 and 5 to compare footprint scores for every TF-motif occurrence across the genome (providing an aggregate pattern across all occurrences of a motif within peaks of interest). This revealed that some of the largest increases in motif binding correspond to EF-up TFs, most notably *Rax*, *Pax6* and *Lhx2* motifs (Fig. 4A), and, similarly, that some of the most significant decreases in motif binding correspond to EF-down TFs, such as *Pou5f1* and *Dmrt1*. Consistent with our

motif analysis in the previous section, *Six3* did not have a strong increase in binding across the transition from day 3 to day 5, perhaps indicating that *Six3* plays a regulatory role in eye-field specification that does not involve direct binding to regulatory elements of EF genes (e.g. *Six3*-mediated *Wnt8* suppression has been shown to be important for neuroretina development; Liu et al., 2010). Interestingly, we found evidence of increased binding of *Otx2*, consistent with the hypothesis that, together with *Sox2*, *Otx2* is required for coordinated upregulation of the EFTFs and, in particular, leads to activation of *Rax* expression (Danno et al., 2008).

We assume that a TF playing a regulatory role by direct binding to DNA should display well-correlated footprint-score and expression trajectories. To investigate this, for each motif we computed the correlation of the footprint score of each occurrence with expression

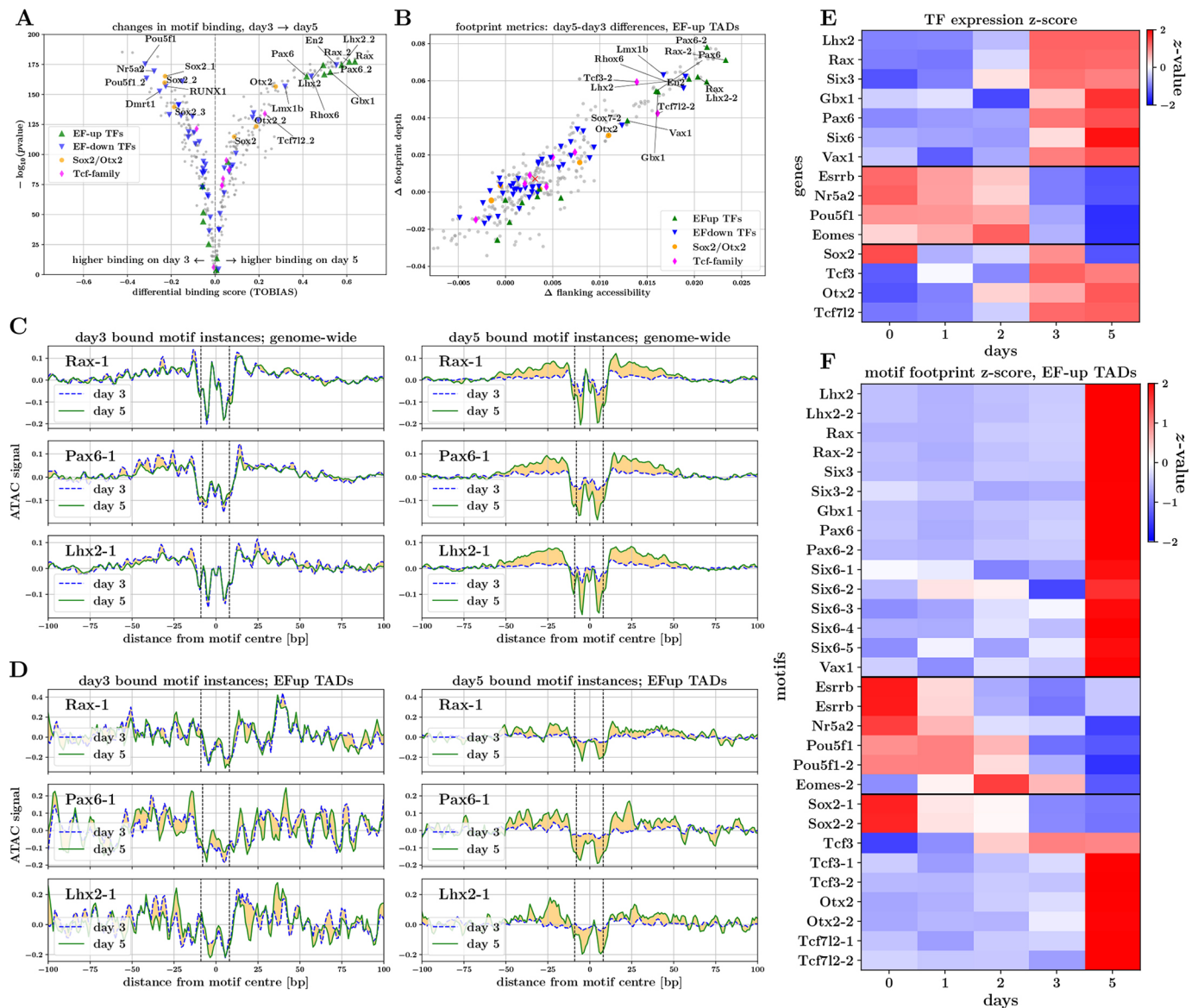


Fig. 4. Footprinting analysis provides evidence of direct interactions of EFTFs with DNA. (A) Differential binding scores for days 3 and 5 for motifs of expressed TFs. (B) Scatterplot of day 3-day 5 differences in TF-motif footprint depth versus flanking accessibility, for peaks in EF-up TADs. (C,D) Aggregate footprint signals for *Rax*, *Pax6* and *Lhx2* motifs. Plots illustrate aggregate signal on day 3 and day 5 for motif instances predicted to be bound on day 3 (left) and day 5 (right). Aggregate footprint signals are shown for genome-wide (C) and EF-up TAD (D) motif instances. (E,F) Expression and motif footprint score trajectories for TFs displaying strongest expression-footprint correlations in EF-up TADs. (E) Heatmap of TF-expression values (average over triplicates) z-transformed across the time course. (F) Heatmap of TF-motif footprint scores in EF-up TAD peaks, z-transformed across the time course (median across peaks).

of the corresponding TF. The set of motifs that displayed a median correlation >0.5 in EF-up-TAD peaks contained all key EF-up TFs (Fig. 4E,F), plus Sox2, Otx2, Tcf3 and Tcf712. When comparing footprint score dynamics in EF-up-TAD peaks and genome wide, we found that there is generally good agreement for most motifs. Notably, however, footprint scores for Six3 and Six6 motifs broke this trend and genome-wide occurrences did not correlate well with their corresponding expression level (Fig. S4B).

Aggregating ATAC-seq signal around multiple occurrences of a given motif allows visualization and quantification of TF footprints. To unpick large changes in footprinting between days 3 and 5 in an unbiased manner, we computed the depth and flanking accessibility for each motif, for genome-wide and EF-up TAD peak occurrences, on each day. Comparing the changes in footprint depth and flanking accessibility between these time points, we found that in EF-up peaks (Fig. 4B) as well as genome wide (Fig. S4A), motifs of the upregulated TFs Rax, Pax6 and Lhx2 (as well as Vax1 and Gbx1) display some of the largest positive changes amongst all motifs of expressed TFs. Similar large increases in these scores were also found for the downregulated TFs Lmx1b, En2 and Rhox6, drawing parallels with our modelling of opening/closing chromatin peaks using motifs (Fig. 3E).

Finally, to confirm that the increases in footprint depth scores correspond to visually discernible footprints, we visualized the aggregate ATAC-seq signal at and around occurrences of a given motif. To pull apart changes between days, we compared the aggregate signal for motif occurrences that are predicted by TOBIAS to be 'bound' on day 3 and on day 5. Although there was little difference in aggregate signal between days, for occurrences predicted to be bound on day 3, we found a striking increase in footprint depths for Rax, Pax6 and Lhx2 motif occurrences predicted to be bound on day 5. This observation applied both to genome-wide peaks as well as to peaks contained in EF-up gene TADs (Fig. 4C,D). This provides further evidence that these TFs are regulating the onset of eye-field specification by binding elements in the genomic locality of EF-up genes. We also found that aggregate footprints for Pou5f1, Sox2 and Otx2 show differences (albeit less striking) that are consistent with eye-field establishment being correlated with loss of pluripotency factor binding and gain of Otx2 binding (Fig. S4C,D). Finally, the aggregate footprints for one of the Tcf3 motifs showed evidence of an increase in binding on day 5, both genome wide and within EF-up peaks (Fig. S4E,F), providing further hints at a direct regulatory role of this TF (or the Tcf family more generally) in eye-field specification.

Peak and TF-specific footprint-score changes identify candidate *cis*-regulatory elements

The motif and footprint analyses described above provide consistent evidence that the canonical EFTFs Rax, Pax6 and Lhx2 play a direct regulatory role in establishing a stable eye field, specifically through binding non-promoter elements in accessible regions within TADs containing EF-up genes. Next, we moved from this broad picture of regulation towards identifying candidate gene-specific CREs through which these key TFs act. For a given EF-up gene, we approached this by searching for ATAC-seq peak regions within the TAD containing the gene, which displayed large changes in footprint scores for motifs of any of these three key TFs, as well as of Sox2, Otx2, Tcf3, Tcf7 and Tcf712. We performed this focused intra-TAD analysis for the five core EFTFs (*Rax*, *Pax6*, *Lhx2*, *Six3* and *Six6*), looking for changes between days 1, 3 and 5 that either mirror the upregulation of the corresponding gene or which point to a loss of repression.

This approach is visualized in Fig. 5 for peaks within the *Rax* (Fig. 5A,B) and *Six6* (Fig. 5C,D) TADs, across the day 3-5 and day 1-3 time points. Focusing on changes from day 3 to day 5 within the *Rax* TAD (Fig. 5A), the heatmap revealed highly suggestive footprint-score changes across the majority of the key TFs, in three peak regions (peak IDs 165713, 165714 and 165716), mirroring upregulation of the *Rax* gene itself (Fig. S5A). A similar pattern was seen within the *Six6* TAD (Fig. 5C) where two peak regions (peak IDs 73704 and 73706) showed high footprint-score changes and increasing accessibility (Fig. S5B). No striking increases in footprint scores were identified across the day1-3 transition; however, there were peaks that showed a decrease in scores particularly for the Sox2-related motifs (most notably peak ID 165695 proximal to *Rax*). We found similar signal patterns, highlighting a small number of peaks for each EF-up gene (results for *Pax6*, *Lhx2* and *Six3* are shown in Fig. S6).

As a step towards validating our approach for identifying gene-specific CREs, we selected two of the EFTFs for follow-up testing: *Rax*, because of the GFP readout of *Rax* expression within the organoid model system, and *Six6*, because of the relative simplicity of the intra-TAD footprinting patterns. We then selected one representative candidate CRE for *Rax* and one for *Six6* (Fig. 5, red peak IDs) to test for enhancer-like behaviour. Although accessibility of the candidate peak proximal to *Rax* remained relatively stable over time (Fig. 6A; Fig. S5A), the footprint-score changes indicate loss of Sox2 binding from day 1 to day 3 and then gain of binding from day 3 to day 5, suggesting this peak's function may switch from repressive to activating, a pattern not seen in other peaks in this TAD. This region was further prioritised as it overlaps with a putative CRE identified by Danno et al. (2008) in *Xenopus* (discussed further in the next section). The candidate peak selected for *Six6* was in a closed chromatin state at day 1, became accessible by day 5 (Fig. 6D; Fig. S5B) and had the largest increases in footprint scores from day 3 to day 5 (within the *Six6* TAD).

We used CRISPR-Cas9 (Ran et al., 2013) to introduce deletions encompassing the TF-binding motifs identified in the candidate CREs into the *Rax*-GFP cell line (see supplementary Materials and Methods). In the *Rax* peak, we introduced a 113-bp homozygous deletion (chr18:65941339-65941452) (Fig. S7A) that disrupts motifs for Sox2, Sox2-Pou5f1, Otx2, Pax6 and Tcf3. This includes the Sox2 and Otx2 motifs identified as regulating *Rax* expression in *Xenopus* (Danno et al., 2008). In the *Six6* peak, a 140-bp homozygous deletion (chr12:72852423-72852563) (Fig. S7B) disrupted one site predicted to be bound by Lhx2, Rax and Pax6, and additionally disrupts Sox2- and Tcf3-binding motifs. We used these cell lines to generate day 5 OV organoids and compared their differentiation with control cell lines that had been transfected with CRISPR-Cas9 reagents but retained the wild-type sequence (Fig. 6B,E). The organoids grown from all CRISPR-Cas9-transfected cell lines were considerably more variable in terms of the OV-like structures formed and the levels of GFP expression than previous batches of organoids, (Figs S8, S9), likely owing to the stress the cells were subjected to throughout the genome-editing process, meaning no clear conclusions could be drawn from imaging alone. To test the hypothesis that these genomic regions act as gene-specific CREs, and that perturbing them has a measurable effect on expression of the target gene, we performed digital gene expression (NanoString) analysis (Geiss et al., 2008) in triplicate on organoids at day 5 of differentiation, grown from cell lines that had no detected mutations introduced by the CRISPR machinery (referred to as wild type) and cell lines with the mutations described above.

Comparing the expression of key genes in wild-type and mutant organoids, we found that disrupting the candidate CREs for both *Rax* and *Six6*, results in downregulation of the associated target gene

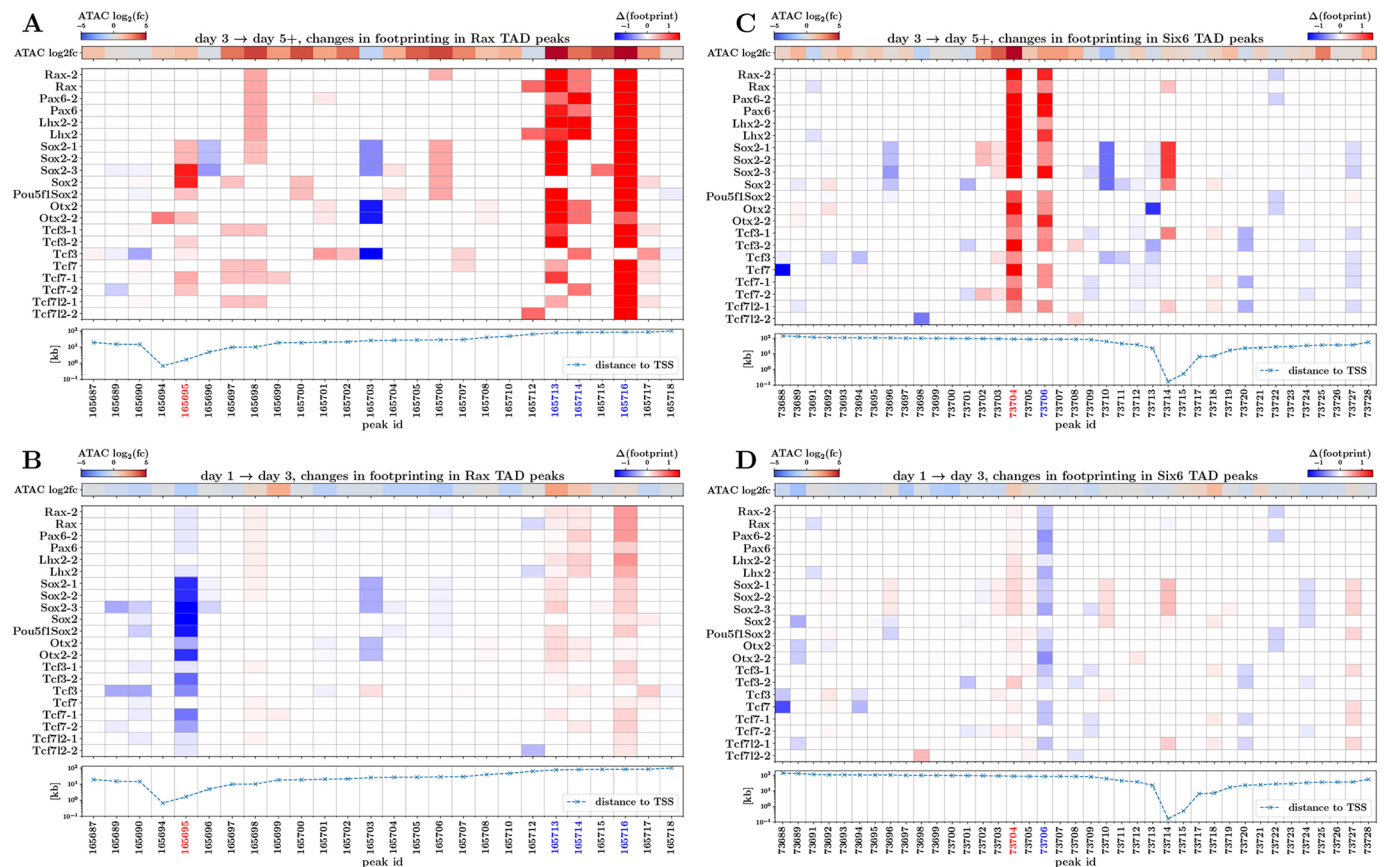


Fig. 5. Changes in motif footprint scores of key TFs identify candidate *Rax* and *Six6* CREs. (A) Heatmap of changes between days 3 and 5 in footprint scores for key TF motifs at peaks in the *Rax* TAD region. Also illustrated are changes in ATAC-seq signal (top), and distance from the *Rax* TSS (bottom) for each peak. (B) As in A, but displaying changes between days 1 and 3. (C,D) As in A,B, but displaying changes in peaks contained in the *Six6* TAD. Red peak IDs indicate those chosen for CRISPR perturbation; blue peak IDs indicate other putative CREs with interesting footprint score changes (see also Fig. S5).

(Fig. 6C,F; -11% for *Rax* and -16% for *Six6* log₂ expression, respectively). We highlight that developmental genes are generally regulated by more than one CRE (Buecker and Wysocka, 2012; Spitz and Furlong, 2012; Long et al., 2016), and therefore we do not expect complete ablation of *Rax* or *Six6* expression. Notably, for both mutants, *Tbx2* was one of the most downregulated genes, further supporting our hypothesis that *Tbx2* is the mammalian equivalent of *Tbx3* in *Xenopus*. Furthermore, we also found that in both *Rax* and *Six6* putative CRE mutant organoids, the expression of all EFTFs is reduced, as might be expected if this set of key TFs co-regulate each other (directly or indirectly) within a GRN. Interestingly, in line with the notion that the EFTFs play a role in suppression of the pluripotency programme, we found evidence of pluripotency genes (e.g. *Zfp42* and *Pou5f1*) being more highly expressed in mutant compared with wild-type organoids. Conversely, markers of later eye development (e.g. *Vsx2* and *Vax2*) were significantly downregulated in mutant organoids.

Regulatory switches responding to ratios of TF concentration may trigger initial EFTF activation

With the resolution of our data, the mechanism actually triggering the expression of key EFTFs is difficult to disentangle. To answer this, we note that experiments (albeit in model systems different to mESC-derived organoids) in which just one of the TFs *Rax*, *Pax6*, *Lhx2* and *Six3* has been knocked out or rendered null, have shown that expression of the non-perturbed genes is activated in the

appropriate regions (Zhang et al., 2000; Fish et al., 2014; Grindley et al., 1995; Bernier et al., 2001; Yun et al., 2009; Roy et al., 2013; Liu et al., 2010). Maintenance of this expression is observed to be significantly de-stabilized. This promotes the hypothesis that the events leading to onset of individual EFTF expression are not dependent on the EFTFs themselves.

These observations indicate that molecular events resulting in *Rax* being switched on, and specifically, the changes at CREs triggering *Rax* expression, do not necessarily involve other EFTFs (*Pax6*, *Lhx2*, *Six3*). Related to this, the work of Danno et al. (2008) showed, in *Xenopus* animal cap cells, that Sox2 and Otx2 proteins interact to activate *Rax* expression, by co-binding a conserved 35-bp region within a CRE ~ 2 kb upstream of the *Rax* promoter. The authors propose that activation of this element is sensitive to the ratio of Sox2 and Otx2 concentrations (Fig. S10A).

As mentioned, the OV organoid ATAC-seq data we have generated in this study contains a peak that overlaps with the conserved element identified by Danno et al. (2008). Intriguingly, we find that this putative CRE is accessible (with no large changes in overall accessibility) throughout the organoid development time course, and thus appears to be primed for activation. Furthermore, if it is the ratio of Sox2 and Otx2 that leads to the activation of this CRE, then we would expect to see sizeable differences in Sox2 and Otx2 binding between OV development days. Indeed, this is what we did find for Sox2; the footprint score for Sox2 motifs decreased from day 1 to day 3 and increased again from day 3 to day 5,

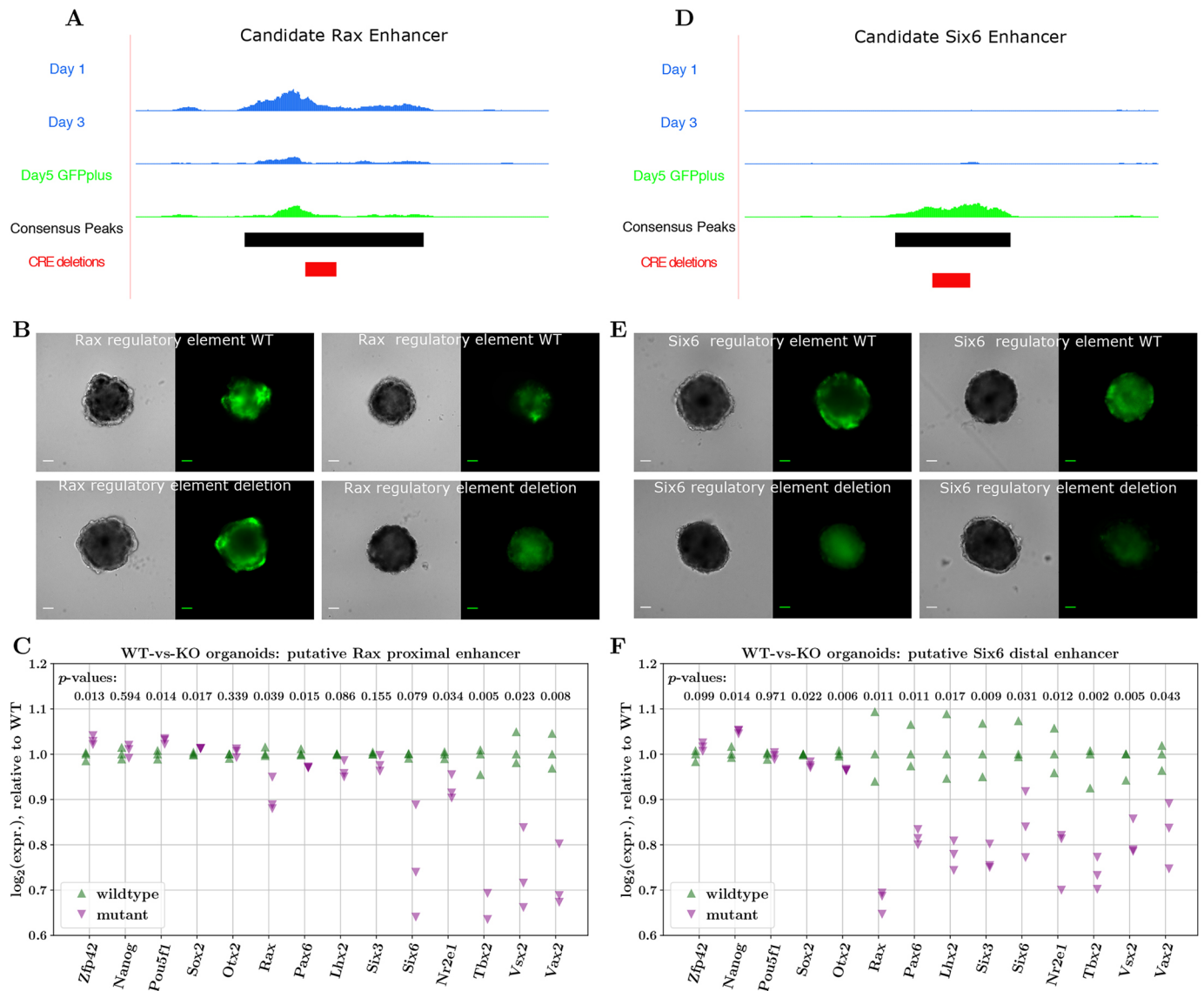


Fig. 6. Changes in expression of key TFs upon disruption of putative EFTF CREs. (A,D) Bigwig tracks of candidate *Rax* and *Six6* enhancers, displaying accessibility changes across days 1, 3 and 5, consensus peaks and regions deleted by CRISPR. (B,E) Brightfield and GFP images of day 5 organoids derived from CRISPR-Cas9-edited cell lines with wild-type (WT; top) or mutant sequences for the predicted *Rax* and *Six6* CREs. Images represent the variability seen in organoid structure and GFP expression. Twenty-four organoids were grown in three replicates, and all showed similar levels of variation between organoids. Scale bars: 100 μ m. (C,F) NanoString quantification of gene expression between WT and mutant (KO) organoids for the proximal *Rax* (C) and distal *Six6* (F) enhancers (WT/mutant differences in \log_2 expression tested using Welch's *t*-test).

specifically at this regulatory element, possibly indicating changing of Sox2 binding partners across the time course (Fig. 5A,B). The predicted binding ratios for Otx2 at this element were less pronounced, although they did increase from day 3 to day 5, in line with the working hypothesis.

To gain further insights into the activation of this element, we mined the UniBind database (Puig et al., 2021) of TF–DNA interactions for robust (combined ChIP-seq and computational modelling) evidence of binding sites of specific TFs in this genomic region, across all (mouse) experimental models. This revealed several TF-binding sites that may be relevant to our system (Fig. S10B), including sites for Sox2 and Otx2 in close proximity. Interestingly, the Otx2 site proximal to the Sox2 site overlaps a UniBind Pou5f1 TF-binding site, which, taken with the corresponding expression of these TFs and our footprint scores, provide further suggestion that

swapping of Sox2 binding partners is a plausible hypothesis for the activation of this regulatory element. Furthermore, we report that by exploring two publicly available Sox2 and Pou5f1 ChIP-seq datasets (Xiong et al., 2022; Avsec et al., 2021) generated from E14-mESCs (roughly equivalent to day 0 organoids), we found peaks of Pou5f1 and Sox2 signal at this CRE (Fig. S10C). In all, this suggests a mechanism whereby, although this element is accessible early in development, it is repressed by Pou5f1 binding and thus not able to activate *Rax* expression. As the expression of *Pou5f1* is downregulated and the expression of *Otx2* increases, Sox2 can exchange binding partners (from Pou5f1 to Otx2) and, in so doing, activate this putative enhancer and consequently also the expression of *Rax*.

Further experiments (in particular a CHIP-seq or Cut&Run time course for the relevant TFs) beyond the scope of our current study

would be required to validate this hypothesis. We do, however, find it both fascinating and of considerable value that specific hypotheses of gene regulation can be pieced together through systematic and integrated analysis of data across modalities, such as the one presented above. We also point out that, similar to the mechanics of regulation we hypothesize above, Delás et al. (2023) have recently provided evidence of a general mechanism whereby groups of genes are regulated via ‘differential binding’ of TFs at regulatory elements (rather than differential accessibility of these elements), in the context of neural progenitor differentiation.

DISCUSSION

By using an OV organoid model to enrich for cells transitioning to an optic fate, we have generated bulk transcriptomic and chromatin accessibility time-course data from a very early developmental event that is crucial for mammalian eye development. Joint computational analyses of these data have revealed three key messages regarding the molecular events controlling eye field specification. First, the transition to the eye field in mESC-derived OV organoids is characterized by the upregulation of a small set of genes, including a core set of TFs known to be necessary for eye development through studies in *Xenopus*. Second, the patterns extracted from our RNA-seq data indicate that when this state transition occurs, not only is this core set of EFTFs upregulated, but, in parallel, other gene expression programmes, such as those controlling cellular pluripotency and cellular signalling, are downregulated or carefully balanced. Third, analysis of the dynamics and sequence composition of chromatin accessibility peaks strongly suggests that the key EFTFs – particularly *Rax*, *Pax6* and *Lhx2* – act to control the upregulation of the EF genes through binding to (primarily) non-promoter regulatory elements in the vicinity (TADs) of those genes. By performing footprinting analyses for key TFs in TADs of important genes, we have additionally identified putative gene-specific CREs that appear to be important for regulating the gene expression changes characterizing the onset of the eye field. Furthermore, by perturbing a subset of these candidate regulatory elements using CRISPR and generating mutant cell lines, we have started to validate the regulatory potential of these elements in controlling EFTF expression in OV organoids.

In all, our study complements the early characterization of the eye field from *Xenopus*, extends this characterization to mammalian models and begins to shed light on the roles of key EFTFs in the transition of cells to the ocular fate. Additionally, our work demonstrates the power and practicality of using an organoid system to generate and test specific hypotheses regarding the regulation of cell-state transitions relevant to crucial events in mammalian development.

Although our data does not have the resolution to identify definitively what may initially trigger the expression of individual EFTFs, we have presented evidence to suggest that this may arise, at least in the case of *Rax*, from the replacement of repressive factors (such as *Pou5f1*) with activating factors (such as *Otx2*) at regulatory elements that are ubiquitously open across the time course. We have also outlined that disentangling the mechanisms behind gene-expression or cellular-state changes that are triggered through the action of TFs at regulatory elements can indeed be very challenging to understand through standard analyses correlating peak accessibility with gene expression. However, footprinting analyses do have the potential to provide richer insights into these more subtle mechanisms of regulation.

The approach to characterize the OV-organoid eye field presented in this work does have some limitations, which indicate areas for

future work. First, at each time point the organoids are likely to be composed of relatively heterogeneous collections of cells. As a result, the bulk RNA-seq and ATAC-seq datasets generated for this study are expected to be powerful enough to reveal the larger changes in gene expression programmes, such as the upregulation of the eye field programme or the suppression of pluripotency, but may be limited in their power to distinguish more subtle effects. To address the confounding issues introduced by cellular heterogeneity, an approach profiling the organoids at the single-cell level would be required. Second, although motif and footprinting analyses of our ATAC-seq data provide consistent evidence to build hypotheses regarding the putative regulatory roles and binding of important TFs, this evidence is indirect (the occurrence of a consensus TF motif within a region of enriched accessibility does not imply the binding of that TF at that genomic location). To gain more direct evidence of changes in TF binding it would be useful to perform time-course ChIP-seq or Cut&Run experiments for a set of TFs identified as important.

Finally, in undertaking the work presented here we have encountered a number of intriguing open questions regarding eye-field specification that require further study. In particular, it is pertinent to ask whether the eye field directly stems, in a sequential manner, from a group of well-defined progenitor cells within the neural plate, or whether individual neuroectodermal cells spontaneously acquire ocular fate and are only then organized into a coherent eye field through migration and cell-signalling events? Furthermore, are these processes similar *in vivo* and in the organoid system? Using a collection of markers Agnès et al. (2022) have recently quantified, in 3D, the emergence of the eye field in dimorphic teleosts and it would be informative to gain a similar spatial molecular map of the eye field in mammalian systems. Related to our observations that eye-field specification seems to require a careful balance of different, potentially competing gene expression programmes, an interesting avenue of further study would be to quantify the effects that perturbations of gene dosage have on this balance (particularly as dosage of key genes has been shown to affect stability of the eye field; Beccari et al., 2012; Bernard et al., 2014). To probe further the emergence of the eye field, it would be fascinating to investigate the extent to which the EFTFs act in a cell-autonomous manner, regulating the genes of the cells in which they are endogenously produced, or whether there are non-cell-autonomous components to gene regulation (paracrine signalling). Recent rapid technological progress in single-cell technologies – both multi-omics and imaging – will enable exciting future research in these directions. Finally, in order to bring this research closer to having translational potential, similar analyses must be performed using organoids derived from human ESCs; indeed, such a line of work has the potential to identify and test mechanisms crucial to early human eye development and disease that are currently impossible to study.

MATERIALS AND METHODS

ESC maintenance and organoid culture

Rx-GFP mouse ESCs were maintained and SFEBq (serum-free floating culture of embryoid body-like aggregates with quick aggregation) differentiation of OV organoids was performed as described previously (Wataya et al., 2008; Eiraku et al., 2011). Briefly, mESCs were dissociated and 4500 cells were plated in differentiation media [GMEM supplemented with 1.5% KOSR (KSR media) or a growth factor-free chemically defined medium (CDM media) for knockout organoids and initial organoids, respectively] in a low cell adhesion 96-well plate. This was defined as day 0. On day 1, growth factor-reduced Matrigel was added to a final concentration of 2%. For further details, see supplementary Materials and Methods.

Differentiation status was assessed using fluorescence imaging to detect the presence of GFP as a measure of *Rax* expression.

RNA-seq assays

On days 0-5 of optic vesicle differentiation, triplicate samples of cells were collected from 24 organoids, which were dissociated and sorted using FACS to select for live single cells. Samples at days 4 and 5 were also sorted on the basis of GFP expression giving a GFP-positive and -negative sample for each of these time points. The FACS gating placement was designed to maximize the number of cells in each sample, such that some cells with low GFP signal were included in the GFP-negative sample, and the GFP-positive samples contained cells with both high and moderate levels of GFP expression. Day 3 samples were analysed with FACS and found to contain a small proportion of GFP-expressing cells, but not enough to extract sufficient RNA. To obtain a more accurate view of the diverging cell types at this stage, all live cells from these organoids, regardless of their GFP expression were included in this sample. RNA was isolated using the Zymo Direct-zol RNA MicroPrep kit following the manufacturer's instructions. Quality and integrity of the RNA samples were analysed, libraries were prepared using the TruSeq kit (Illumina), and sequenced on the NextSeq 550 platform as per the manufacturer's protocols. For further details, see supplementary Materials and Methods.

RNA-seq analysis

Mapping reads and quantification

Coding regions of DNA and non-coding RNA from the mouse genome build 38 (mm10), downloaded from the Ensembl database, were used to create the transcriptome index. kallisto (Bray et al., 2016) was then used to pseudo-align RNA-seq reads for each library to these regions and quantify transcript abundance.

Differential expression analysis

Differential expression analysis between time points and expression normalization was performed using DESeq2 (Love et al., 2014). Specifically, the kallisto abundance estimates were used as inputs to DESeq2, using a design matrix of time points. The 'estimateSizeFactors' function was used to compute size factors for normalization, and the 'lfcShrink' function, with type='ashr' for log2-fold-change shrinkage (Stephens, 2016), was used to contrast time points. To select differentially expressed genes, we used an FDR threshold of 0.001 and an (absolute) log2-fold-change threshold of 1.5. To select genes of particular relevance to the transition between day 3 cells and cells on the RaxGFP-positive lineage (days 4 and 5), we performed differential expression analysis for the time points day 3/day 4GFP⁺, day 4GFP⁻/day 4GFP⁺, day 5GFP⁻/day 5GFP⁺. Genes found to be significantly upregulated across all these comparisons were termed 'EF-up' genes, whereas those significantly downregulated across these comparisons were termed 'EF-down' genes. We also used DESeq2 to select a set of 'housekeeping' or stably expressed genes, by requiring these genes to display absolute log2-fold-change <0.1 for every successive day comparison, and have a mean (across days) normalized expression >30 (this procedure identified 190 such stably expressed genes).

Clustering

To cluster gene expression trajectories, we used a Gaussian-mixture model (GMM) approach applied to *z*-transformed log2 values of normalized gene expression for EF-up and EF-down genes. In detail, we used the GMM implementation available in the Python scikit-learn library (Pedregosa et al., 2011), fitting GMMs on the trajectories for 20 random initializations. The number of clusters that best described the data was determined using the minimum value of the average Bayesian Information Criterion (BIC) across random seeds, and the cluster allocation of each gene was achieved by consensus clustering, using the ClusterEnsembles Python library.

GO enrichment analysis

We applied a GO enrichment analysis on the three sets of genes corresponding to the broad patterns discovered using clustering, using the Metascape online tool (Zhou et al., 2019; <https://metascape.org>).

ATAC-seq assays

On days 0-5 of optic vesicle differentiation, cells from 48 organoids were pooled and sorted using FACS, as described above for the RNA-seq. One replicate was generated for each time point. ATAC-seq sample preparation was performed as described by Buenrostro et al. (2015). Briefly, 50,000 cells were lysed, and the transposition reaction carried out using the Tn5 transposase enzyme. PCR amplification of samples was conducted using a universal forward primer and a unique barcoded reverse primer for each time point. Samples were then purified and size selected. Bioanalysis showed that the libraries contained the expected fragment size distribution (Buenrostro et al., 2013). ATAC-seq samples were paired-end sequenced on the Illumina Hi-seq platform, with day 4 samples run separately from the other time points. For further details, see supplementary Materials and Methods.

ATAC-seq analysis

Mapping reads, quantification and peak-calling

We used the FastQC tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and cutadapt (Martin, 2011) to perform quality control on and trim adapter sequences from the raw sequencing reads for all samples. Then we used bowtie2 (Langmead and Salzberg, 2012) to align trimmed reads to the mm10 genome. Duplicate reads, unmapped reads, and reads mapping to the mitochondrial genome were removed using samtools (Danecek et al., 2021). Filtered reads were shifted +4 bp for the positive strand and -5 bp for the negative strand to show the centre of the transposition site. bedtools (Quinlan and Hall, 2010) was used to generate the final bam and bed files used for downstream analysis. To enrich for fragments indicative of binding of TFs, we filtered out reads corresponding to fragment sizes >100 bp (i.e. retained only sub-nucleosomal reads for downstream analysis). To identify genomic regions enriched for accessible chromatin, we computationally called peaks using the callpeak function of macs2 (Zhang et al., 2008; <https://github.com/taoliu/MACS>), with parameters -f BAMPE -g mm --keep-dup all -B -SPMR -nomodel, on paired-end bam files of each sample (day) separately. Using the narrowPeak macs2 peak coordinates, we created a consensus set of peaks with which to study accessibility dynamics across days, using the bedtools merge command. From this consensus set of peaks, we removed peaks overlapping blacklisted regions of the mm10 genome, as listed on <http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/>. We used conditional quantile normalization (Hansen et al., 2012) to normalize the signal in consensus ATAC-seq peaks while correcting for the effects of peak width and GC content. Owing to the separate sequencing runs, day 4 samples showed considerable batch effects and were removed from subsequent analysis. Bedgraph files from macs2 were converted to BigWig files for visualization in the UCSC genome browser. Finally, we used ChIPseeker (Yu et al., 2015) to annotate peaks by their genomic context, collapsing peak annotations into four categories: promoter, intergenic, intronic and exonic.

Peaks by TAD regions

To assign peaks to TAD regions, we downloaded the coordinates for TAD regions from the work of Bonev et al. (2017; publicly available at <https://github.com/aertslab/mucistarget/tree/master/data/tads>). We noted that there are often gaps between genomically adjacent TAD regions, and in these cases we took these inter-TAD regions to also define relevant regions to restrict the search-space of putative peak-gene links (interestingly, for the TAD data we used, the *Rax* and *Six6* genes were contained within such inter-TAD regions). EF-up and EF-down TADs were defined as those TADs containing the genes determined to be differentially up- and downregulated, respectively, across the transition from day 3 to day 5. Housekeeping TADs were defined as TAD regions containing genes that are stably expressed across the full time course.

GREAT analysis

To perform a first putative linking of dynamic peaks to genes we employed the online GREAT tool (McLean et al., 2010; <http://great.stanford.edu>) to generate region-gene associations computationally. We used the coordinates of dynamic peak regions (regions of increasing and decreasing accessibility across the transition from day 3 to day 5) as test regions, and the whole genome as the background regions (default setting of GREAT).

Motif analysis of ATAC-seq dynamics

We obtained the DNA sequence of each consensus peak using bedtools getfasta and the mm10 genome. Following this, we used fimo with parameters `-text -max-strand -thresh 1e-4` (i.e. a P -value threshold of $1e^{-4}$), and we scanned each peak region for occurrences of all motifs in the JASPAR2020 database (*Mus musculus* TF motifs), using the presence/absence of motifs for further downstream analysis. We also restricted our attention to motifs of TFs that were expressed in the OV system, which we defined to be the mean (across RNA-seq replicates) DEseq2-normalized counts of a TF's expression >50 , at any point in the time course. To compute enrichment of motifs for our defined signal peak sets [peaks with increasing or decreasing accessibility signal above a log₂-fold-change of 1.5], we employed the hypergeometric test (Fisher test) using the set of peaks in TADs of housekeeping genes as a background.

Finally, to quantify the importance of each motif in a multivariate modelling approach, we fitted an L_2 -regularized logistic regression model to predict the binary behaviour of each peak (increasing or decreasing accessibility signal) using the presence of motifs corresponding to EF-up genes, EF-down genes, day2-day3 differentially expressed genes, as well as Sox2 and Otx2, as input features. Where there were multiple motifs for a TF, we collapsed the input feature into a single feature in which a TF motif was deemed to be present in a peak if that peak contained a significant occurrence of any of its possible motifs. We trained models on two sets of peaks: dynamic peaks [those with log₂-fold-change >1] within TADs containing EF-up genes and genome-wide dynamic peaks. We first used 5-fold cross-validation to determine an appropriate regularization parameter for each model, and then trained models using these optimal parameters on the full sets of peaks. In this second step, we fitted ten models with different random initial seeds, and took the mean for each coefficient across each run. To perform this modelling, we used the 'LogisticRegression' functions implemented in the sklearn Python library (Pedregosa et al., 2011).

We performed our *de novo* motif discovery analysis using streme (Bailey, 2021) and matched the inferred motifs to the consensus ones in the JASPAR2020 database using tomtom (Gupta et al., 2007). Results of these analyses applied on peak regions increasing and decreasing in accessibility across the transition from day 3 to day 5 can be found in Tables S5 and S6.

ChIP-seq evidence-backed sites of TF–DNA interactions were obtained from the UniBind database, (Puig et al., 2021; <https://umibind.uio.no/>), and specifically downloading data from the *Mus musculus* genome tracks.

Footprinting analysis

To quantify differences in putative TF binding between time points, we used the TOBIAS framework (Bentsen et al., 2020). We first used the bam files (with subnucleosomal reads) from each time point (days 0, 1, 2, 3 and 5) as input to TOBIAS-ATACcorrect to correct for Tn5-insertion bias, resulting in 'corrected' bigwig files of ATAC-seq signal. Together with the coordinates for the consensus peak set, the latter were used as input to TOBIAS-ScoreBigwig to compute base-pair footprint scores for these genomic regions, again in bigwig format. Finally, these footprint bigwig files were used as input to TOBIAS-BINDetect (in time-series mode, using the `-time-series flag`) to identify motif occurrences in the consensus peak set and compute associated footprint scores for each day, for all mouse motifs in the JASPAR2020 database (Fornes et al., 2020). This final step also quantified differences in putative binding of TFs across the full consensus peak set between each pair of successive days (used to produce the results of Fig. 4A). To generate the data for aggregate footprints (used for Fig. 4B and Fig. S4B), we used TOBIAS-PlotAggregate to average corrected ATAC-seq signal in ± 100 bp windows around occurrences of a given motif.

To quantify changes in aggregate footprints for a given TF motif, we first computed two metrics, 'flanking accessibility' and 'footprint depth' (closely following the approach of Baek et al., 2017), on the aggregate signal at each time point. For each motif, we took the ± 12 bp window around the motif centre as the 'central region' (putative region of binding), and first computed flanking accessibility as the mean aggregate signal in the 60 bp either side of the central region. Footprint depth was then computed as the mean aggregate signal within the 24-bp central region minus the flanking accessibility. Finally, we computed the difference between time points of

each of these metrics to obtain the change in flanking accessibility and of footprint depth.

CRISPR-Cas9-mediated disruption of candidate CREs

Mutations in the candidate regulatory elements within the Rax and Six6 TADs were introduced into the Rax-GFP cell line using CRISPR-Cas9. Guide RNAs, targeting the sites of TF-binding motifs within the identified peaks, were individually cloned into the pSpCas9(BB)-2A-GFP (PX458, Addgene plasmid #48138) vector. Rx-GFP cells were transfected with the resulting plasmid using Lipofectamine 3000 and cells selected based on GFP expression from the plasmid 48 h later. Cells were plated at low density, individual clones were expanded and resulting cell lines were assessed for deletions over the CREs in question using Sanger sequencing (Fig. S7). One cell line with a deletion encompassing motifs of interest along with one cell line without any deletions were maintained and used for optic vesicle organoid culture for each of the two selected putative CREs. At day 5, triplicate samples of 24 organoids were pooled, and RNA isolated as described for the RNA-seq assay. The NanoString nCounter analysis system (Geiss et al., 2008) was used to quantify RNA expression for a panel of 180 genes for which expression levels changed through the time course of differentiation. For further details, see supplementary Materials and Methods.

Plotting and visualization

Plots and visualizations were produced using the matplotlib Python (Hunter, 2007) library and Inkscape (<https://inkscape.org>).

Acknowledgements

We are grateful to Elisabeth Freyer and the IGC FACS facility, to the staff of the IGC Imaging facility for imaging support, to the IGC technical services team, to the genetics core at ECRF for sequencing, and to Alison Munro within the HTPU at the IGC for running the NanoString. We also thank Wendy Bickmore for her feedback on the manuscript.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: L.J.O., H.B., D.R.F., A.S.P.; Methodology: L.J.O., D.R.F., A.S.P.; Formal analysis: L.J.O., A.S.P.; Investigation: L.J.O., J.R., H.B., F.K.; Data curation: L.J.O., A.S.P.; Writing - original draft: L.J.O., A.S.P.; Supervision: D.R.F.; Funding acquisition: D.R.F.

Funding

L.J.O. was funded by a PhD studentship from the UK Medical Research Council. A.S.P. is a cross-disciplinary post-doctoral fellow supported by funding from the University of Edinburgh and Medical Research Council (MC_UU_00009/2). D.R.F. is supported by a Medical Research Council University Unit grant (MC_UU_00007/3). Open Access funding provided by the University of Edinburgh. Deposited in PMC for immediate release.

Data availability

Raw sequencing data generated in this work has been deposited in Gene Expression Omnibus under accession number GSE213273). We have explored Sox2 and Pou5f1 ChIP-seq and ChIP-nexus datasets that are publicly available in GEO (accession numbers GSE174774, GSM4072776 and GSM4072777). Analysis scripts and codes, as well as processed data are available at https://github.com/andrewpapa/optic_vesicle_eye_field_analysis.

Peer review history

The peer review history is available online at <https://journals.biologists.com/dev/lookup/doi/10.1242/dev.201432.reviewer-comments.pdf>.

References

- Agnès, F., Torres-Paz, J., Michel, P. and Rétaux, S. (2022). A 3D molecular map of the cavefish neural plate illuminates eye-field organization and its borders in vertebrates. *Development* **149**, dev199966. doi:10.1242/dev.199966
- Avsec, Ž., Weilert, M., Shrikumar, A., Krueger, S., Alexandari, A., Dalal, K., Fropf, R., McAnany, C., Gagneur, J., Kundaje, A. et al. (2021). Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366. doi:10.1038/s41588-021-00782-6

- Baek, S., Goldstein, I. and Hager, G. L. (2017). Bivariate genomic footprinting detects changes in transcription factor activity. *Cell Rep.* **19**, 1710-1722. doi:10.1016/j.celrep.2017.05.003
- Bailey, T. L. (2021). STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834-2840. doi:10.1093/bioinformatics/btab203
- Beccari, L., Conte, I., Cisneros, E. and Bovolenta, P. (2012). Sox2-mediated differential activation of Six3.2 contributes to forebrain patterning. *Development* **139**, 151-164. doi:10.1242/dev.067660
- Bernard, C., Kim, H.-T., Torero Ibad, R., Lee, E. J., Simonutti, M., Picaud, S., Acampora, D., Simeone, A., Di Nardo, A. A., Prochiantz, A. et al. (2014). Graded Otx2 activities demonstrate dose-sensitive eye and retina phenotypes. *Hum. Mol. Genet.* **23**, 1742-1753. doi:10.1093/hmg/ddt562
- Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T. et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat. Commun.* **11**, 4267. doi:10.1038/s41467-020-18035-1
- Berest, I., Arnold, C., Reyes-Palomares, A., Palla, G., Rasmussen, K. D., Giles, H., Bruch, P.-M., Huber, W., Dietrich, S., Helin, K. et al. (2019). Quantification of differential transcription factor activity and multiomics-based classification into activators and repressors: diffTF. *Cell Reports* **29**, 3147-3159. doi:10.1016/j.celrep.2019.10.106
- Bernier, G., Vukovich, W., Neidhardt, L., Herrmann, B. G. and Gruss, P. (2001). Isolation and characterization of a downstream target of Pax6 in the mammalian retinal primordium. *Development* **128**, 3987-3994. doi:10.1242/dev.128.20.3987
- Bilioni, A., Craig, G., Hill, C. and McNeill, H. (2005). Iroquois transcription factors recognize a unique motif to mediate transcriptional repression in vivo. *Proc. Natl Acad. Sci. USA* **102**, 14671-14676. doi:10.1073/pnas.0502480102
- Bray, N., Pimentel, H., Melsted, P. and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525-527. doi:10.1038/nbt.3519
- Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A. et al. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557-572. doi:10.1016/j.cell.2017.09.043
- Buecker, C. and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276-284. doi:10.1016/j.tig.2012.02.008
- 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. *Nat. Methods* **10**, 1213-1218. doi:10.1038/nmeth.2688
- Buenrostro, J. D., Wu, B., Chang, H. Y. and Greenleaf, W. J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109
- Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J. et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193-1203. doi:10.1038/ng.3646
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008. doi:10.1093/gigascience/giab008
- 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. *Proc. Natl Acad. Sci. USA* **105**, 5408-5413. doi:10.1073/pnas.0710954105
- Delás, M. J., Kalaitzis, C. M., Fawzi, T., Demuth, M., Zhang, I., Stuart, H. T., Costantini, E., Ivanovitch, K., Tanaka, E. M. and Briscoe, J. (2023). Developmental cell fate choice in neural tube progenitors employs two distinct cis-regulatory strategies. *Dev. Cell* **58**, 3-17. doi:10.1016/j.devcel.2022.11.016
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380. doi:10.1038/nature11082
- Dorsky, R. I., Itoh, M., Moon, R. T. and Chitnis, A. (2003). Two tcf3 genes cooperate to pattern the zebrafish brain. *Development* **130**, 1937-1947. doi:10.1242/dev.00402
- Eiraku, M., Takata, N., Ishibashi, H., Kawada, M., Sakakura, E., Okuda, S., Sekiguchi, K., Adachi, T. and Sasai, Y. (2011). Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51-56. doi:10.1038/nature09941
- Eiraku, M. and Sasai, Y. (2012). Mouse embryonic stem cell culture for generation of three-dimensional retinal and cortical tissues. *Nat. Protoc.* **7**, 69-79. doi:10.1038/nprot.2011.429
- Fish, M. B., Nakayama, T., Fisher, M., Hirsch, N., Cox, A., Reeder, R., Carruthers, S., Hall, A., Stemple, D. L. and Grainger, R. M. (2014). Xenopus mutant reveals necessity of rax for specifying the eye field which otherwise forms tissue with telencephalic and diencephalic character. *Dev. Biol.* **395**, 317-330. doi:10.1016/j.ydbio.2014.09.004
- Fitzpatrick, D. R. and van Heyningen, V. (2005). Developmental eye disorders. *Curr. Opin. Genet. Dev.* **15**, 348-353. doi:10.1016/j.gde.2005.04.013
- Furukawa, T., Kozak, C. A. and Cepko, C. L. (1997). rax, a novel paired-type homeobox gene, shows expression in the anterior neural fold and developing retina. *Proc. Natl Acad. Sci. USA* **94**, 3088-3093. doi:10.1073/pnas.94.7.3088
- Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghie, M., Baranašić, D. et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **48**, D87-D92.
- Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T. et al. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317-325. doi:10.1038/nbt1385
- Grant, C. E., Bailey, T. L. and Stafford Noble, W. (2011). FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **7**, 1017-1018. doi:10.1093/bioinformatics/btr064
- Grindley, J. C., Davidson, D. R. and Hill, R. E. (1995). The role of Pax-6 in eye and nasal development. *Development* **121**, 1433-1442. doi:10.1242/dev.121.5.1433
- Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. and Noble, W. S. (2007). Quantifying similarity between motifs. *Genome Biol.* **8**, 24. doi:10.1186/gb-2007-8-2-r24
- Hansen, K. D., Irizarry, R. A. and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Bioinformatics* **13**, 204-216. doi:10.1093/bioinformatics/kxr054
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90-95. doi:10.1109/MCSE.2007.55
- Kim, C.-H., Oda, T., Itoh, M., Jiang, D., Artinger, K. B., Chandrasekharappa, S. C., Driever, W. and Chitnis, A. B. (2000). Repressor activity of Headless/Tcf3 is essential for vertebrate head formation. *Nature* **407**, 913-917. doi:10.1038/35038097
- Klemm, S. L., Shipony, Z. and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207-220. doi:10.1038/s41576-018-0089-8
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. doi:10.1038/nmeth.1923
- Liu, W., Lagutin, O., Swindell, E., Jamrich, M. and Oliver, G. (2010). Neuroretina specification in mouse embryos requires Six3-mediated suppression of Wnt8b in the anterior neural plate. *J. Clin. Invest.* **120**, 3568-3577. doi:10.1172/JCI43219
- Long, H. K., Prescott, S. L. and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187. doi:10.1016/j.cell.2016.09.018
- Love, M. I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. doi:10.1186/s13059-014-0550-8
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10-12. doi:10.14806/embnet.17.1.200
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M. and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495-501. doi:10.1038/nbt.1630
- Minnoye, L., Marinov, G. K., Krausgruber, T., Pan, L., Marand, A. P., Secchia, S., Greenleaf, W. J., Furlong, E. E. M., Zhao, K., Schmitz, R. J. et al. (2021). GREAT improves functional interpretation of cis-regulatory regions. *Chromatin Accessibility Profiling Methods* **1**, 10. doi:10.1038/s43586-020-00008-9
- Nakano, T., Ando, S., Takata, N., Kawada, M., Mugaruma, K., Sekiguchi, K., Saito, K., Yonemura, S., Eiraku, M. and Sasai, Y. (2012). Self-formation of optic cups and storable stratified neural retina from human ESCs. *Cell Stem Cell* **10**, 771-785. doi:10.1016/j.stem.2012.05.009
- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N. L., Meisig, J., Sedat, J. et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385. doi:10.1038/nature11049
- Ong, C.-T. and Corces, V. G. (2012). Enhancers: emerging roles in cell fate specification. *EMBO Rep.* **13**, 423-430. doi:10.1038/embor.2012.52
- 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. doi:10.1186/s12864-021-07760-6
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825-2830.
- Quinlan, A. R. and Hall, I. R. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842. doi:10.1093/bioinformatics/btq033
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A. and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281-2308. doi:10.1038/nprot.2013.143
- Ray-Jones, H. and Spivakov, M. (2021). Transcriptional enhancers and their communication with gene promoters. *Cell. Mol. Life Sci.* **78**, 6453-6485. doi:10.1007/s00018-021-03903-w

- Reddington, J. P., Garfield, D. A., Sigalova, O. M., Karabacak Calviello, A., Marco-Ferreres, R., Girardot, C., Viales, R. R., Degner, J. F., Ohler, U. and Furlong, E. E. M. (2020). Lineage-resolved enhancer and promoter usage during a time course of embryogenesis. *Dev. Cell* **55**, 648-664. doi:10.1016/j.devcel.2020.10.009
- Roy, A., de Melo, J., Chaturvedi, D., Thein, T., Cabrera-Socorro, A., Houart, C., Meyer, G., Blackshaw, S. and Tole, S. (2013). LHX2 is necessary for the maintenance of optic identity and for the progression of optic morphogenesis. *J. Neurosci.* **33**, 6877-6884. doi:10.1523/JNEUROSCI.4216-12.2013
- Russell, R., Ilg, M., Lin, Q., Wu, G., Lechel, A., Bergmann, W., Eiseler, T., Linta, L., Kumar, P. P., Klingenstein, M. et al. (2015). A dynamic role of TBX3 in the pluripotency circuitry. *Stem Cell Reports* **5**, 1155-1170. doi:10.1016/j.stemcr.2015.11.003
- 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. *Mech. Dev.* **141**, 90-99. doi:10.1016/j.mod.2016.05.001
- Spitz, F. and Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613-626. doi:10.1038/nrg3207
- Stephens, M. (2016). False discovery rates: a new deal. *Biostatistics* **18**, 275-294.
- Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W. and Spitz, F. (2016). The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39**, 529-543. doi:10.1016/j.devcel.2016.10.015
- Takata, N., Abbey, D., Fiore, L., Acosta, S., Feng, R., Gil, H. J., Lavado, A., Geng, X., Interiano, A., Neale, G. et al. (2017). An eye organoid approach identifies Six3 suppression of R-spondin 2 as a critical step in mouse neuroretina differentiation. *Cell Reports* **21**, 1534-1549. doi:10.1016/j.celrep.2017.10.041
- Tétreault, N., Champagne, M.-P. and Bernier, G. (2009). The LIM homeobox transcription factor Lhx2 is required to specify the retina field and synergistically cooperates with Pax6 for Six6 trans-activation. *Dev. Biol.* **327**, 541-550. doi:10.1016/j.ydbio.2008.12.022
- Thomson, M., Liu, S. J., Zou, L.-N., Smith, Z., Meissner, A. and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell* **145**, 875-889. doi:10.1016/j.cell.2011.05.017
- Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. and Jaynes, J. B. (1998). Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol. Cell. Biol.* **18**, 2804-2814. doi:10.1128/MCB.18.5.2804
- Wang, Z., Oron, E., Nelson, B., Razis, S. and Ivanova, N. (2012). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* **10**, 440-454. doi:10.1016/j.stem.2012.02.016
- 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. *Proc. Natl Acad. Sci. USA* **105**, 11796-11801. doi:10.1073/pnas.0803078105
- Whissell, G., Montagni, E., Martinelli, P., Hernando-Mombalona, X., Sevillano, M., Jung, P., Cortina, C., Calon, A., Abuli, A., Castells, A. et al. (2014). The transcription factor GATA6 enables self-renewal of colon adenoma stem cells by repressing BMP gene expression. *Nat. Cell Biol.* **7**, 695-707. doi:10.1038/ncb2992
- Xiong, L., Tolen, E. A., Choi, J., Velychko, S., Caizzi, L., Velychko, T., Adachi, K., MacCarthy, C. M., Lidschreiber, M., Cramer, P. et al. (2022). Oct4 differentially regulates chromatin opening and enhancer transcription in pluripotent stem cells. *eLife* **11**, e71533. doi:10.7554/eLife.71533
- Young, R. M., Hawkins, T. A., Cavodeassi, F., Stickney, H. L., Schwarz, Q., Lawrence, L. M., Wierzbicki, C., Cheng, B. Y. L., Luo, J., Mayela Ambrosio, E. et al. (2019). Compensatory growth renders Tcf7l1a dispensable for eye formation despite its requirement in eye field specification. *eLife* **8**, e40093. doi:10.7554/eLife.40093
- Yu, G., Wang, L.-G. and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383. doi:10.1093/bioinformatics/btv145
- Yun, S., Saijoh, Y., Hirokawa, K. E., Kopinke, D., Murtaugh, L. C., Monuki, E. S. and Levine, E. M. (2009). Lhx2 links the intrinsic and extrinsic factors that control optic cup formation. *Development* **136**, 3895-3906. doi:10.1242/dev.041202
- Zhang, L., Mathers, P. H. and Jamrich, M. (2000). Function of Rx, but not Pax6, is essential for the formation of retinal progenitor cells in mice. *Genesis* **28**, 135-142. doi:10.1002/1526-968X(200011/12)28:3/4<135::AID-GENE70>3.0.CO;2-P
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoutte, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W. et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. doi:10.1186/gb-2008-9-9-r137
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C. and Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523. doi:10.1038/s41467-019-09234-6
- Zuber, M. E., Gestri, G., Viczian, A. S., G., B. and Harris, W. A. (2003). Specification of the vertebrate eye by a network of eye field transcription factors. *Development* **130**, 5155-5167. doi:10.1242/dev.00723
- Zuber, M. E. (2010). Invertebrate and vertebrate eye development, chapter two – eye field specification in *Xenopus laevis*. *Curr. Top. Dev. Biol.* **93**, 29-60. doi:10.1016/B978-0-12-385044-7.00002-3

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 μ 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 μ 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 μ M 2-Mercaptoethanol and 1.5% KOSR. CDM differentiation media was made up of Iscove's Modified Dulbecco's Medium (IMDM) GlutaMAXTM Supplement (Thermo, 31980022) and Ham's F12 Nutrient Mix, GlutaMAXTM 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 μ 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% CO₂ 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 μ 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 μ l of cell suspension was pelleted, resuspended in 100 μ l and centrifuged again before the supernatant was removed. Cells were then resuspended in 100 μ l of cold ATAC lysis buffer (10 mM TrisCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 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 μ l ChIPmentation buffer (10 mM Tris pH8, 5 mM MgCl₂, 10% Dimethylformamide) and 2.5 μ 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 μ l elution buffer.

PCR Amplification. PCR amplification was conducted in a 32 μ l reaction volume for each sample, made up of 10 μ l transposed DNA, 4.7 μ l of 10 μ M custom barcoded nextera primer mix consisting of universal forward primer and one of the barcoded reverse primers, 1.3 μ l 50X SYBR Green (Invitrogen, S7563) and 16 μ 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 μ l of PCR reaction was made up to 50 μ l with nuclease free H₂O.

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 μ 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 μ l 80% ethanol for 30 seconds. The samples were left to air dry for 10 minutes. 20 μ 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 α cells and purified from liquid culture using the QIAprep Spin Miniprep Kit as per manufacturer's instructions.

Rx-GFP mESCs were transfected with 2 μ g of plasmid DNA diluted in Opti-MEM (ThermoFisher, 31985062) and 6 μ 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 μ l of Hybridisation Buffer was added to each vial of the Reporter CodeSet, 8 μ l of this was added to each of the hybridisation tubes. 5 μ l of 20 ng/ μ l RNA (100 ng total) was added to the appropriate hybridisation tube, followed by 2 μ 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

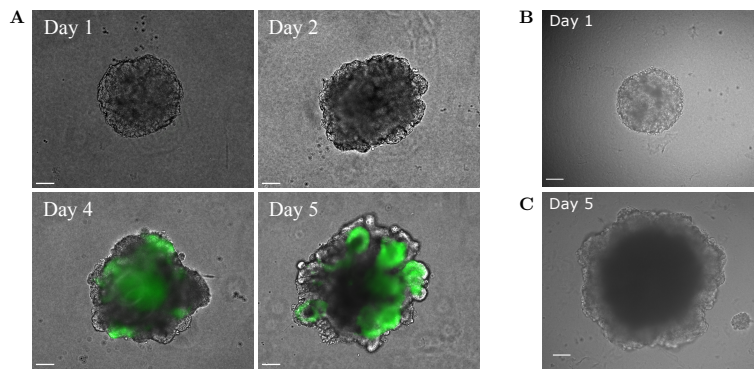


Fig. S1. Organoid culture controls for OV timecourse. A. Representative images of organoids cultured in KSR media at days 1, 2, 4 and 5 of growth. The same cell line was used to generate these organoids as was used to grow the organoids in Fig.1A. B. Representative image of organoid cultured in maintenance 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 μ m.

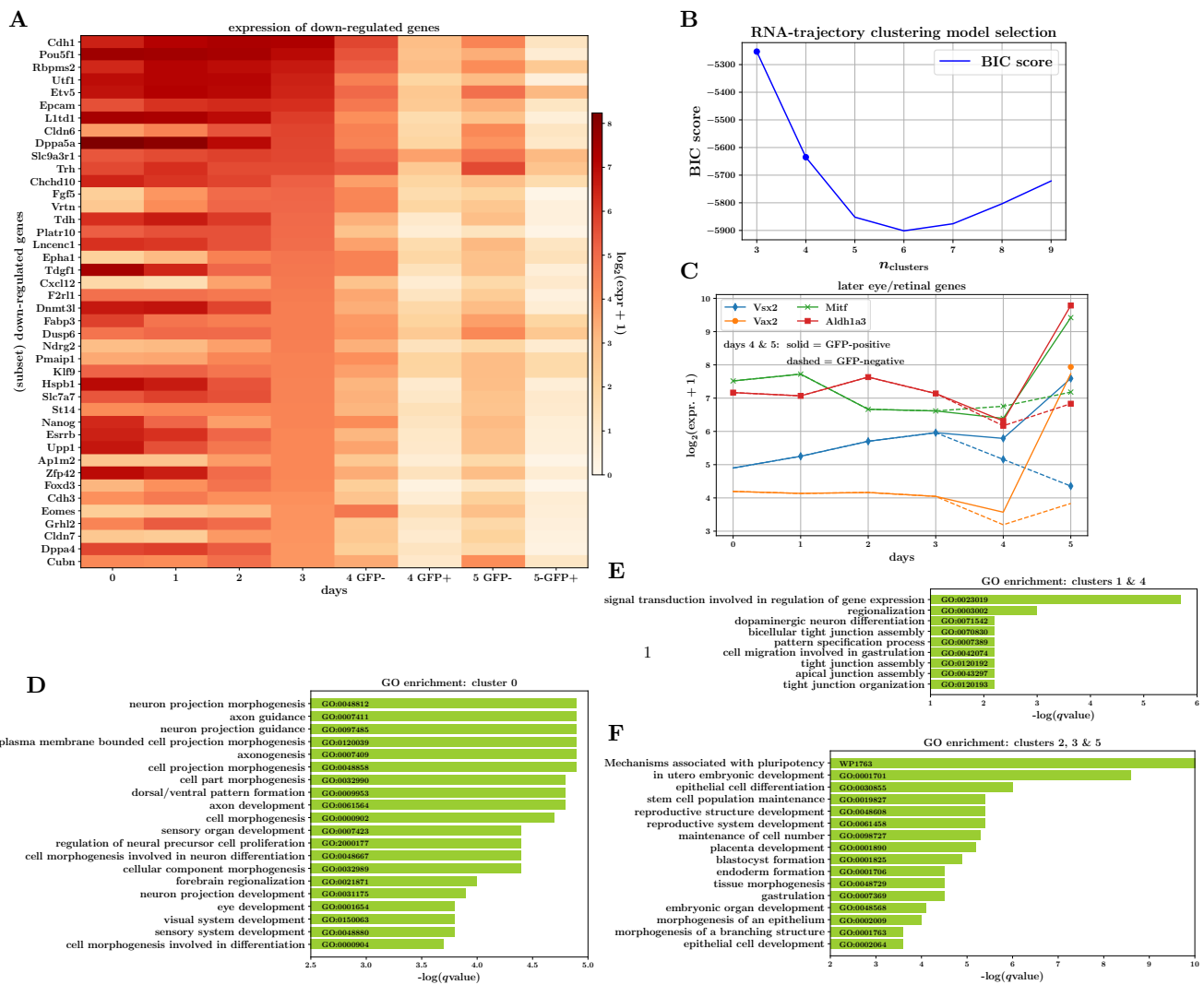


Fig. S2. Down-regulated genes indicate pluripotency and other gene-expression programs are suppressed during eye field specification. A. Expression heatmap displaying expression changes across the OV timecourse, for a selection of genes down-regulated across day3-day4/5 transition. B. BIC model selection 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.

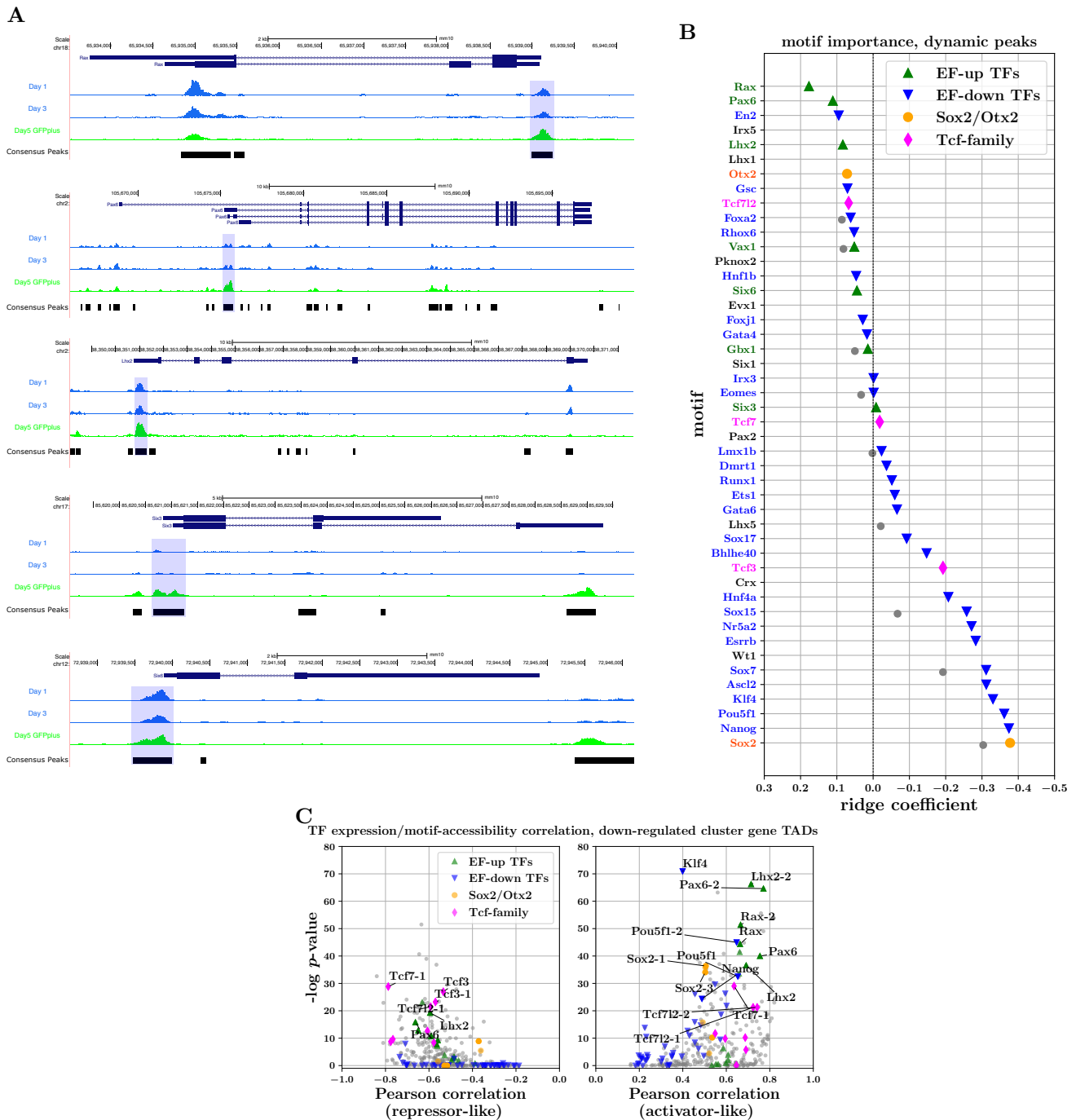


Fig. S3. Promoters of canonical eye field TFs are accessible from early in OV timecourse. Motif importance for genome-wide dynamic peaks mirrors importance in EF TADs. A. Bigwig tracks displaying changes in accessibility for regions around canonical EFTFs across the OV development timecourse. Promoters 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.

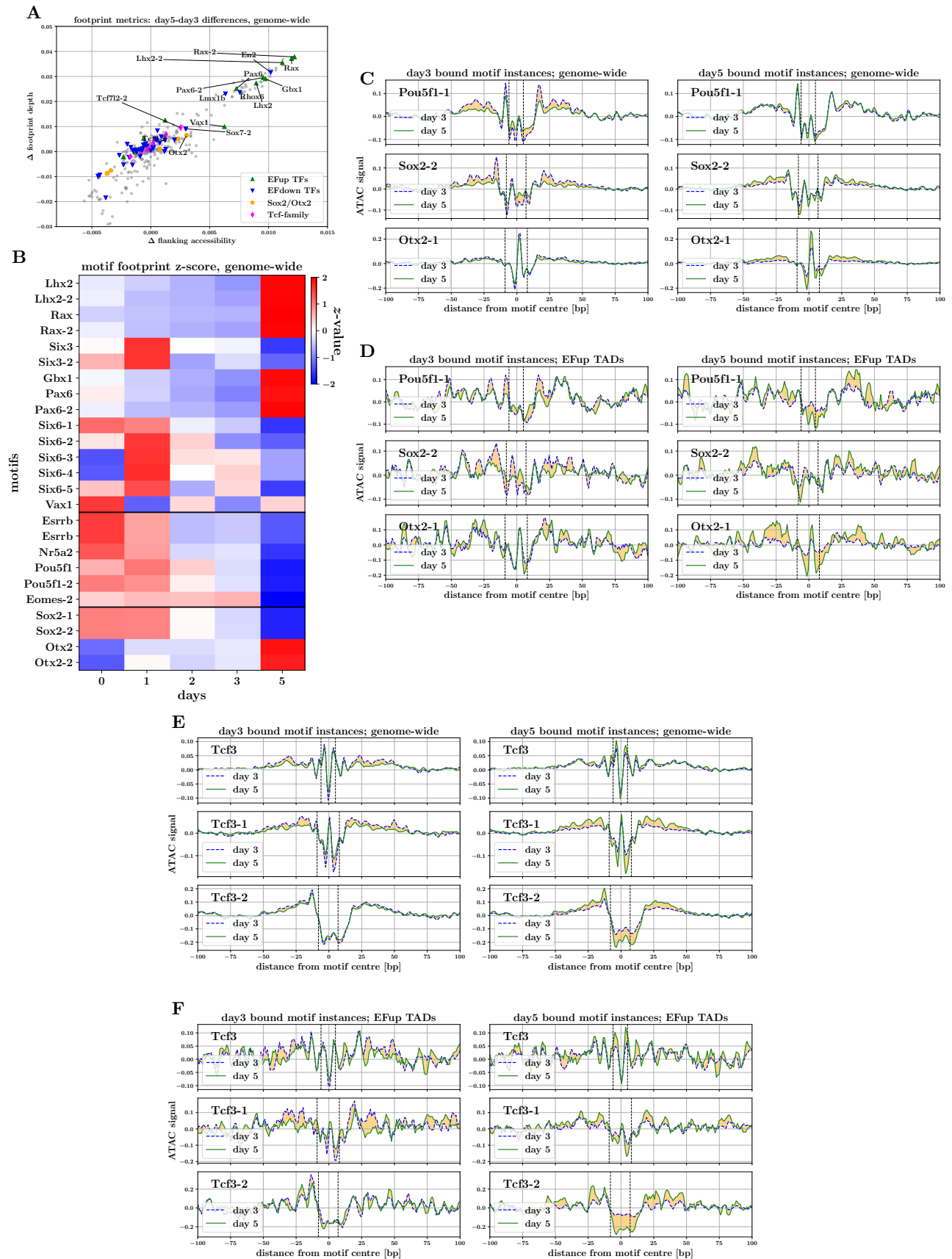


Fig. S4. Footprinting patterns genome-wide. **A.** Scatterplot of day3-day5 differences in TF-motif footprint depth versus flanking accessibility, for peaks genome-wide. **B.** Heatmap of TF-motif footprint scores in peaks genome-wide, z-transformed across time-course (median across peaks). **C. & D.** 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.

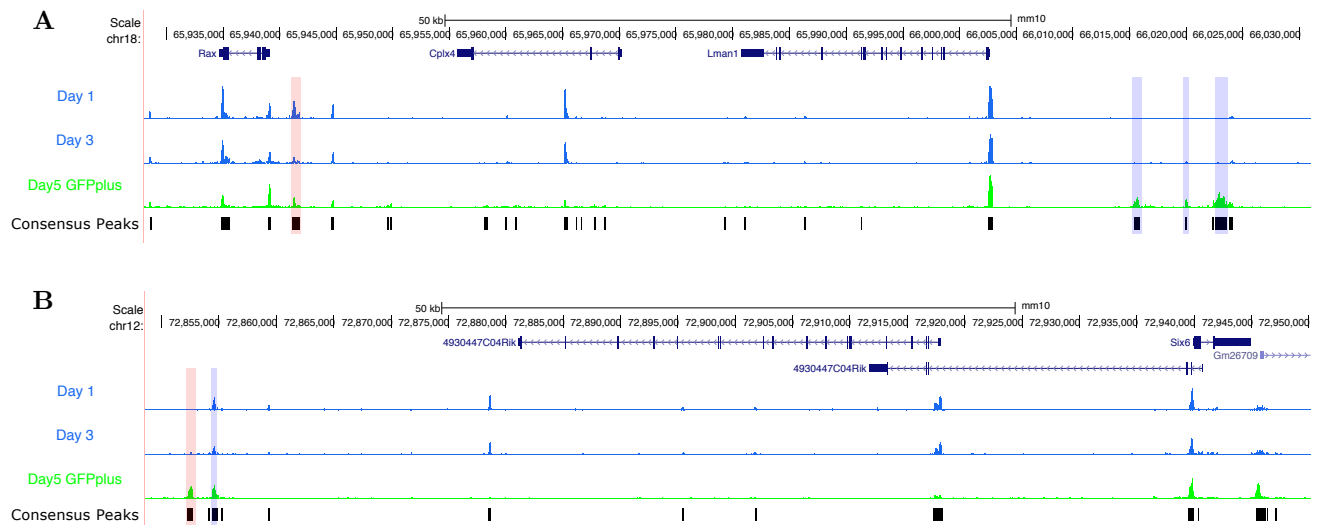


Fig. S5. Accessibility changes around candidate *Rax* and *Six6* enhancer elements. A. & B. Changes in ATAC-seq signal around the *Rax* and *Six6* loci, with consensus peaks illustrated by black bars. Peaks showing interesting footprint score changes mentioned in text are highlighted in blue, and the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

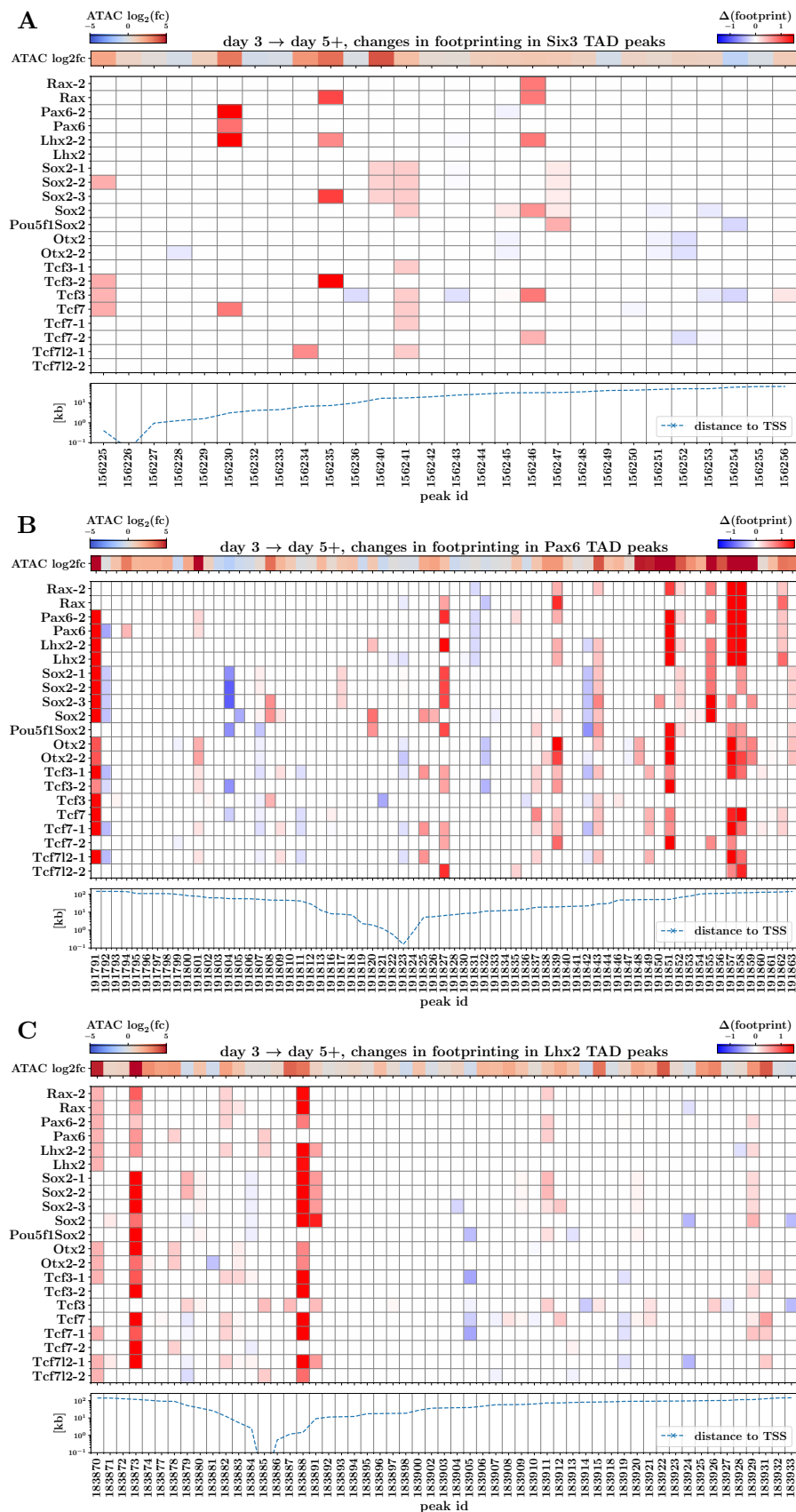


Fig. S6. Changes in motif footprint scores of key TFs identify candidate *Pax6*, *Lhx2* and *Six3* cis-regulatory elements. A. Heatmap of day3-to-day5 changes in footprint scores for motifs of key TFs at peaks contained in *Six3* TAD region. Also illustrated are associated changes in normalized ATAC-seq 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.

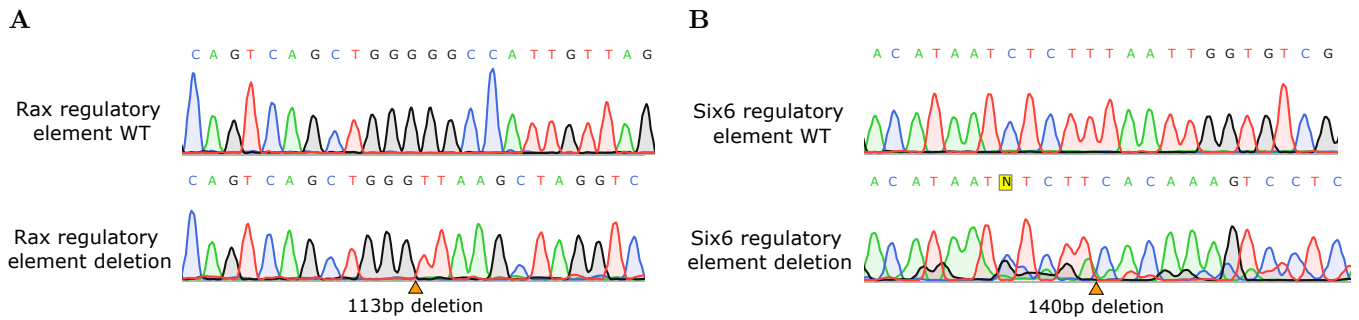


Fig. S7. DNA sequencing analysis of CRISPR targeted regions. Sanger sequencing traces of the region around the CRISPR-Cas9 guide target site in the **A. *Rax*** and **B. *Six6*** predicted enhancer regions. The sequence of the WT cell line used as the unedited control is shown in the top panel, and the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

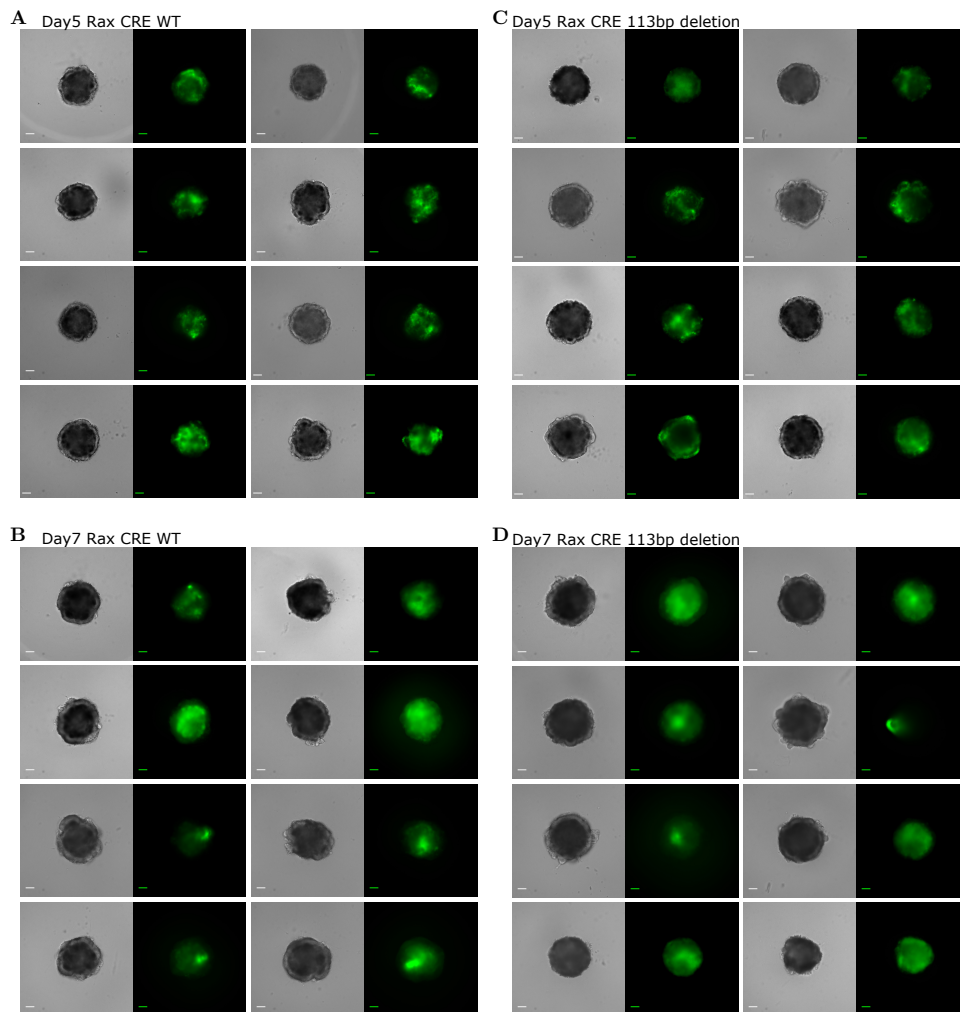


Fig. S8. Organoids derived from candidate *Rax* enhancer CRISPR-disrupted cell lines. **A. & B.** Organoids derived from wildtype *Rax* enhancer CRISPR cell-line, at days 5 and 7. 3 replicates of 24 organoids were grown for each cell line, to minimise the stress the organoids were subjected to, 8 were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

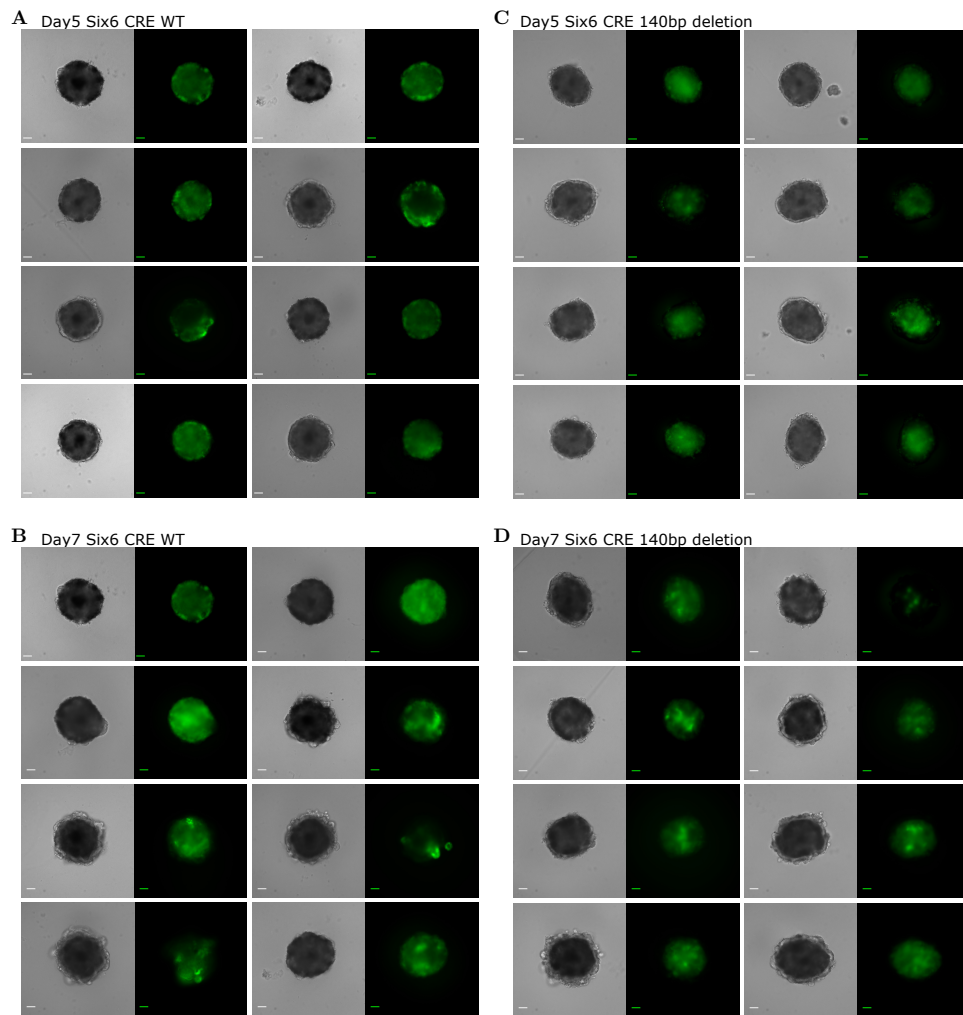


Fig. S9. Organoids derived from candidate *Six6* enhancer CRISPR-disrupted cell lines. A. & B. Organoids derived from wildtype *Six6* enhancer CRISPR cell-line, at days 5 and 7. 3 replicates of 24 organoids were grown for each cell line, to minimise the stress the organoids were subjected to 8 were imaged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

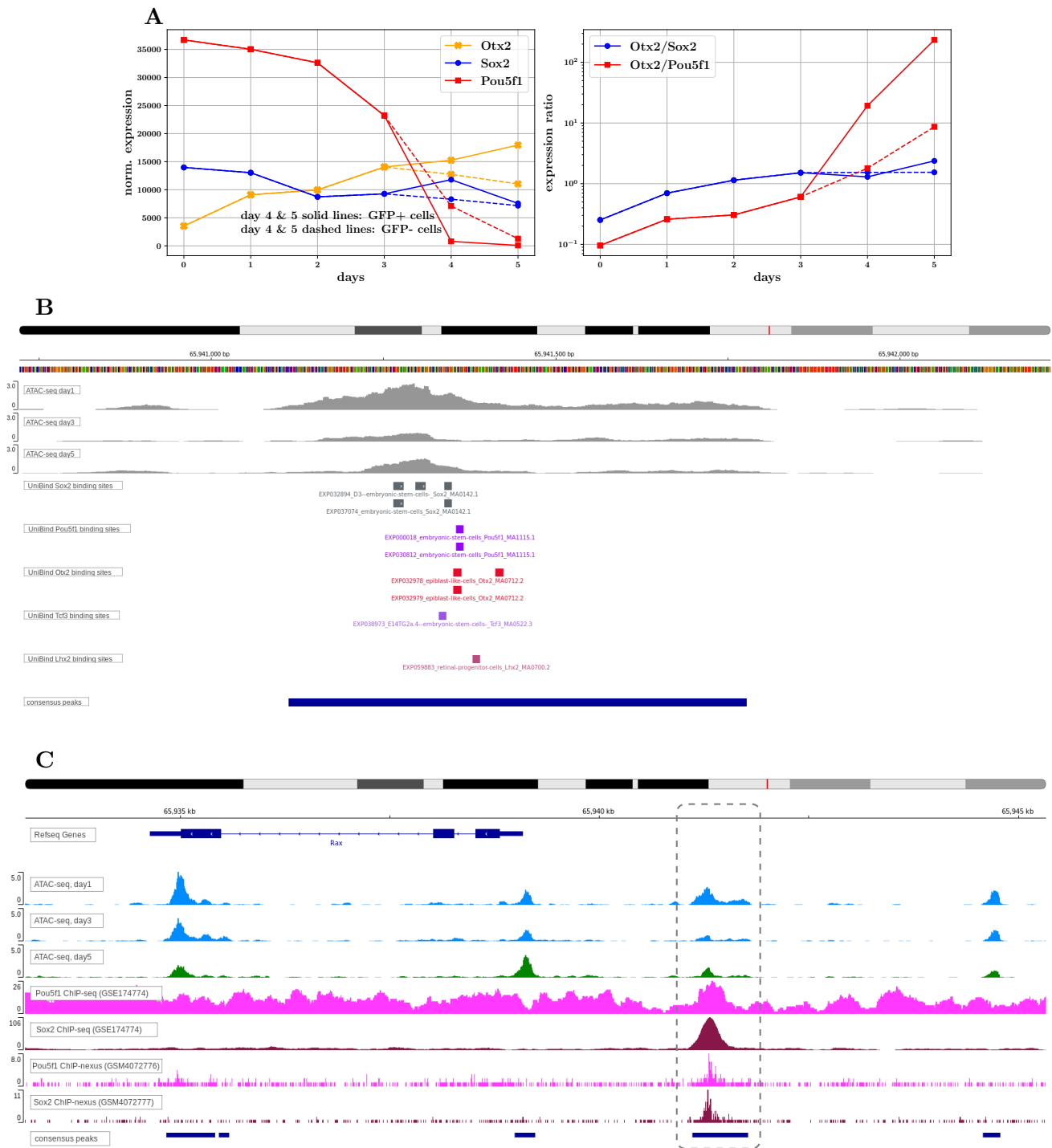


Fig. S10. ATAC-seq, ChIP-seq and RNA-seq data lead to hypothesis of *Rax* element activation. A. Expression (left) and expression-ratios (right) for key TFs hypothesized to play a role in the activation of the proximal *Rax* element identified by [Danno et al. 2008](#), and perturbed in the OV-organoid 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

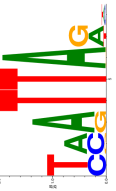

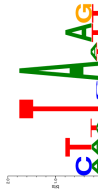
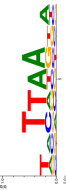
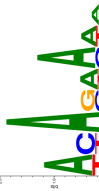

streme motif	visualization	e-value (streme)	similar motifs & p-values (tomt.com)
1-TMATTARH		3.0×10^{-34}	Lhx2-2 (1.8×10^{-5}), Rax (2.1×10^{-5}), Akl1 (4.1×10^{-5}), Alx1-1 (4.1×10^{-5}), Esx1 (4.8×10^{-5}), Alx3 (4.8×10^{-5}), En2 (4.8×10^{-5}), En1 (6.2×10^{-5}), Phox2a (6.2×10^{-5}), Lhx2 (6.6×10^{-5}), Prrx1 (6.6×10^{-5}), Pax7 (6.6×10^{-5}), Pou3f2 (7.7×10^{-5}), Nkx1-2 (8.9×10^{-5}), Emx2 (8.9×10^{-5}), Vax2 (1.1×10^{-4}), Lhx6-1 (1.4×10^{-4}), Nkx6-1 (1.6×10^{-4}), Arx (1.6×10^{-4}), Arx (1.6×10^{-4}), Msx2 (1.8×10^{-4}), Alx1-2 (1.8×10^{-4}), Dlx3 (1.8×10^{-4}), Pou2f1 (1.9×10^{-4}), Evx1 (2.4×10^{-4}), Prrx2 (2.7×10^{-4}), Rhox6 (2.7×10^{-4}), Vax1 (2.9×10^{-4}), Rax-2 (3.1×10^{-4}), Lhx6-2 (3.4×10^{-4}), Msx1 (3.7×10^{-4}), Pax6 (3.7×10^{-4}), Prrx2 (3.8×10^{-4}), Gbx1 (3.9×10^{-4}), Nkx6-1-2 (4.1×10^{-4}), Pax6-2 (4.2×10^{-4}), Lhx1 (4.4×10^{-4}), Prrx2 (4.8×10^{-4}), Gbx2 (5.5×10^{-4}), Dlx3 (6.0×10^{-4}), Pou6f1-1 (7.7×10^{-4}), Lmx1b (8.6×10^{-4}), Nkx6-1-1 (1.1×10^{-3}), Isl1 (1.1×10^{-3}), Atrd3b (1.1×10^{-3}), Pou3f4 (1.6×10^{-3}), Lhx5 (1.8×10^{-3}), Atrd3a-1 (3.7×10^{-3}), Pou6f1-2 (3.7×10^{-3}), Pou3f1 (4.5×10^{-3}), Barhl2 (6.0×10^{-3}), Hmx1 (9.6×10^{-3})
2-ATGAACWMAVWDRAA		2.4×10^{-23}	Mtf1-2 (2.8×10^{-4}), Foxj3 (5.0×10^{-4}), Foxj3-1 (5.0×10^{-4}), Lmx1b (7.8×10^{-4}), Bbx-1 (9.7×10^{-4}), Foxa2-2 (1.1×10^{-3}), Six4 (1.1×10^{-3}), Glis2-2 (3.8×10^{-3}), Sox7-1 (4.1×10^{-3}), Sox7-1 (4.1×10^{-3}), Tef3-2 (4.5×10^{-3}), Atrd3a-1 (6.2×10^{-3}), Srf-2 (7.0×10^{-3}), Sox4-2 (7.0×10^{-3}), Foxk1 (7.0×10^{-3}), Foxk1-1 (7.0×10^{-3}), Dmbx1 (7.3×10^{-3}), Dmbx1-1 (9.6×10^{-3}), Mybl1-1 (9.7×10^{-3})
3-CTWTSAWAG		1.0×10^{-8}	No significant JASPAR motif matches.
4-TVTTAABA		2.3×10^{-5}	No significant JASPAR motif matches.
5-ACASAAWW		6.4×10^{-4}	No significant JASPAR motif matches.
6-ATNSATCA		1.9×10^{-2}	Tcf3-1 (2.2×10^{-4}), Tcf7-1 (2.5×10^{-4}), Pbx1 (3.4×10^{-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		7.6×10^{-4}	No significant JASPAR motif matches.
2-CCATATCCTGCA		1.5×10^{-3}	No significant JASPAR motif matches.
3-CTGCAGGTGACK		1.5×10^{-3}	Arnt1 (1.8×10^{-4}), Bhlhe40 (9.3×10^{-4}), Arnt (1.1×10^{-4}), Tef3 (1.3×10^{-3}), Zic1::Zic2 (1.4×10^{-3}), Mycn (1.7×10^{-3}), Aif3 (3.4×10^{-3}), Mif1 (4.0×10^{-3}), Eomes-2 (4.6×10^{-3}), Zic1-2 (4.7×10^{-3}), Zic2-2 (4.7×10^{-3}), Zic2 (4.8×10^{-3}), Mixp (6.0×10^{-3}), Mycn (6.4×10^{-3}), Ase12 (6.5×10^{-3}), NR4A2 (6.9×10^{-3}), Rang(var.2) (7.7×10^{-3}), Creb5 (8.1×10^{-3}), FOS::JUN (8.1×10^{-3}), CREB1 (8.8×10^{-3}), Srebfl(var.2) (9.1×10^{-3}), Tef12 (9.7×10^{-3})
4-GAGGAAGGCTCCAC		1.5×10^{-3}	No significant JASPAR motif matches.
5-AACAATAGAGTGA		2.5×10^{-3}	Sox5 (2.1×10^{-5}), Sox15-1 (3.3×10^{-4}), Sox7-1 (4.0×10^{-4}), Sox5-1 (7.1×10^{-4}), FOXO3 (8.5×10^{-4}), Sox8-1 (8.7×10^{-4}), Foxj2 (1.0×10^{-3}), Sox12-1 (1.0×10^{-3}), Sox21-2 (1.6×10^{-3}), Sox13-1 (1.6×10^{-3}), Nanog (2.4×10^{-3}), Sox17-1 (3.4×10^{-3}), Sox2 (4.6×10^{-3}), Sox11-1 (4.6×10^{-3}), Sox4-1 (6.1×10^{-3})
6-GAGATTACYTA		3.8×10^{-3}	No significant JASPAR motif matches.
7-ATTCAITCCW		2.6×10^{-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](#)