DISSERTATION | DOCTORAL THESIS

Titel | Title

Enhancer cooperativity in gene regulation over large genomic distances

verfasst von | submitted by

Songjie Feng

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 794 620 490
Dissertationsgebiet lt. Studienblatt | Field of study as it appears on the student record sheet:	Molekulare Biologie
Betreut von | Supervisor:	Assoz. Prof. Christa Bücker PhD

Acknowledgements

I would like to express my gratitude to Christa Buecker for this exciting project, which has been a wonderful journey from beginning to end. Your support and advice are instrumental in shaping both the project and my own growth as a researcher.

I would also like to thank, danken, ringraziare, agradecer, and tacka all the members of the lab. The lab atmosphere is the most inspiring I have ever known, and the memories from this time will always remain precious to me. I would like to thank Henry. You are not only a fantastic supervisor but also a very close friend. It has been a great pleasure working with you and moving the project forward together. Thank you for organizing so many movie nights. These were wonderful memories beyond the lab. I also want to thank Swathi, my deskmate. I still remember the legendary shared snack drawer, a true lifeline with its endless supply of emergency food. The Indian music in TC will last forever in my memory. Isa, mother of Luna and Nala, always brought joy to the lab. Sharing cat photos became part of our daily routine, and cat topics were a constant highlight during lab lunchtime. Thank you for organizing the Halloween party. It was a wonderful lab tradition that I will never forget. Felix, working with you on developing our new system has been such a rewarding time. Your sharp scientific insight and genuine passion have truly inspired me. Lab would be different without you. Maria, having you in my life is truly a blessing, and I am deeply grateful for it. It is such a rare coincidence that we first interned in the same lab and then ended up working together in another one. For my postdoc, I will make sure to choose the lab wisely, just in case you decide to join as well. I would also like to thank Tamara, the goddess of plants and gastruloids. Beyond being an excellent scientist, you have been a wonderful and deeply trustworthy friend. The trip to Italy is a memory I will always treasure. Many thanks to Mattias, the king of CUT&Tag. You brought not only fancy new techniques but also a fresh and critical perspective to the lab. Thank you, Tara and Florian, the rising stars of the enhancer team. Although you joined the lab only for a short time, you brought so much joy and happiness to it. Last but not least, I am deeply grateful to Marie for your dedication to lab management, which demanded so much time and energy. Thank you also for handling all the orders of reagents and consumables, the lab could not have run without you. Thank you also for your support and help in my project. I really appreciate it.

I would also like to thank Martin Leeb and all the members of the Leeb lab — Anna, Luis, Laura, Mattia, Michelle, Michi B., Michi O., Moritz, and Tania. I am very grateful for Martin's and the lab's constructive input on my project during our joint meetings. The Christmas market and the lab retreat were also unforgettable moments that made me feel truly welcomed as part of the team. I will always cherish those experiences together with you. Special thanks to Luis, who is like a brother. Your support and thoughtful advice in both scientific work and everyday life have greatly helped me throughout this journey.

I would also like to thank my TAC members, Alexander Stark and Ulrich Elling. Your constructive input during each TAC meeting has been instrumental in shaping the project and ultimately bringing it to completion. I am deeply grateful for the time, expertise, and guidance you generously shared throughout this process. Special thanks to the FACS facilities of the Max Perutz labs. Thank you, Kitti Csalyi, Thomas Saurer, and Endre Kiss. I still remember the many times I had to call you for help outside of working hours, and I am truly grateful for your patience and support.

Many thanks to my dear friend and former lab colleague, Hana Marvanová. You were my very first friend in Vienna, and I cannot imagine my PhD journey without your support. I am so lucky and grateful to have you in my life. I would also like to thank my friends in Vienna: Weiqiang Guo, Yiying Ding, Kangli Guo, Wentao Lee, Tuoyu Song, Zhou Ren, Giovannie Sestini, and flatmate Karim. Thank you all for the friendship and support during the PhD journey. I also would like to thank my friends in other countries: Hao Kang, Ziming Wang, Xuexia Xu, Wen Jiang, Ming Xing and He Wang.

At last, I would like to thank my parents, my sister and her daughter. Thank you all for your trust and and support. I also give my special thanks to Jasper and Lylla. Their enthusiastic greetings after work never failed to cheer me up. They added joy, laughter, and love to my PhD journey.

Table of Content

Abstract

Enhancers are a class of DNA sequences that regulate gene transcription in a spatial-temporal manner. Recent studies show that enhancer numbers exceed genes numbers in the genome. In many genomic regions, enhancers are clustered to regulate the expression of target genes. Nevertheless, the mechanisms by which individual enhancer cooperates within such clusters to modulate gene expression are still not well understood. Manipulating enhancers in their native genomic context is labor-intensive and not well suited for high-throughput studies. In particular, for cis-regulatory elements with weak enhancer activity, conventional loss-of-function approaches are often insufficient to fully capture their function within endogenous regions.

In this thesis, I generated a cell line with a synthetic locus that enables efficient integration of enhancer sequences at three defined distances upstream of a fluorescent reporter gene, thereby allowing systematic reconstruction of regulatory landscapes from the bottom up. With this platform, I inserted multiple enhancers individually at defined distances relative to the same promoter. I found that all tested enhancers exhibited reduced promoter-activating capacity when the enhancer–promoter distance increased. However, the magnitude of this effect varied among enhancers and depended on their intrinsic enhancer strength.

To gain deeper insight into enhancer cooperativity in gene regulation, I introduced enhancer pairs into the synthetic locus, with each enhancer placed at a defined distance relative to the promoter. The results show that when weak and strong enhancers were combined, expression levels were markedly elevated if the weak enhancer was positioned between the strong enhancer and the promoter, indicating that enhancer synergy depends both on intrinsic enhancer strength and on relative genomic positioning. CTCF-binding sites and other structural elements are known to modulate enhancer–promoter interactions. To evaluate their contribution to enhancer cooperativity in our synthetic system, I removed two CTCF-associated regions in multiple cell lines harboring synergistic enhancer pairs. Deletion of the CTCF regions led to a further increase in reporter gene expression. This indicates that enhancer synergy is affected by CTCF sites.

Taken together, these findings summarized in this thesis highlight that transcriptional output is not governed by a single factor but rather emerges from the interplay between enhancer cooperativity and structural elements.

Zusammenfassung

Enhancer sind eine Klasse von DNA-Sequenzen, die die Genexpression in räumlich-zeitlicher Weise regulieren. Jüngste Studien zeigen, dass die Anzahl der Enhancer im Genom die Zahl der Gene deutlich übersteigt. In vielen Genomregionen sind Enhancer zu Clustern angeordnet, die gemeinsam die Expression ihrer Zielgene steuern. Dennoch sind die Mechanismen, durch die einzelne Enhancer innerhalb solcher Cluster kooperieren, um die Genexpression zu modulieren, bislang nur unzureichend verstanden. Die Manipulation von Enhancern in ihrem nativen genomischen Kontext ist arbeitsintensiv und eignet sich für High-Throughput-Studien kaum. Insbesondere für cis-Elemente mit schwacher Enhancer-Aktivität reichen herkömmliche Loss-of-Function-Ansätze oft nicht aus, um ihre Funktion in endogenen Regionen vollständig zu erfassen.

In dieser Dissertation habe ich eine Zelllinie mit einem synthetischen Lokus etabliert, der die effiziente Integration von Enhancer-Sequenzen in drei definierten Abständen zu einem fluoreszenten Reportergens ermöglicht. Auf diese Weise lassen sich regulatorische Landschaften systematisch von Grund auf rekonstruieren. Mit diesem System habe ich mehrere Enhancer einzeln in definierten Abständen zu demselben Promotor integriert. Dabei zeigte sich, dass alle getesteten Enhancer mit zunehmender Distanz zum Promotor eine verringerte Aktivierungskapazität aufwiesen, wobei das Ausmaß dieses Effekts je nach intrinsischer Stärke der Enhancer variierte.

Um tiefere Einblicke in die Kooperativität von Enhancern bei der Genregulation zu gewinnen, wurden Enhancer-Paare in den synthetischen Lokus integriert, wobei jeder Enhancer in einem definierten Abstand zum Promotor positioniert war. Die Ergebnisse zeigen, dass die Kombination von schwachen und starken Enhancern zu einer deutlich erhöhten Expression führte, wenn der schwache Enhancer zwischen dem starken Enhancer und dem Promotor lag. Dies weist darauf hin, dass Enhancer-Synergie sowohl von der intrinsischen Aktivität als auch von der relativen genomischen Positionierung abhängt. CTCF-Bindestellen und andere strukturelle Elemente sind dafür bekannt, Enhancer–Promotor-Interaktionen zu modulieren. Um ihren Beitrag zur Enhancer-Kooperativität in unserem synthetischen System zu

untersuchen, habe ich zwei CTCF-assoziierte Regionen in mehreren Zelllinien mit synergistischen Enhancer-Paaren entfernt. Die Deletion dieser CTCF-Regionen führte zu einem weiteren Anstieg der Reportergenexpression, was darauf hinweist, dass Enhancer-Synergie durch CTCF-Bindestellen beeinflusst wird.

Zusammenfassend zeigen diese Ergebnisse, dass die Transkription eines Gens nicht durch einen einzelnen Faktor bestimmt wird, sondern aus dem Zusammenspiel von Enhancer-Kooperativität und strukturellen Elementen hervorgeht.

Introduction

The human body contains over 37 trillion cells, encompassing a wide variety of cell types. Proper development and cell differentiation require precise spatial and temporal control of gene expression. Gene transcription is precisely regulated through the coordinated actions of gene promoters and distal regulatory elements known as enhancers. Enhancers were first discovered in transient transfection assays by using plasmid reporters (Banerji et al., 1981), where the SV40 enhancer was shown to increase the expression of the β-globin reporter gene from various distances and in either orientation. These findings formed the textbook definition of an enhancer: a DNA sequence that activates a target promoter independently of distance and orientation. Enhancers are now recognized as fundamental regulators of transcription in eukaryotic genomes that play a crucial role in cell-type-specific gene regulation. In addition to their developmental roles, enhancers carry the majority of non-coding variants associated with human disease, highlighting their significance in both evolution and pathology. For instance, a single nucleotide change in the *Duffy* enhancer confers resistance to malaria (Tournamille et al., 1995). Point mutations in the ZRS enhancer cause ectopic Sonic hedgehog expression in the developing limb and lead to polydactyly (Kvon et al., 2020; Lettice et al., 2008). What's more, structural rearrangements that disrupt enhancer–promoter specificity are frequently implicated in oncogene activation in cancer (Ryan et al., 2015). Together, these findings underscore enhancers as key nodes in the regulatory genome and emphasize the importance of understanding their mechanisms of action to fully decipher gene regulation in health and disease.

1. Enhancer Grammar

Although the importance of enhancers in gene regulation is well established, a key question is how their activity is encoded in the underlying DNA sequence and interpreted by regulatory proteins. Central to this process is the interaction between transcription factors and their cognate DNA motifs. The affinity, arrangement, and combinatorial use of these motifs together constitute the "grammar" of enhancers, which dictates the timing, location, and strength of gene expression. Examining how transcription factor (TF) binding, motif organization

contribute to enhancer grammar provides a foundation for understanding the regulatory logic of transcription.

1.1 Transcription Factor Binding and Pioneer Activity

Enhancers are made up of multiple sequence motifs, which can be recognized by TFs that usually contain a DNA-binding domain and a transactivation domain. The DNA-binding domain can specifically identify its sequence motif and bind the DNA, and the transactivation domain can facilitate activation by recruiting cofactors. The activity of enhancers is largely determined by the binding of TFs and the subsequent recruitment of coactivators and histone modifiers that facilitate transcriptional initiation (Slattery et al., 2014; Spitz & Furlong, 2012).

How do TFs interact with enhancers? One hypothesis indicates that pioneer transcription factors are the first to engage silent chromatin, where they induce local accessibility and nucleosome displacement. This chromatin remodeling facilitates the recruitment of other factors, which mediate the assembly of coactivators and the transcriptional machinery to activate gene expression (**Figure1 A**) (de la Serna et al., 2005; Gassler et al., 2022; McDaniel et al., 2019). A well-studied pioneer transcription factor is FoxA1, which was identified to occupy nucleosome-bound enhancer regions of the serum albumin gene in endodermal cells, preceding the gene's transcriptional activation (Bossard & Zaret, 1998; Cirillo et al., 2002). Pioneer transcription factors can also play a potential bookmarking role during mitosis as they remain bound to mitotic chromatin (Caravaca et al., 2013; Deluz et al., 2016; Festuccia et al., 2016; Kadauke et al., 2012). During mitosis, chromatin becomes highly condensed, and most TFs dissociate from DNA (Gottesfeld & Forbes, 1997). Pioneer factors bound to mitotic chromatin can facilitate rapid rebinding of mitotically-excluded TFs and reactivation of their target genes in G1 (Kadauke et al., 2012; Liu et al., 2017).

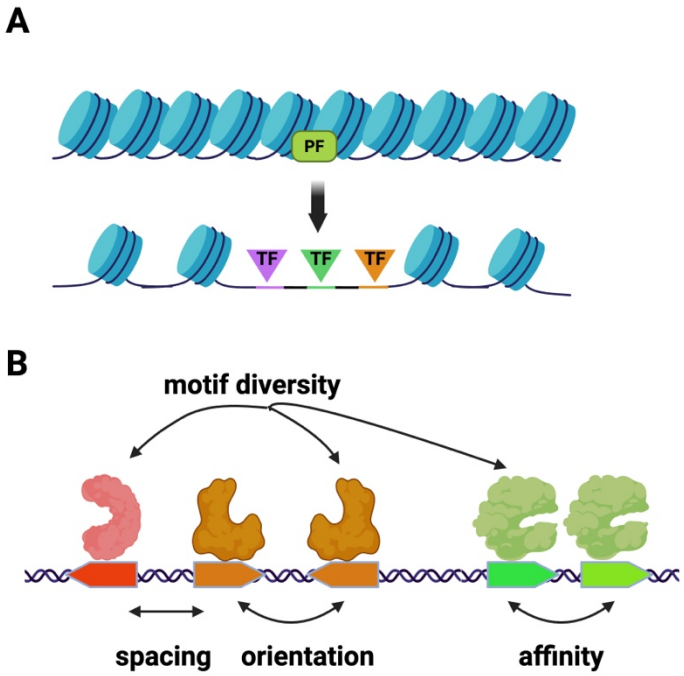


Figure 1. Interaction between pioneer transcription factors, TFs and enhancer. **A**. Binding of pioneer transcription factors to enhancers promotes chromatin accessibility, enabling recruitment of additional TFs. **B**. Enhancers comprise TFBSs for multiple TFs, which differ in their orientation, affinity, spacing, and arrangement. **Figure 1B** adapted from Jindal G& Farley E, 2021 (Developmental Cell); modified. © Elsevier (2021), used with permission (RightsLink License No. 6145960216212). Created with BioRender.com.

It remains controversial whether pioneering is a unique feature of select TFs or a broader capability that depends on cellular context. A recent study challenged the conventional classification of pioneer transcription factors by comparing FoxA1 and the non-pioneer TF HNF4A in naive K562 cells, assessing the concentrations required for each to bind accessible versus inaccessible chromatin (Hansen & Cohen, 2022). Overexpressed HNF4A displayed stronger binding to inaccessible regions than FOXA1. These findings suggest that pioneering activity may not be exclusive to a distinct TF subclass but instead reflects intrinsic DNA-binding affinity and expression level (Stoeber et al., 2024).

Once an enhancer adopts an open chromatin conformation, sequence-specific TFs can bind to their cognate transcription factor binding sites (TFBSs) within the enhancer. The efficiency of TF binding is modulated by the intrinsic affinity of the binding motifs, the DNA methylation status, and the presence of cooperative partner TFs.

1.2 DNA Methylation and Motif Affinity

In mammalian cells, de novo DNA methylation is catalyzed by Dnmt3a and Dnmt3b, whereas maintenance methylation is mediated by Dnmt1 (Bestor et al., 1988; Okano et al., 1999). The DNA-binding domains of TFs exhibit differential sensitivity to DNA methylation, which can be assessed using methylation sensitivity assays such as methylation-sensitive microarrays (Mann et al., 2013) and systematic evolution of ligands by exponential enrichment (SELEX) (Yin et al., 2017). Analysis of 542 human TFs by SELEX revealed that while 23% of tested TFs were inhibited by DNA methylation, 34% displayed increased binding affinity to methylated motifs.

1.3 Low- and High-Affinity Motifs in Enhancer Specificity

The affinity to the DNA-binding motif affects the dwell time of TFs. Imaging-based studies have demonstrated that TFs often transiently interact with numerous genomic loci, with dwell times of less than one second, before engaging a site for an extended period (~10 s), likely due to the presence of a high-affinity binding motif (J. Chen et al., 2014). Although interactions at low-affinity sites are transient, they may still play a critical role in enhancer function. One study showed that in *Drosophila*, *shavenbaby* enhancers contain clusters of low-affinity binding sites of Ubx-Exd. When replacing these sites with high-affinity sites, the specificity of the enhancer is lost. Such low-affinity Ubx binding sites offer a general mechanism for achieving both enhancer specificity and robustness, highlighting the importance of detecting and characterizing such motifs within enhancers (Crocker et al., 2015). Follow up study showed that this low-affinity enhancer can be activated in nuclear regions where Ubx and Hth are locally concentrated, indicating that microenvironmental TF enrichment promotes their transcriptional activity (Tsai et al., 2017). Another study demonstrated that the FoxF enhancer contains several low-affinity ETS binding sites. Mutations that increase the binding affinity of these sites lead to elevated enhancer activity, resulting in misregulation of gene expression and consequent defects in heart development of ciona embryos (Jindal et al., 2023). A seminal finding was reported for the ZRS enhancer, which contains five ETS binding sites. A mutation within the ETS-A site that increases binding affinity from 0.15 to 0.25 is sufficient to cause polydactyly. Further increases in binding affinity result in phenotypes with higher penetrance

and greater severity (Lim et al., 2024). Together, these show that the affinity of TFBSs is essential for establishing tissue-specific patterns of gene expression during development.

1.4 Cooperative Binding and Transcription Factor Syntax

Enhancers have multiple TFBSs (**Figure 1B**). A central question in gene regulation is the contribution of multiple TFBSs to enhancer activity. One model named billboard model was proposed to explain how TFs coordinate enhancer activity. In this model, TFs bind independently to an enhancer, with any single TF capable of maintaining a nucleosome-free state(Reiter et al., 2017; Spitz & Furlong, 2012). A recent CAP-SELEX study systematically analyzed 9,400 human TF–TF–DNA interactions and identified hundreds of TF pairs that recognize different heterodimeric motifs, many of which exhibited clear spacing and orientation preferences between their binding sites (Jolma et al., 2015). This finding indicates that DNA-guided cooperativity is common among TFs and suggests that, in most cases, cooperative binding arises not from direct protein–protein interactions but from the formation of DNA-mediated complexes. Building on this concept, a more recent CAP-SELEX screen of over 58,000 TF–TF combinations identified 2,198 interacting pairs and 1,131 composite motifs. These composite motifs, often bound by developmentally co-expressed TFs, were enriched in cell-type-specific regulatory elements (Xie et al., 2025).

Another model, the TF-collective model, proposes that TFs occupy an enhancer collectively through a combination of TF–DNA interactions at TFBSs and TF–TF interactions (Junion et al., 2012). The binding of one TF could impact the binding of additional ones (Carey, 1998), and single-nucleotide polymorphisms (SNPs) or mutations in the binding site of one TF can modulate the binding behavior of another (Deplancke et al., 2016). An analysis of five TFs critical for cardiac development showed that they frequently co-occupy a set of enhancers that contain motifs for different transcription factors subsets. Strikingly, removal of any one TF in in vitro assays eliminated the ability of the others to activate these enhancers (Junion et al., 2012). This might be facilitated by protein–protein interactions between these TFs. However, recently researchers performed systematic co-deletion assays to investigate potential dependencies in TF binding across promoters. Surprisingly, such binding dependencies are found to be uncommon, one-sided. For example, Msn2/4 can promote the recruitment of

multiple TFs to their cognate promoters, whereas the occupancy of Msn2 itself remains unchanged even when as many as 14 other TFs are simultaneously deleted (Lupo et al., 2023). These findings highlight the need to re-examine the contribution of TF-binding cooperativity to binding specificity, as well as the mechanisms by which non-DNA-binding domains shape this process.

High-resolution nuclease footprinting combined with single-molecule methylation profiling confirmed that transcription factor cooperativity is prevalent at active enhancers, with quantitative analyses indicating that cooperative occupancy represents the dominant binding mode at most active enhancers (Rao et al., 2021). Further integrative analyses of large-scale datasets such as ATAC-seq, TF ChIP-seq, and footprinting of mammals and *Drosophila* further revealed two distinct TF groups that frequently colocalize with the majority of expressed factors, forming characteristic “stripe” patterns. The first group comprises lineage-determining TFs, consistent with their roles in tissue-specific transcription. The second group, termed universal stripe factors, can recognize overlapping GC-rich motifs and enhance chromatin accessibility for partner TFs and extend their binding residence times (Zhao et al., 2022).

TFs cooperative interactions are essential for the activation of developmental enhancers in a cell–type–specific manner (Lagha et al., 2012). One recent study focused on coordinator, a 17-bp DNA motif, which was identified through comparative analysis of enhancer landscapes in human and chimpanzee cranial neural crest cells (Prescott et al., 2015). The coordinator contains common motifs by many homeodomain (HD) TFs and basic helix–loop–helix (bHLH) TFs, and it facilitates cooperative binding between HD TFs and TWIST1, which is bHLH family mesenchymal regulator. While expression of TWIST1 is broad across the developing mesenchyme, the expression of HD TFs is more regionally restricted. Through this interplay, Coordinator-mediated TF cooperativity regulates genes that govern cell-type and positional identity, thereby contributing to the developmental and evolutionary diversification of craniofacial morphology (Kim et al., 2024).

1.5 Disease-Associated Mutations in Enhancers

Investigating disease-associated mutations within cis-regulatory elements offers critical insights into the grammar underlying enhancer function. Such mutations can disrupt or create TFBSs, alter motif affinities, and ultimately influence enhancer strength and specificity. A well-known example is the human ZRS limb enhancer of Shh, located ~1 Mb from its target promoter, where sequence variants lead to ectopic Shh expression in the anterior limb bud and cause polydactyly across multiple vertebrate species (Kvon et al., 2020). Similarly, large-scale mutagenesis of a Ciona developmental enhancer has shown that increasing TF motif affinity can lead to loss of enhancer specificity, highlighting the importance of finely tuned binding site strength in maintaining spatial regulation (Farley et al., 2015). In another example, a single-nucleotide variant (SNV) in a non-regulatory region of the human α-globin locus creates a de novo promoter that downregulates α-globin expression, leading to α-thalassaemia (Bozhilov et al., 2021).

However, recent high-throughput studies suggest that enhancer strength is generally robust to mutations and does not depend on strict sequence conservation. For example, a survey of 5.9 million SNPs across four human genomes identified only ~30,000 variants that significantly altered the regulatory activity of cis-regulatory elements such as enhancers and promoters (van Arensbergen et al., 2019). Another study conducted functional assays in transgenic mice to investigate 23 ultraconserved enhancers function by introducing 1,547 mutations. Surprisingly, 83% (19/23) retained enhancer activity at one developmental stage, and over half remained functional later in development (Snetkova et al., 2021). These findings underscore that gene expression is not strictly determined by sequence conservation. Instead, it depends on key features of enhancers. These include the types and combinations of TFBSs, as well as their binding affinity, number, order, orientation, and spacing. Together, these features define the regulatory architecture of enhancers. These parameters collectively define the enhancer syntax or enhancer grammar, which describes how the arrangement and sequence context of TFBSs modulate enhancer activity (Jindal & Farley, 2021).

1.6 High-Throughput Approaches to Enhancer Grammar

Several studies applied massively parallel reporter assays (MPRAs) to investigate the TF syntax in cis-regulatory elements like enhancers and promoters and showed that motif grammar

is often flexible; however, mutations within TFBSs can still disrupt TF binding and impair regulatory activity (Barbadilla-Martínez et al., 2024; Georgakopoulos-Soares et al., 2023; King et al., 2020; Reiter et al., 2023; Singh et al., 2021; Song et al., 2023).

One example is notochord enhancers of Ciona, which contain low-affinity ETS and Zic TFBSs (Jindal & Farley, 2021). To test how enhancer grammar encodes notochord activity, the Ciona Brachyury enhancer (BraS) was analyzed. A library of 24.5 million variants was generated with the Zic and two ETS sites fixed, while all other nucleotides were randomized. Compared with wild-type BraS (73% notochord expression), the randomized version (BraS rZE) drove expression in only 28% of embryos, indicating the importance of additional sequence features. In a second library of 45 million variants, the Zic, two ETS, FoxA, and Bra sites were fixed, yielding a construct (BraS rZEFB) with expression levels comparable to wild type (62% vs. 73%). This result demonstrates that these five sites are necessary and sufficient for BraS activity. This study highlights the importance of enhancer grammar in notochord regulation (Song et al., 2023). In another work, a MPRA was applied to 209,440 synthetic sequences to systematically evaluate the arrangement and orientations of 18 liver-related TFBSs. The experiments revealed that the orientation and sequential order of TFBSs can substantially alter enhancer or promoter activity. Consistent patterns were observed in human promoter datasets, and a second MPRA involving 164,307 candidate liver regulatory elements validated similar orientation- and order-dependent effects. Incorporating TFBS orientation information into a sequence-to-expression prediction model led to a 7.7% improvement in accuracy. Together, these results indicate that TFBS orientation and order are critical parameters in transcriptional regulation and should be integrated into analyses of sequence variant function (Georgakopoulos-Soares et al., 2023).

However, enhancer activity is influenced not only by the TFBSs themselves, but also by the surrounding sequence context. A recent study in *Drosophila melanogaster* S2 cells used STARR-seq based high-throughput strategies to investigate how sequence flexibility and local context shape TF motif function. Eight representative TF motif types were inserted into hundreds of positions across nearly 500 enhancers to test how their activity changes with sequence environment. The results showed that, although key motifs could be substituted with alternative sequences from a limited number of motif classes, only a small subset of all possible sequences were functionally active. Moreover, motif activity was strongly influenced by the

local sequence context, including flanking regions, motif diversity, and spacing. As a result, the same motif did not exhibit equivalent activity across different positions (Reiter et al., 2023).

1.7 Deep Learning Approaches to Enhancer Grammar

Recently, deep learning models have been used to further investigate the regulatory grammar of cis-regulatory elements, combined with high-throughput reporter data such as MPRA and large multi-omics datasets. These models accurately capture regulatory grammar (de Almeida et al., 2022; Sahu et al., 2022; Vaishnav et al., 2022). DeepSTARR, a deep learning model, predicts the activity of thousands of developmental and housekeeping enhancers directly from DNA sequence in *Drosophila melanogaster* S2 cells. In addition to recognizing canonical TF motifs, the model also learned higher-order regulatory syntax, including the context-dependent functionality of identical motifs influenced by surrounding sequences and spacing (de Almeida et al., 2022). Another study applied deep residual networks (ResNets) to predict enhancer activity and used model-interpretation methods to show that these networks can learn regulatory grammars that recover known heterotypic TF clusters, illustrating how deep learning can reveal biologically meaningful regulatory rules (L. Chen & Capra, 2020).

Beyond TFBSs themselves, the regulatory influence of surrounding sequence context can also be quantitatively learned by deep learning models. One recent work introduced the concept of a transcription factor binding unit (TFBU), which explicitly incorporates the sequence context surrounding TFBSs. Based on this concept, a model named DeepTFBU was designed to rationally generate enhancers by tuning TFBS context sequences to modulate activity and achieve cell–type–specific responses. To validate this strategy, the authors measured the activity of over 36,000 DeepTFBU-designed sequences across diverse conditions by MPRA. The results showed that TFBS-context preferences are often TF- and cell type–specific, and that modifying the context region within TFBU modules significantly altered enhancer activity for the majority of TFs tested (82.9%, 97 out of 117). This approach demonstrates that context design is as critical as motif choice (J. Li et al., 2025).

Deep learning models have emerged as powerful tools for decoding the cis-regulatory code and predicting regulatory activity from DNA sequence (Bu et al., 2017; Erwin et al., 2014; Quang & Xie, 2016), More recently, generative deep learning approaches have been used to design synthetic enhancers and promoters with precise, cell type–specific activity (Barbadilla-Martínez et al., 2024; de Almeida et al., 2024; Gosai et al., 2024; Hecker et al., 2025; Schreiber et al., 2025, 2025; Taskiran et al., 2024). For instance, convolutional neural networks have been used to design de novo enhancers in *Drosophila* embryos targeting specific tissues, including the central nervous system and muscle, achieving high validation rates in vivo (de Almeida et al., 2024). Similarly, frameworks such as CREsted support end-to-end enhancer modeling, encompassing data preprocessing, sequence design, and motif analysis, and have demonstrated applicability in diverse systems, including the mouse cortex and zebrafish development (Hecker et al., 2025).

Further studies have shown that deep learning guided optimization can generate fully synthetic enhancers from random sequences, including compact or dual-cell-type enhancers, while also revealing fine-scale syntax rules such as motif arrangement and flanking sequence effects. In human cells, synthetic cis-regulatory elements designed through neural network models and massively parallel reporter assays have exhibited superior tissue specificity compared to natural sequences (Gosai et al., 2024; Taskiran et al., 2024).

In the context of promoters, a recent model named PARM was trained by integrating convolutional neural networks with custom-designed MPRAs to predict cell-type-specific promoter activity directly from promoter sequence. Beyond prediction, PARM facilitates the identification of TFBSs that shape promoter activity. It further captures detailed regulatory grammar, revealing positional preferences of activators and repressors and context-dependent motif–motif interactions (Barbadilla-Martínez et al., 2024). Together, these approaches establish a generalizable paradigm for learning enhancer and promoter grammar, understanding transcriptional regulation, and engineering functional cis-regulatory elements for basic and applied purposes.

2. Enhancer–Promoter Communication

Although enhancers are central regulators of mammalian gene expression, the mechanisms by which they specifically communicate with target promoters to activate mRNA transcription remain largely unclear. Enhancers can reside within introns of their target genes or be located tens to hundreds of kilobases away, and are capable of bypassing nearby non-target genes to interact with their target promoters specifically (Mifsud et al., 2015; Sanyal et al., 2012; Schoenfelder & Fraser, 2019). This long-range communication is facilitated by the 3D architecture of genomic DNA in nuclear space, with each level of spatial architecture contributing uniquely to enhancer–promoter interactions (Andrey et al., 2017; Javierre et al., 2016).

2.1 3D Genome Organization

The development of genomics assays to map genome-wide 3D chromatin interactions has revealed that the genome is hierarchically organized at several levels (**Figure 2**) (Dixon et al., 2012; Kempfer & Pombo, 2020; Lieberman-Aiden et al., 2009; McKay et al., 2020, 2020; Sexton et al., 2012). Within the interphase nucleus, individual chromosomes occupy distinct territories. At a megabase scale, transcriptionally active euchromatic regions tend to localize in the nuclear interior, where they interact with one another to form compartment A, often clustering near nuclear speckles, whereas heterochromatic and silenced regions associate with the nuclear lamina or nucleolus and form compartment B. At an intermediate scale from tens to hundreds kilobases, the genome is partitioned into self-interacting topologically associating domains (TADs), enriched for internal chromatin contacts but insulated from neighboring regions. The boundaries between TADs in mammals are frequently demarcated by CTCF or other insulator proteins. At the finest level of organization, cohesin-mediated loop extrusion brings enhancers and promoters into proximity to facilitate their regulatory interactions.

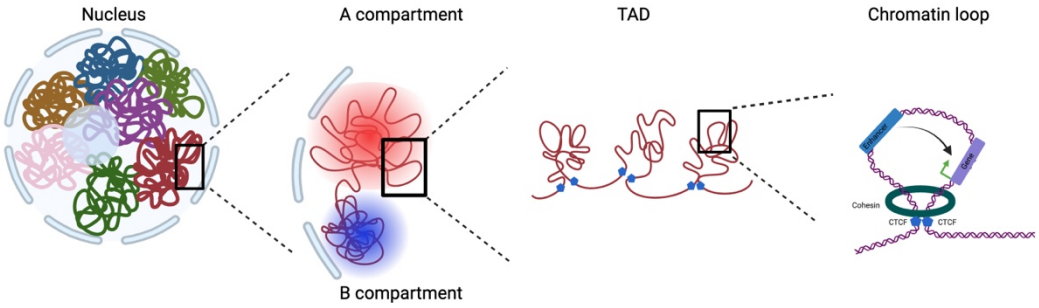


Figure 2. Genomic DNA in the nucleus is folded in a hierarchical manner across multiple scales. This three-dimensional organization gives rise to distinct architectural features, including chromosome territories, A/B compartments, and chromatin loops. Created with BioRender.com.

2.2 3D Genome Alterations and Misregulation of Enhancer–Promoter Interactions

Alterations in three-dimensional genome architecture can disrupt enhancer–promoter communication (**Figure 3**), resulting in enhancer adoption or "hijacking," and thereby contribute to developmental disorders and cancer (Lettice et al., 2011).

For instance, deletion of boundary elements may result in the fusion of neighboring TADs, enabling enhancers from one domain to ectopically activate genes in another, which can cause misexpression and disease (Bianco et al., 2018). At the *PITX2* locus, deletion of a 15-kb region spanning a TAD boundary allows *PITX2* to hijack pacemaker-specific enhancers, leading to its ectopic expression in the sinoatrial node and resulting in sinoatrial node dysfunction, atrial fibrillation, and developmental defects (Baudic et al., 2024). Another well-documented case is the *EPHA4* locus, where deletion of CTCF-associated boundaries permits inappropriate enhancer–gene contacts and causes congenital limb malformations (Lupiáñez et al., 2015). Likewise, duplications involving boundaries and adjacent genes may generate neo-TADs, placing duplicated genes under the influence of enhancers from nearby regions (Franke et al., 2016). A recent study reported that a 0.5-Mb duplication at the *Fgf8* locus resulted in the formation of neo-TADs and increased interactions between *Fgf8* and enhancers from a neighboring TAD, thereby driving ectopic activation in both forelimbs and hindlimbs through enhancer hijacking (Socha et al., 2021).

Structural variants such as inversions or translocations that cross TAD boundaries can reorganize regulatory domains. This phenomenon, known as TAD shuffling, can enable ectopic enhancer-promoter interactions and thereby drive abnormal gene regulation (**Figure 3C**). An illustrative example is an inversion affecting the *EPHA4* locus, which repositions an enhancer cluster, that normally regulates *EPHA4*, close to *WNT6*. As a result, *WNT6* becomes ectopically activated in an *EPHA4*-like expression domain, leading to limb malformations (Lupiáñez et al., 2015). Although alterations in 3D genome architecture can disturb enhancer–promoter specificity, leading to aberrant gene regulation, this outcome is not universal. Several studies have shown that chromosomal rearrangements do not necessarily coincide with changes in gene expression. For instance, analysis of the mouse *HoxD* cluster, which comprises two TADs, demonstrated that even large deletions spanning a TAD boundary had little effect on limb bud expression (Rodríguez-Carballo et al., 2017). Another study in *Drosophila* embryo shows that TAD shuffling or fusion rarely perturb the expression of most genes, indicating that gene regulation is buffered against large-scale topological alterations (Ghavi-Helm et al., 2019).

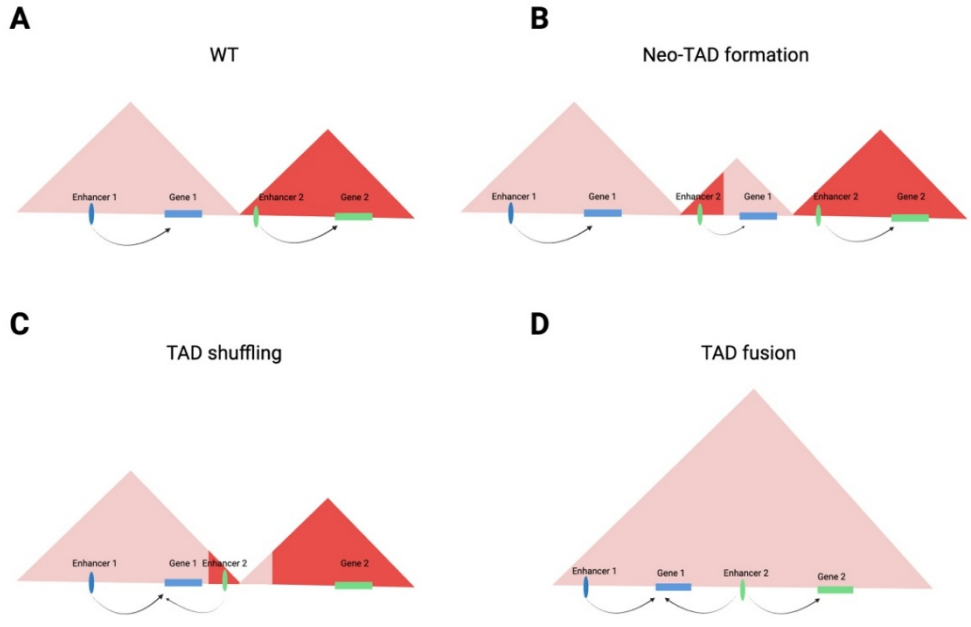


Figure 3. Structural variations can lead to TAD reorganization. **A**. Illustration of WT TAD structure. **B.** Illustration of neo-TAD generated by duplication of a boundary-containing region. **C.** Illustration of TAD shuffling generated by inversion of a boundary-containing region. **D.** Diagram illustrating TAD fusion triggered by boundary deletion. **Figure 3** adapted from Ibrahim DM and Mundlos S. (2020) modified. © Elsevier (2020), used with permission (RightsLink License No. 6145961227068). Created with BioRender.com.

2.3 Methods for Mapping 3D Genome Architecture

The investigation of three-dimensional genome architecture has been largely propelled by two major categories of technologies: imaging-based methods, most notably fluorescence *in situ* hybridization (FISH), and sequencing-based techniques derived from chromosome conformation capture (3C).

DNA-FISH was used for measuring the localization and distance between genomic loci in single cells. However, its relatively low throughput and limited spatial resolution constrain its applicability in large-scale and high-resolution analyses of chromatin architecture (Speicher et al., 1996). Subsequent advancements in super-resolution microscopy, including structured illumination microscopy (SIM) (Heintzmann & Cremer, 1999), single-molecule localization microscopy (SMLM) (Betzig et al., 2006), and stimulated emission depletion (STED) microscopy (Klar et al., 2000), have surpassed the classical diffraction limit, providing nanometer-scale resolution and enabling detailed visualization of chromatin organization. More recent innovations, including Oligopaints (Beliveau et al., 2012) and OligoFISSEQ (Nguyen et al., 2020), have further enhanced the multiplexing capabilities of FISH-based methods, allowing precise and high-throughput tracing of genomic regions across nuclear space.

Chromosome conformation capture (3C) can map chromatin interactions (“one versus one”) through proximity ligation techniques (Dekker et al., 2002). However, due to its inherently limited throughput and scope, the 3C method gave rise to more comprehensive variants. For instance, circularized 3C (4C) enabled genome-wide detection of interactions anchored at a single locus (“one-versus-all”) (Simonis et al., 2006), while Hi-C, a high-throughput extension, provided an unbiased, genome-wide view of chromatin contact landscapes (“all-versus-all”) (Andrey et al., 2013). Hi-C was pivotal in uncovering major principles of nuclear architecture, including the segregation of chromatin into A/B compartments (Lieberman-Aiden et al., 2009), the delineation of TADs (Dixon et al., 2012), and the formulation of the loop extrusion model (Schwarzer et al., 2017). Despite its strengths, Hi-C's resolution is inherently constrained by restriction enzyme digestion and sequencing depth. This limitation has been addressed by Micro-C, a modified method employing micrococcal nuclease for more uniform chromatin

fragmentation down to the nucleosome level, thereby enabling finer-scale interaction maps (Hsieh et al., 2015).

In parallel, hybrid approaches that combine proximity ligation with chromatin immunoprecipitation, including ChIA-PET, HiChIP, and PLAC-seq, have been developed to selectively capture chromatin interactions mediated by specific TFs or epigenetic modifications (Fang et al., 2016; Fullwood et al., 2009; Mumbach et al., 2016). Additionally, targeted capture methods such as Capture-C and Capture Hi-C have been introduced to focus on interactions anchored at predefined genomic loci (Davies et al., 2016; Jäger et al., 2015). These complementary imaging and sequencing-based approaches have established a foundation for elucidating the hierarchical organization of the genome and have provided unprecedented resolution into the spatial mechanisms that regulate enhancer–promoter communication.

2.4 Mechanistic Models of Enhancer–Promoter Communication

Although imaging- and sequencing-based methods have revealed the spatial organization of the genome and enhancer–promoter proximity, such findings remain largely correlative. Physical contact does not necessarily imply functional communication, and transcriptional activation can occur without detectable proximity. These limitations highlight the need for mechanistic explanations, and multiple mechanistic frameworks have been proposed to explain how enhancers communicate with their target promoters. In early one-dimensional (1D) tracking models, TFs or RNA polymerase II were thought to bind enhancers and move linearly along the DNA fiber until reaching the promoter(Yang & Hansen, 2024). However, this model cannot account for the frequent observation that enhancers bypass nearby promoters to regulate more distal targets. More recent three-dimensional (3D) models provide a better explanation(Furlong & Levine, 2018; Yang & Hansen, 2024). The best-characterized is the loop extrusion model, in which structural maintenance of chromosome (SMC) protein complexes such as cohesin actively extrudes DNA loops until constrained by boundary elements such as CTCF-bound sites in a convergent orientation, thereby bringing enhancers and promoters in close spatial proximity. A recent study demonstrated that loss of cohesin reduces enhancer–promoter contact frequency and consequently downregulates a subset of genes with distal enhancers (>50kb), providing support for a contact-based model of enhancer

function (Guckelberger et al., 2024). Another 3D mechanism is the direct bridging model, where Mediator and RNA polymerase II interaction can assemble the preinitiation complex (PIC), and connect enhancers with promoters to facilitate transcription initiation and elongation (Abdella et al., 2021; Richter et al., 2022; Soutourina, 2018). Furthermore, dimerization or oligomerization of TFs, such as LDB1 and YY1, can provide an additional means of bridging enhancers and promoters in 3D space (Krivega & Dean, 2017; Weintraub et al., 2017). A well-characterized example is the β-globin locus, where long-range interactions between the locus control region (LCR) and the target promoter are mediated by the self-associating factor LDB1, which is simultaneously recruited to both regulatory sites to facilitate enhancer–promoter communication (Deng et al., 2012).

Beyond direct physical contact, action-at-a-distance models were brought out as recent live-cell imaging studies have shown that enhancer–promoter distance does not decrease upon transcription activation. One such action-at-a-distance mechanism is the phase separation model, in which condensates formed through multivalent, low-affinity interactions mediated by intrinsically disordered regions of TFs, co-activators, Mediator, and RNA polymerase II generate transcriptionally active hubs. These hubs can promote and stabilize enhancer–promoter communication without requiring stable physical proximity (Chong et al., 2018; Hnisz et al., 2017). For example, super-resolution microscopy reveals that the distance between the *Shh* promoter and the neural enhancer SBE6, located approximately 100 kb away, was found to be bigger in transcriptionally active neural progenitor cells compared to silent embryonic stem cells (Benabdallah et al., 2019), challenging classical contact-based models. This phenomenon of increased enhancer–promoter distance was not limited to a single gene, but extended to multiple Estrogen-receptor alpha (ERα) regulated loci (Acuña et al., 2024). Consistent with this finding, an independent live-cell imaging study investigating the Sox2 locus in mouse embryonic stem cells also reported that physical proximity between enhancer and promoter is not required for transcriptional activation (Alexander et al., 2019).

Another proposed action-at-a-distance mechanism is the TF activity gradient (TAG) model, which suggests that enhancer-bound co-activators, such as CBP or the histone acetyltransferase p300, establish a local gradient of acetylated TFs. These modified TFs can then diffuse away from the enhancer, with promoters in closer proximity more likely to encounter them and

become transcriptionally activated (Karr et al., 2022). Supporting this view, several studies shows that transcriptional output declines as the genomic distance between enhancer and promoter increases (H. Chen et al., 2018; Nagano et al., 2025; Rinzema et al., 2022; Zuin et al., 2022).

Collectively, these models indicate that enhancer–promoter communication is not controlled by a single mechanism. Instead, it involves a range of strategies, including loop formation, dynamic molecular bridging, and condensate assembly. These processes may act together in a context-dependent manner to regulate gene expression.

2.5 Multi-Way Enhancer–Promoter Hubs and Regulatory Modules

Beyond single enhancer–promoter pairs, emerging 3C-based techniques such as multi-contact 4C, Hi-C, Tri-C, and Micro-C have revealed that enhancers and promoters often interact within complex multi-way hubs, particularly in the context of long-range regulatory interactions (Allahyar et al., 2018; Lando et al., 2024; Oudelaar et al., 2018). Regulatory hubs can take various forms: promoter hubs, where multiple enhancers regulate a single promoter (**Figure 4A**); enhancer hubs, where one enhancer interacts with several promoters (**Figure 4B**); and multi-connected hubs, which involve multiple enhancers and multiple promoters engaged in complex interaction networks (**Figure 4C**). Recent studies have revealed that multiple cis-regulatory elements within a TAD can spatially cluster to form sub-TAD regulatory modules. Depending on the experimental approach used, these structures have been described under different names, including Variable chromatin modules (VCMs) (Waszak et al., 2015), microcompartments (Goel et al., 2023), cis-regulatory domains (CRDs) (Delaneau et al., 2019), regulation associated modules (RAMs) (Zheng & Wang, 2022), chromatin nanodomains (CNDs) (Szabo et al., 2020), and super-enhancers (SE) (Sabari et al., 2018).

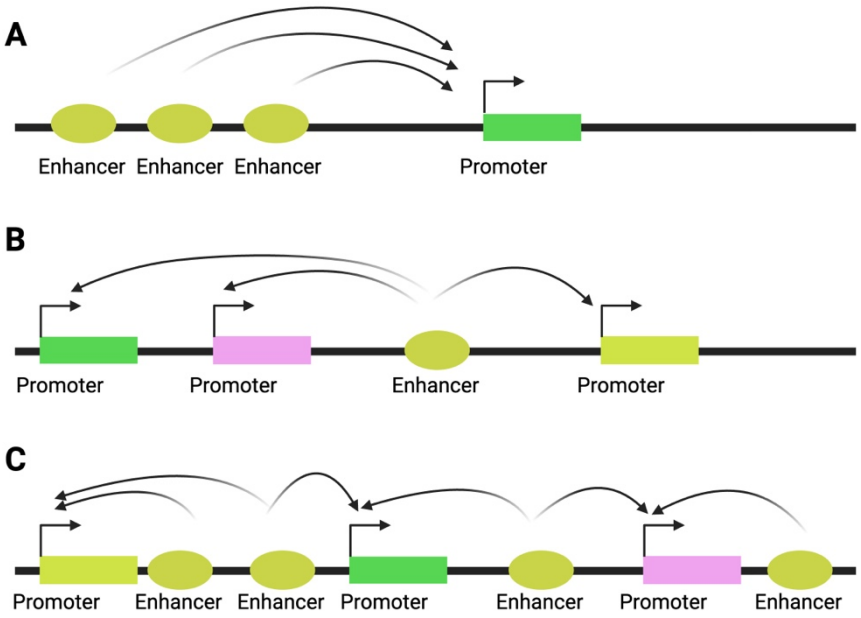


Figure 4. Graphical representation of multi-way enhancer–promoter Hubs. The top panel illustrates multiple enhancers regulating one promoter. The middle panel depicts a shared enhancer regulates several promoters. The bottom panel shows a complex network where multiple enhancers and promoters are interconnected. Figure 4 adapted from Uyehara CM and Apostolou E. (2023), modified. © Elsevier (2023), used with permission (RightsLink License No. 6145970847394). Created with BioRender.com.

One proposed model for the formation of 3D enhancer–promoter (E–P) hubs is phase separation, as mentioned earlier. Super-enhancers, which are composed of clusters of highly active enhancer elements, have been widely observed to be occupied by phase-separated condensates containing TFs, BRD4, and Mediator. These condensates are thought to create a concentrated microcompartment that facilitates transcription and have been frequently associated with multi-way interactions involving multiple enhancers and promoters (Sabari et al., 2018; Wang et al., 2021). Another example is enhancer hubs enriched with KLF4, whose structural integrity and associated regulatory interactions are disrupted upon KLF4 depletion (Di Giammartino et al., 2019). Interestingly, recent studies have demonstrated that active promoters and enhancers within such hubs are spatially confined to small A compartments enriched in H3K27ac, surrounded by B compartments with inactive chromatin (Harris et al., 2023). Consistently, an independent study has shown that multi-way chromatin interaction hubs are enriched for active histone modifications, including H3K4me3 and H3K27ac, supporting the notion that transcriptionally active hubs can spatially segregate from inactive chromatin in a manner analogous to A/B compartments, but at a finer scale (Lando et al., 2024).

Another model for such rosette-like hubs formation is the cohesin-mediated loop extrusion. A study shows that three-way chromatin contacts anchored at CTCF binding sites (CBSs) are critically dependent on both CTCF and cohesin. Experimental depletion of CTCF or cohesin degradation diminished the tendency of CBSs and their flanking regions to engage in three-way interactions. This indicates that both architectural proteins are essential for establishing and maintaining multi-way chromatin hubs, likely established through cohesin-mediated loop stacking (Allahyar et al., 2018; Hafner et al., 2023). Another study systematically examined the structural organization of super-enhancers and found that subsets of constituent enhancers within these regions can assemble into spatial hubs. ChIP-seq analysis further revealed that these hub enhancers frequently overlap with CTCF binding sites and cohesin peaks (Huang et al., 2018).

2.6 Shared Enhancers and Competition between Promoters

Building on the concept of multi-way regulatory hubs, an important scenario arises when a single enhancer is engaged by more than one promoter. In such promoter hubs, enhancers can coordinate the expression of multiple genes, yet this raises a central mechanistic question: how does a shared enhancer regulate multiple targets? Using a dual-reporter system in which two genes were positioned equidistant from a shared enhancer, it was shown that both promoters exhibit highly coordinated transcriptional bursting dynamics (Fukaya et al., 2016). Further study revealed that this coordinated bursting is mediated by clusters of TFs and co-activators assembled at the shared enhancer, which serve as a transcription hub to synchronously activate transcription at both promoters (Kawasaki & Fukaya, 2023). Consistent with this study, Tri-C experiment in erythroid cells demonstrated that multiple promoters can interact with α-globin LCR simultaneously. Deletion of a TAD boundary allows promoters located outside the TAD domain to interact with the LCR, in addition to the promoters within the TAD. The observations imply that promoters are not mutually exclusive in their enhancer interactions (Oudelaar et al., 2018, 2019).

However, the ectopic promoter–enhancer contacts established upon boundary deletion have been shown to attenuate the expression of intra-TAD genes (Hanssen et al., 2017). Two additional studies have demonstrated that a single-nucleotide change within the α-globin locus

can give rise to a novel promoter, which perturbs downstream globin gene expression and leads to α-thalassaemia (Bozhilov et al., 2021; De Gobbi et al., 2006). These findings are consistent with observations at the β-globin locus, where β-like globin genes appear to compete for LCR engagement in a mutually exclusive manner. Specifically, experimental redirection of the LCR toward the γ-globin gene cluster led to an increased burst frequency of γ-globin and a corresponding decrease in β-globin burst frequency, without affecting burst size (Bartman et al., 2016; Deng et al., 2012, 2014).

2.7 Enhancer Cooperation in Gene Regulation

There exist many examples of enhancer hubs with one promoter interacting with multiple active enhancers (Kvon et al., 2021; Q. Li et al., 2002; Whyte et al., 2013). A key question is how these enhancer hubs regulate target genes, as the enhancer strength of each enhancer can be different. It has been reported that multiple enhancers can regulate gene expression in a sub-additive, additive, or super-additive (synergistic) manner, depending on their strength, spatial arrangement, and chromatin context. In the additive manner, the transcriptional output of a promoter reflects the sum of activities from individual enhancers, suggesting minimal interaction among them. Super-additive (or synergistic) effects arise when combined enhancer activity exceeds the expected total, indicating cooperative interactions that amplify gene expression. Sub-additive behavior is observed when the combined effect falls short of the sum, often due to the saturation of the promoter (Bothma et al., 2015; Choi et al., 2021; Loubiere et al., 2024; Scholes et al., 2019; Thomas et al., 2021).

For example, a recent study examined how combinations of primary and shadow enhancers regulate the developmental genes *hunchback* (*hb*), *knirps* (*kni*), and *snail* (*sna*) in early *Drosophila* embryos. At the *kni* locus, the primary and shadow enhancers showed additive, and in some cases super-additive, effects on transcriptional output. In contrast, *hb* enhancers showed context-dependent interactions: they functioned sub-additively in anterior regions of the embryo where the *Bicoid* activator is abundant but shifted toward additive behavior in posterior regions with lower *Bicoid* levels. Interestingly, the *sna* enhancers also demonstrated sub-additive regulation (Bothma et al., 2015). An independent study shows that deletion of individual shadow enhancers of snail did not result in observable developmental defects under

standard conditions. However, when exposed to environmental stress, such as elevated temperature, these deletions led to impaired gastrulation (Perry et al., 2010). Several other studies show that shadow enhancers are seemingly redundant and often drive overlapping spatial or temporal expression patterns (X. Li & Levine, 2024). However, genome editing and live-cell imaging have revealed that shadow enhancers play important roles in fine-tuning gene expression, ensuring developmental precision, and enhancing phenotypic robustness, particularly under environmental or genetic stress (Kvon et al., 2021).

Beyond shadow enhancers, additive and synergistic interactions among other enhancers have also been observed. At the *Fgf5* locus, individual enhancer elements display low intrinsic activity in reporter assays; however, targeted deletion of these elements results in substantial reductions in endogenous *Fgf5* expression. This indicates that enhancers exhibiting low intrinsic activity can function cooperatively in a super-additive manner to drive robust gene expression at their native genomic locus (Thomas et al., 2021). A separate study examining the transdifferentiation of human B cells into macrophage-like cells also identified both additive and super-additive modes of enhancer cooperation. While the majority of enhancers acted additively and independently influenced promoter activity, a subset of target genes (~20%) exhibited synergistic enhancer interactions, where combined activity exceeded the expected sum of individual effects. Notably, the synergistic enhancers are enriched for cell-type-specific TFs and can activate macrophage-specific genes, thereby facilitating efficient cell fate transitions (Choi et al., 2021).

Extending these observations to a systematic, high-throughput setting, researchers discovered distinct modes of enhancer–enhancer cooperation with STARR-seq: housekeeping enhancers tend to function additively, with their combined activity closely matching the sum of their individual effects. In contrast, developmental enhancers exhibit super-additive behavior. Notably, this super-additive behavior does not depend on the enhancers' native genomic positions or specific transcription factor binding motifs (Loubiere et al., 2024). However, another study systematically tested about 69000 enhancer-enhancer-promoter combination in mouse embyronic stem cells and found that enhancer pairs generally combine their activities in a near-additive manner. This behavior was consistently observed across multiple developmental promoters. The study further demonstrated that cases of super-additivity were

rare and occurred predominantly in the context of weak promoters (Martinez-Ara et al., 2024). The discrepancy between these findings underscores the influence of promoter context, enhancer class, and methodological approach on the observed modes of enhancer cooperation.

2.8 New Class of Regulatory Elements

In addition to canonical enhancers within enhancer hubs, a subset of cis-regulatory elements has been identified that bears active enhancer-like chromatin marks but exhibits little or no intrinsic activity. This new class of regulatory elements act more like architectural components, facilitating enhancer–promoter communication or boosting the activity of distal enhancers. Depending on the genomic locus where they were reported, these elements have been termed "facilitator elements", "tethering elements", or "range extenders." (Blayney et al., 2023; Bower et al., 2025; Brosh et al., 2023; X. Li & Levine, 2024; Z. Zhou et al., 2025).

In *Drosophila*, structural elements known as tethering elements were first identified in *Hox* gene *Scr* (Calhoun et al., 2002). Although tethering elements lack intrinsic enhancer activity, they promote spatial proximity between distal enhancers and promoters and contribute to *Scr* gene activation (Levo et al., 2022). Similarly, Orphan CGIs (oCGIs) have also been proposed to act as tethering elements that facilitate physical communication between poised enhancers and target genes and contribute to gene expression (Pachano et al., 2021).

Such structural elements are also reported in super-enhancer regions. A recent study identified facilitator elements within the α-globin LCR. The α-globin LCR comprises five enhancers: R1, R2, R3, Rm, and R4. While distal enhancers R1 and R2 strongly activate α-globin transcription, the proximal elements R3, Rm, and R4 show minimal activity on their own. R3/ Rm/R4-only locus failed to drive expression, but each of these elements significantly boosted R1/R2-mediated activation in a position-dependent manner, indicating their function as "facilitators" that enhance the ability of R1 and R2 to activate the α-globin promoter (Blayney et al., 2023).

In the *Sox2* locus, which contains the *Sox2* Control Region (SCR) super-enhancer, the functional contribution of individual regulatory elements was dissected using the "Big-IN" strategy, in which different combinations of DNase hypersensitive sites (DHSs) were used to replace the entire SCR. This combinatorial analysis revealed that certain DHSs, although inactive on their own, can act as facilitator elements that increase the activity of neighboring elements through cooperative interactions (Brosh et al., 2023). Facilitator elements are present not only within the SCR, but also in regions proximal to the *Sox2* promoter. For example, DHS15 was reported to show weak enhancer activity by luciferase reporter assay or deletion by CRISPR/Cas9 system (H. Y. Zhou et al., 2014). Live-cell imaging showed that in the WT cells, condensate proximity to Sox2 gene can increase burst size and frequency. However, deletion of DHS15 abolished this effect, as burst size and frequency became comparable between proximal and distal condensates. This indicates that the "weak" enhancer is required for burst enhancement (Du et al., 2024).

Similarly, a recent study reported the identification of "Range Extenders" (REX), which lack intrinsic enhancer activity but facilitate long-range enhancer–promoter communication during limb development (Bower et al., 2025). REX exists in the surrounding region of a limb-specific enhancer HS72, which failed to activate Shh transcription when relocated to the endogenous position of the ZRS enhancer. However, when HS72 was fused with REX, long-range activation of Shh was restored. This demonstrated that REX could extend the functional range of HS72, effectively converting it into a long-range-acting enhancer. Similar range-extending effects were observed when other limb-specific enhancers were paired with REX, highlighting its role in facilitating long-distance enhancer–promoter communication. Motif analysis of REX has revealed an enrichment of binding sites for LHX2. Disruption of LHX2 motif impairs long-range enhancer function without affecting enhancer strength. LHX2 interacts with LDB1, a cofactor previously implicated in mediating long-range chromatin interactions at the β-globin locus, as well as in the formation of multi-enhancer hubs at the Myc locus and inter-chromosomal enhancer hubs(Aboreden et al., 2025; Deng et al., 2012; Monahan et al., 2019). LHX2 and LDB1 may contribute to REX-mediated enhancer–promoter communication. These findings suggest that REX elements may promote spatial genome organization by engaging architectural proteins to support distal regulatory contacts (Bower et al., 2025).

3. Methods for Enhancer Discover

The studies discussed above demonstrate that enhancers regulate transcriptional output by modulating bursting dynamics, thereby influencing both the precision and robustness of gene expression. These findings highlight the importance of enhancers as central control elements of transcription. Yet, a fundamental challenge remains: How to reliably identify enhancers across the genome and determine which sequences are functionally active in a given cellular or developmental context. For example, an element that drives robust transcription in one tissue or cell type may remain completely inactive in another, and some enhancers become functional only at defined developmental time points or in response to particular external stimuli. This pronounced context dependency has made the identification of enhancers a persistent challenge. To address this, a wide range of experimental strategies has been developed over the years to discover and characterize enhancers, range from classical reporter-based approaches to modern high-throughput and genome-wide methods, which are discussed in the following section.

3.1 Histone Modification and Transcription Factor ChIP-seq

Active enhancer regions are marked by flanking enrichment of histone modifications such as H3K4me1 and H3K27ac (Shlyueva et al., 2014). This epigenomic signature enables enhancer prediction using ChIP-seq (Heintzman et al., 2007). Active enhancers also contain TF binding motifs, which can be specifically recognized and bound by TFs and their co-activators (e.g. p300) (**Figure 5A**). ChIP-seq profiling of tissue-specific TFs enables the identification of enhancers active in specific cell types or tissues. In particular, p300 ChIP-seq performed in embryonic mouse tissues was able to predict enhancer activity in corresponding tissues with high accuracy (Visel et al., 2009). Further studies have leveraged TF or histone mark ChIP-seq to map tissue-specific enhancer landscapes in multiple developmental and differentiated contexts (**Figure 5B**) (Ko et al., 2017; Zhang et al., 2024; P. Zhou et al., 2017).

Conventional ChIP-seq requires large cell numbers (10^5–10^7), limiting its application in rare populations. CUT&RUN overcomes this by using a Protein A/G–MNase fusion to cleave DNA at protein-binding sites in situ, eliminating the need for crosslinking and sonication. This method achieves high signal-to-noise ratios and works with as few as 100–1000 cells (Skene et al., 2018). It has enabled high-resolution mapping of factors like SOX2, and Nanog in early embryos, revealing distinct binding profiles (Hainer et al., 2019). CUT&Tag is an advanced method for profiling DNA–protein interactions. It utilizes a Protein A/G–fused Tn5 transposase (pA/G-Tn5) to simultaneously cleave and tag DNA at binding sites, enabling efficient library construction without the need for end repair or adapter ligation. CUT&Tag offers improved simplicity, higher sensitivity, and better suitability for limited material (Kaya-Okur et al., 2019).

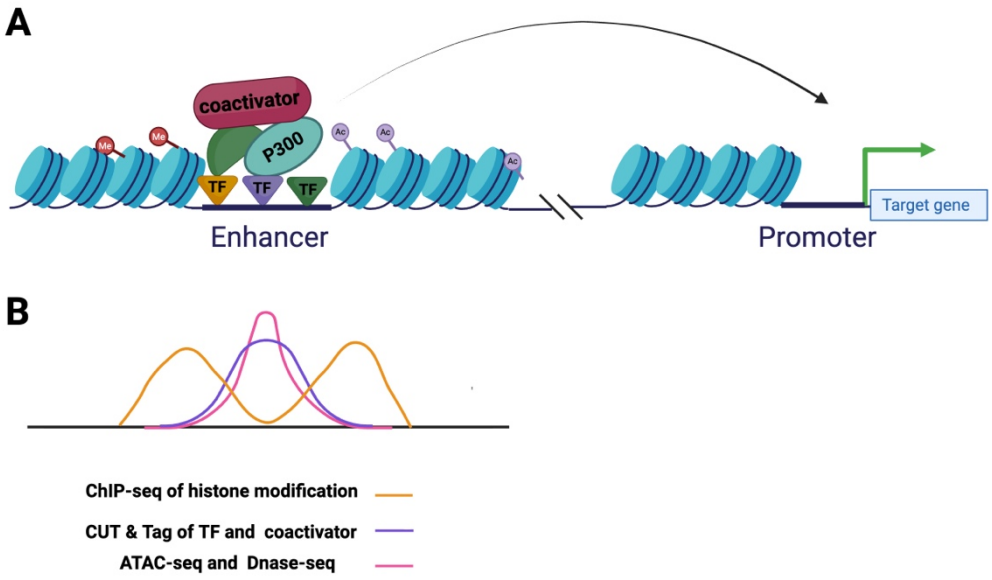


Figure 5. Molecular features and experimental approaches for enhancer characterization. **A.** Schematic representation of features in active enhancer. TFs and co-activators, such as P300, bind to enhancer regions enriched with histone modifications. **B.** Representative genomic profiles used to identify enhancers. Histone modification patterns are detected by ChIP-seq (orange), TF and co-activator binding is mapped by CUT&Tag (purple), and chromatin accessibility is assessed by ATAC-seq and DNase-seq (pink). Created with BioRender.com.

3.2 Chromatin Accessibility Assays (DNase-seq, MNase-seq, ATAC-seq)

Active enhancer regions are marked by increased chromatin accessibility, which can be detected by DNase-seq, MNase-seq, or ATAC-seq (Buenrostro et al., 2013; Kaplan et al., 2009; Thurman et al., 2012). DNase I is an endonuclease and can preferentially cleaves DNA at DNase I hypersensitive sites, which correspond to open chromatin regions. Following digestion,

the resulting DNA fragments are amplified and subjected to high-throughput sequencing (Schones et al., 2008). MNase possesses both endonuclease and exonuclease activities and preferentially digests linker DNA between nucleosomes, which can be used to map nucleosome positioning (Chereji et al., 2019). However, both DNase-seq and MNase-seq require a large number of cells as input, limiting their application in rare samples, such as early embryos. ATAC-seq utilizes a hyperactive Tn5 transposase to simultaneously fragment DNA and insert sequencing adapters, a process termed tagmentation which occurs preferentially in regions of accessible chromatin. As a result, sequencing-ready DNA fragments are predominantly generated from open chromatin regions, enabling the high-resolution identification of regulatory elements, such as active promoters and enhancers. Compared to DNase-seq and MNase-seq, ATAC-seq requires fewer cells (500 to 50,000 cells), and involves only two steps: nuclei isolation followed by tagmentation. This makes ATAC-seq especially valuable for profiling rare cell populations or clinical samples with limited availability (Buenrostro et al., 2013).

3.3 Massively Parallel Reporter Assays (MPRAs and STARR-seq)

While chromatin signatures enable the identification of active enhancers, they do not directly reflect the magnitude of enhancer activity. Reporter assays have been developed to directly detect the enhancer strength of cis-regulatory elements. In such assays, candidate regulatory elements are cloned upstream of a reporter gene (e.g., luciferase) and transiently transfected into cells. The activity of the elements is inferred from the reporter gene expression level (Neumayr et al., 2019). As traditional reporter assays are not suitable for high-throughput analysis, MPRAs were developed to overcome this limitation and enable simultaneous quantification of the activity of numerous candidate enhancers (**Figure 6A**) (Patwardhan et al., 2012). MPRAs have been used to detect cis-regulatory elements in the entire genomes of different species, as well as artificially designed sequences (Melnikov et al., 2012; Patwardhan et al., 2012). In the second study, MPRA technology was used to analyze the function of three liver enhancers via saturation mutagenesis. A library of over 100,000 enhancer variants was constructed by cloning each variant upstream of a minimal promoter and luciferase reporter, with unique 20-bp barcodes inserted into the 3′ UTR. The library was delivered to mouse liver via hydrodynamic tail vein injection, followed by RNA-seq after 24 hours to quantify enhancer

activity based on barcode expression. It was found that, while most mutations had modest effects, TFBSs were not always predictive of functional output, underscoring the need for empirical validation of these sites (Patwardhan et al., 2012).

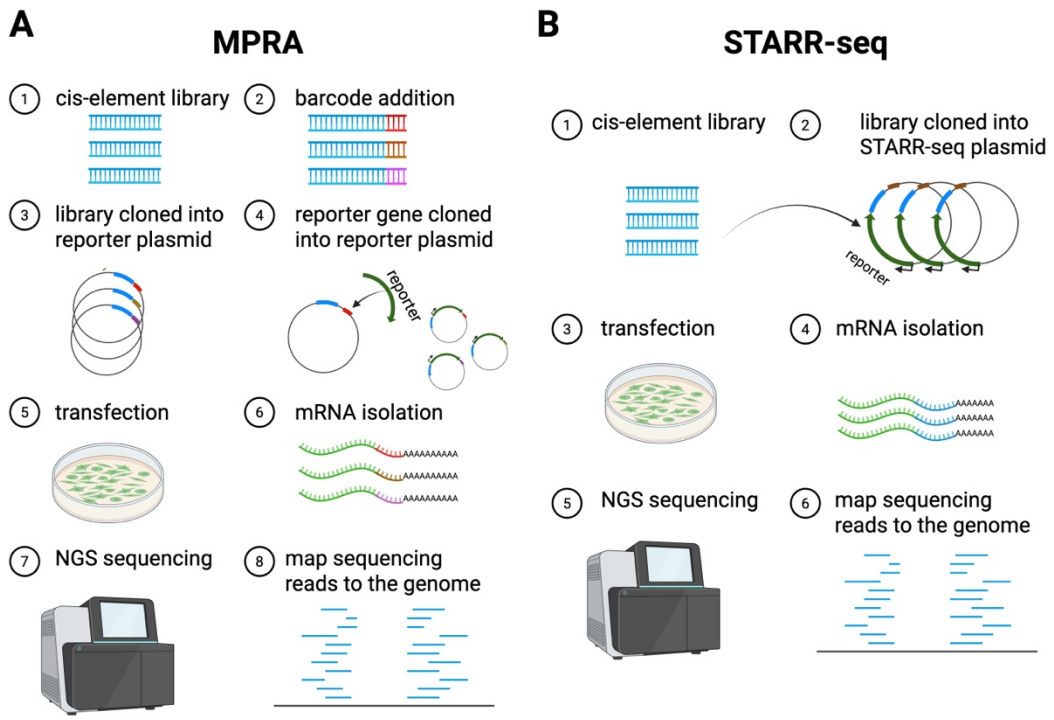


Figure 6. Experimental workflows of MPRA and STARR-seq for high-throughput functional characterization of cis-regulatory elements. **A.** Schematic representation of MPRA strategy. A cis-regulatory element library is barcoded and cloned into plasmids followed by insertion of a reporter gene. The plasmid library is transfected into cells, followed by mRNA isolation, next-generation sequencing (NGS), and mapping of sequencing reads to the genome to quantify enhancer activity. **B.** Schematic representation of STARR-seq strategy. Cis-regulatory elements are cloned into a STARR-seq reporter plasmid downstream of a minimal promoter. After transfection, RNA transcripts derived from active elements are isolated, sequenced, and mapped back to the genome to assess regulatory activity. Created with BioRender.com.

An alternative to traditional MPRAs is STARR-seq, which has been developed to directly measure enhancer activity without relying on barcode sequences (**Figure 6B**). In STARR-seq, genomic DNA fragments are inserted into the 3′ UTR of a reporter gene, allowing active enhancers to drive their own transcription upon vector transfection. Reporter transcripts are then quantified by deep sequencing, enabling direct and high-throughput assessment of enhancer function. This method allows genome-wide, quantitative screening of millions of candidate regulatory elements, regardless of their length or origin (Arnold et al., 2013). STARR-seq has been utilized to investigate enhancer–promoter compatibility (Zabidi et al., 2015). The study revealed that in *Drosophila*, housekeeping promoters are selectively activated

by housekeeping enhancers, whereas developmental promoters predominantly interact with developmental enhancers, indicating a functional specificity between enhancer–promoter pairs.

Despite their power, MPRAs have limitations. They are often restricted to cell lines or model organisms, and the artificial context of the plasmids ignores native three-dimensional chromatin organization (Arnold et al., 2013). Consequently, a positive singal does not necessarily prove its in vivo regulatory activity. Comparison between ChIP-seq profiles of active histone marks or TFs and MPRA-derived functional data reveals a subset of elements that, despite displaying canonical enhancer-associated chromatin features, are insufficient to drive gene expression independently. This suggests that these histone modifications do consistently correlate with enhancer activity. Instead, these elements can contribute to regulatory activity only when acting cooperatively with other enhancers (Brosh et al., 2023).

While reporter-based assays and genome-wide approaches such as MPRAs and STARR-seq have greatly advanced our ability to identify and functionally characterize enhancers, they fall short in capturing a recently described class of cis-regulatory elements, such as facilitator elements, which show little or no intrinsic activity when assayed in isolation. Moreover, while MPRA-based assays are well suited for evaluating the activity of single enhancers or pairs, they lack the capacity to assess interactions among multiple enhancers arranged across extended genomic distances. How these elements cooperate with canonical enhancers to form higher-order regulatory hubs and sustain robust gene expression remains poorly understood. Loss-of-function studies at endogenous loci have yielded important insights, yet they remain limited in their ability to disentangle the specific contributions of individual facilitator elements within complex genomic architectures. In addition, such approaches are inherently labor-intensive and time-consuming, as each perturbation requires precise genome editing and subsequent validation. Their low experimental throughput further constrains systematic interrogation of large numbers of elements or combinatorial configurations, thereby limiting a comprehensive understanding of enhancer cooperation in its native genomic context.

To address these limitations, my main project focused on investigating enhancer cooperation in a controlled setting. I developed a synthetic platform that enables the systematic integration

of enhancers at varying distances and in diverse combinations, thereby allowing complex regulatory architectures to be assembled in a bottom-up fashion. Using this system, I found that enhancer activity declined with increasing enhancer–promoter distance. Furthermore, systematic testing of enhancer pairs revealed that synergistic effects emerged when a strong enhancer was placed distally, with a weaker enhancer situated between the strong enhancer and the promoter.

Chapter I: Enhancer cooperativity can compensate for loss of activity over large genomic distances

Henry F. Thomas, Songjie Feng, Felix Haslhofer, Marie Huber, Maria Garcia Gallardo, Vincent Loubiere, Daria Vanina, Mattia Pitasi, Alexander Stark, Christa Buecker

Thomas et al., 2025, Molecular Cell 85, 362–375

https://doi.org/10.1016/j.molcel.2024.11.008

Status: Published

Contribution:

Enhancers are cis-regulatory elements in the genome that control gene expression in a spatial and temporal manner. At some loci, enhancers form clusters to activate transcription. However, how individual enhancers communicate and cooperate in gene regulation remains largely unknown.

Here, I established a synthetic platform in a mouse embryonic stem cell line that enables the integration of multiple enhancers at three defined distances, either individually or in pairs, at distinct genomic positions. I found that increasing the distance between enhancer and promoter reduces enhancer activity. Interestingly, the introduction of a weak enhancer between a strong and a weak enhancer facilitated the ability of the distal strong enhancer to activate the target gene.

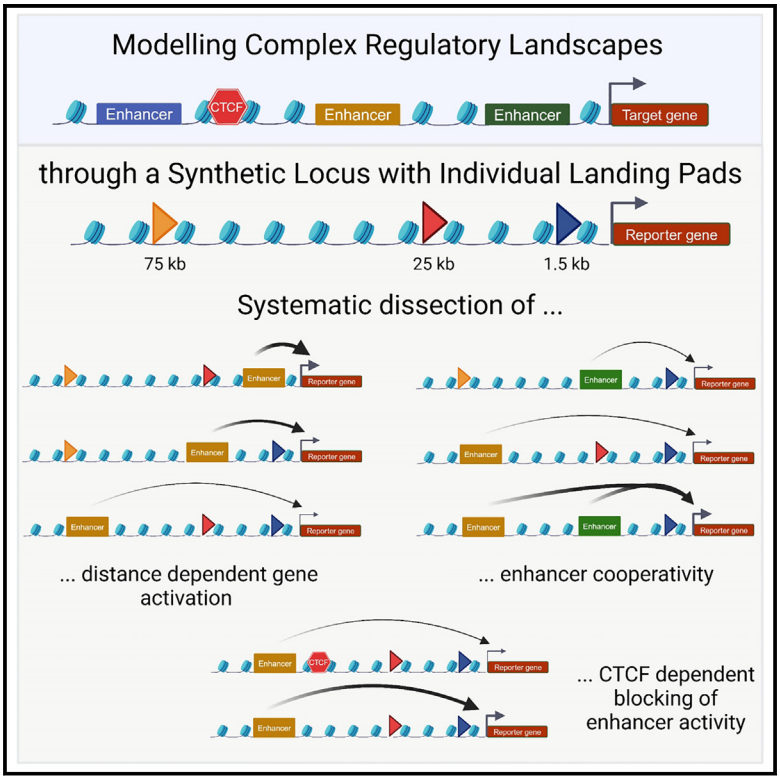
As one of the lead authors, I contributed to this work in several key aspects, including project design, experimental execution, data analysis, and manuscript writing. I integrated the loading pads to generate the synthetic locus and selected three enhancers for subsequent validation and investigation in our system. I also quantified enhancer activity in cell lines containing a single enhancer positioned at different distances using FACS and analyzed the correlation between enhancer–promoter distance and reporter expression. Furthermore, I generated cell lines harboring different enhancer pairs and validated their activity. These experiments revealed that

enhancer synergy occurred only when enhancer pairs were arranged at specific distances and orientations.

Molecular Cell

Enhancer cooperativity can compensate for loss of activity over large genomic distances

Graphical abstract



Highlights

- Synthetic platform with multiple landing pads to build complex regulatory landscapes
- Distance-dependent activation is specific to each enhancer
- Position-dependent cooperation occurs between weak and strong enhancers
- CTCF sites can lower enhancer-dependent activity but do not interfere with synergy

Authors

Henry F. Thomas, Songjie Feng, Felix Haslhofer, ..., Mattia Pitasi, Alexander Stark, Christa Buecker

Correspondence

henry.fabian.thomas@univie.ac.at (H.F.T.), christa.buecker@maxperutzlabs.ac.at (C.B.)

In brief

Thomas, Feng, et al. introduce a synthetic platform that allows the building of complex regulatory landscapes. Integrating the same enhancer at different distances from a promoter uncovered that activation from a distance is an individual characteristic of each enhancer. Combining multiple enhancers revealed extensive cooperativity among individual elements.

Thomas et al., 2025, Molecular Cell *85*, 362–375
January 16, 2025
https://doi.org/10.1016/j.molcel.2024.11.008

CellPress
OPEN ACCESS

Molecular Cell

Article

Enhancer cooperativity can compensate for loss of activity over large genomic distances

Henry F. Thomas,[1,2,3,6,*] Songjie Feng,[1,2,3,6] Felix Haslhofer,[1,2] Marie Huber,[1,2] María García Gallardo,[1,2,3] Vincent Loubiere,[4] Daria Vanina,[1,2] Mattia Pitasi,[1,2,3] Alexander Stark,[4,5] and Christa Buecker[1,2,7,*]

[1]Max Perutz Laboratories, Vienna BioCenter Campus (VBC), Dr.-Bohr-Gasse 9, 1030 Vienna, Austria
[2]University of Vienna, Center for Molecular Biology, Department of Microbiology, Immunobiology, and Genetics, Dr.-Bohr-Gasse 9, 1030 Vienna, Austria
[3]Vienna BioCenter PhD Program, a Doctoral School of the University of Vienna and the Medical University of Vienna, 1030 Vienna, Austria
[4]Research Institute of Molecular Pathology (IMP), Vienna BioCenter (VBC), 1030 Vienna, Austria
[5]Medical University of Vienna, Vienna BioCenter (VBC), Vienna, Austria
[6]These authors contributed equally
[7]Lead contact
*Correspondence: henry.fabian.thomas@univie.ac.at (H.F.T.), christa.buecker@maxperutzlabs.ac.at (C.B.)
https://doi.org/10.1016/j.molcel.2024.11.008

SUMMARY

Enhancers are short DNA sequences that activate their target promoter from a distance; however, increasing the genomic distance between the enhancer and the promoter decreases expression levels. Many genes are controlled by combinations of multiple enhancers, yet the interaction and cooperation of individual enhancer elements are not well understood. Here, we developed a synthetic platform in mouse embryonic stem cells that allows building complex regulatory landscapes from the bottom up. We tested the system by integrating individual enhancers at different distances and confirmed that the strength of an enhancer contributes to how strongly it is affected by increased genomic distance. Furthermore, synergy between two enhancer elements depends on the distance at which the two elements are integrated: introducing a weak enhancer between a strong enhancer and the promoter strongly increases reporter gene expression, allowing enhancers to activate from increased genomic distances.

INTRODUCTION

Transcription needs to be tightly controlled to ensure correct expression levels. In higher eukaryotes, *cis*-regulatory elements such as enhancers control when and in which cells a promoter is active.[1,2] Enhancers are short DNA sequences that consist of multiple transcription factor (TF) binding sites and control the expression of their target genes from a distance.[2] With increasing genomic distance, individual enhancers activate lower levels of expression from the promoter,[3–5] raising the question of how enhancers can bridge the distance to their promoter.

Enhancers are typically studied either in their native genomic context or in reporter assays.[2,6] Dissection of enhancer loci has, e.g., determined the contribution of tissue-specific TFs,[1,7–9] identified a role for 3D genome organization in enhancer-promoter communication,[2,10] and described various modes of enhancer cooperativity.[11–20] However, the complexity of endogenous regulatory landscapes can obstruct their analysis. A recent dissection of the α-globin locus used advances in synthesis of entire genomic loci to remove all enhancers and reintroduce them individually.[20] This approach revealed the super-additive behavior of several enhancers, contrasting previous conclusions of additive behavior at the same locus.[12,20] This underlines the importance of studying individual elements of multi-enhancer clusters both in the presence and absence of other *cis*-regulatory elements. Genetic modification of endogenous loci remains laborious and is limited to few selected enhancer clusters, making it challenging to derive general rules of enhancer cooperativity.

Conversely, reporter assays eliminate regulatory complexity and analyze the activity of individual enhancers and even combinations. Besides testing tissue specificity and enhancer strength,[6,21] reporter assays can be massively parallelized (massively parallel reporter assays, MPRAs) to measure activity of millions of fragments in one experiment and identify enhancer activity in the cell type of interest.[22,23] MPRAs are limited by the size of the analyzed DNA fragment. Therefore, enhancer-promoter communication over large genomic distances cannot be investigated.[24] Furthermore, most genes are controlled by multiple enhancers,[25] and elements with low or undetectable enhancer activity in reporter assays can exert critical regulatory functions at their endogenous loci.[18,20,26,27] While the first studies combined two enhancers to study cooperativity,[28,29]

CellPress
OPEN ACCESS

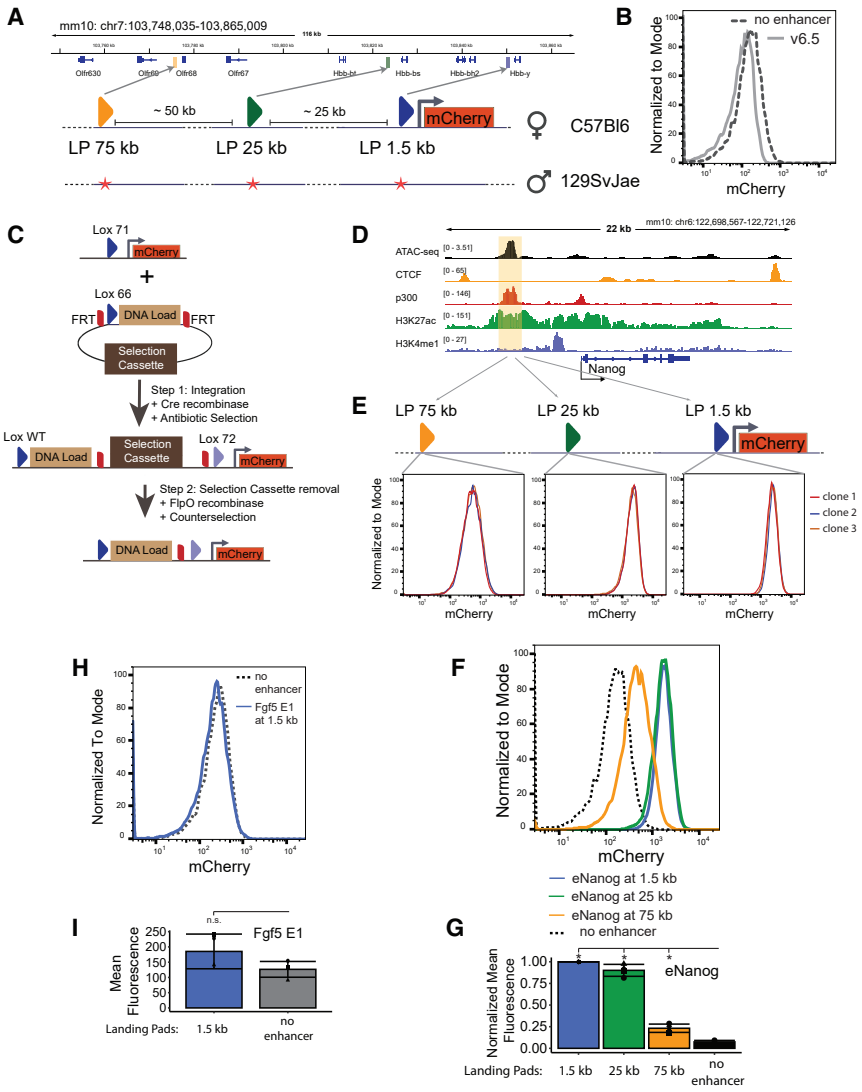


Figure 1. Generation and validation of a synthetic locus to systematically test enhancer activity

(A) Schematic representation of the different components of the synthetic locus generated. Top: UCSC browser track depicting the β-globin locus. The position of the different landing pads (LPs) is indicated. Below: different parts of the synthetic locus including three LP at different distances to the minimal promoter (arrow) and the mCherry reporter gene. All components are integrated on the C57BL6 allele. Bottom: red stars depict single nucleotide polymorphisms (SNPs).

(B) FACS plot showing a selected example of clone 1 no enhancer and the parental cell line v6.5.

(C) Schematic representation of the integration strategy (see text for details).

(D) IGV browser track depicting ATAC-seq,[33] CTCF,[35] p300,[34] H3K27ac,[34] and H3K4me1[34] ChIP-seq tracks for the *Nanog* locus.

(E) Representative FACS plots showing three individual clones with the *Nanog* enhancer (*eNanog*) integrated at the three LPs.

(F) FACS plots of individual examples for mCherry expression of *eNanog* integrated at each distance.

(G) Quantification of mean mCherry expression of *eNanog* integrated at the indicated LPs. All mean expressions were normalized to the 1.5 kb integration. Statistically significant different signals ($p < 0.05$, one-sided paired t tests) are marked by stars.

(H) Representative FACS plots showing an individual clonal cell line with the *Fgf5 E1* enhancer integrated at the 1.5 kb LP.

(I) Quantification of mean mCherry expression of *Fgf5 E1* integrated at the 1.5 kb LP compared with the no-enhancer control. Statistical significance was tested ($p < 0.05$, one-sided paired t tests).

See also Figures S1–S3.

the genomic distance between the elements was negligible. Therefore, new experimental systems are required to study enhancers in a native environment with higher throughput.

Here, we developed a versatile platform for testing enhancer activities and their distance dependencies: a cell line that allows for efficient integration of enhancer sequences at three different distances to a fluorescent reporter gene. This cell line allows building and analyzing complex regulatory landscapes from the bottom up. We compared multiple enhancers selected from different genomic contexts of the mouse genome at different distances to the same promoter. We demonstrate that all enhancers show reduced ability to activate a promoter with increasing distance but to different degrees. Furthermore, combinations of weak and strong enhancers strongly increase expression levels when the weak enhancer is placed between the strong enhancer and the promoter. Thus, synergy between enhancers depends on the individual enhancer activity and on the relative genomic distance between the enhancers and the promoter. Finally, we demonstrate that the interplay of genomic distance, enhancer cooperativity, and CTCF insulation determines gene expression levels.

RESULTS

Design of the reporter system

We aimed to generate a flexible, quantitative reporter system to analyze individual *cis*-regulatory elements and their combinations at different genomic distances from a reporter gene. We introduced the system into the β-globin locus, an inert environment in mouse embryonic stem cells (mESCs)[30] devoid of activating and low on repressive histone marks (Figure S1A[30–35]). We limited the reporter locus to one allele by targeting the C57BL/6 allele of v6.5 mESCs,[36] a C57BL/6 X 129/sv cross with single nucleotide polymorphisms (SNPs) distinguishing the parental alleles (Figure 1A). We inserted an mCherry reporter gene under the control of a minimal TK promoter[37,38] 1.5 kb

from the *Hbb-γ* gene along with three landing pads (LPs) 1.5, 25, and 75 kb upstream of the reporter gene (Figure 1A). After the initial introduction of the reporter gene, we chose two independent clones for the integration of the additional LPs, reaching comparable conclusions with both clones.

mCherry allows for readout of expression levels by fluorescence-activated cell sorting (FACS), and we can obtain information on average expression values and the distribution of expression levels within a population of cells. The final reporter cell lines with the integration of the reporter gene and all three LPs, but without any enhancers, had slightly elevated levels of mCherry fluorescence compared with wild-type (WT) cells (Figures 1B and S1B).

For integrating enhancers at our locus, we used a variation of the Cre/lox recombination system. Recombination between identical lox sites within the genome and a donor plasmid leads to the plasmid integration into the genome. The efficiency is low due to its reversibility but can be increased by using lox66 and lox71 variants,[39,40] which contain mutations that reduce recognition by Cre recombinase after recombination, thereby minimizing excision. We integrated a lox71 site 1.5 kb 5′ of the reporter gene and designed plasmids containing lox66 sites for enhancer integration. Upon recombination, the entire plasmid is integrated into the genome (Figure 1C). To remove the plasmid backbone, we included compatible FRT sites[40] (Figure 1C). Expression of FlpO recombinase excises the entire backbone (Figure 1C), leaving only the enhancer and a single FRT site. For the LPs at 25 and 75 kb, we used different lox variants (here: lox2272-71 and loxm2-71) and designed specific targeting vectors for each LP. These vectors included a compatible lox66 site, a positive selection cassette fused to ΔThymidine Kinase for negative selection, and LP-specific FRT sites distinct from those used in other vectors (see Table S5). We validated that all lox sites and integrations occurred as expected for the two independent clones and multiple cell lines with integrated enhancers using Cas9-seq in combination with Nanopore sequencing.[41]

The e*Nanog* activates gene expression in a distance-dependent manner

We amplified all selected enhancers from the *Mus musculus castaneus* strain to distinguish the integrated from the endogenous enhancer. First, we integrated the proximal enhancer from the *Nanog* locus (Figure 1D) into the targeting plasmid for the 1.5 kb LP. After co-transfection with a plasmid expressing Cre recombinase, roughly 25% of selected colonies had the desired plasmid integration. mCherry expression was slightly reduced upon plasmid integration (Figure S1C, right). We then transfected a plasmid expressing FlpO recombinase and selected for excision of the plasmid backbone with ganciclovir. Now, mCherry expression was strongly activated, and several independently derived cell lines showed almost indistinguishable mCherry levels (Figure 1E).

Interestingly, some clones initially contained few mCherry-negative cells that were quickly lost upon passaging (Figure S1D). We sorted mCherry-positive and -negative cells for one clone and measured mCherry fluorescence by FACS over several passages. While the mCherry-positive population remained stable, the negative population quickly gained mCherry expression (Figure S1D). Initiation of mCherry expression after backbone removal can be delayed but completes within a few passages. Therefore, all FACS analyses in this study were carried out after repeated passaging to ensure full activation.

Next, we integrated the *Nanog* enhancer (*eNanog*) at 25 and 75 kb. Integration efficiencies were similar to integration at 1.5 kb (see Table S5). At both distances, mCherry expression was not increased after initial plasmid integration (Figure S1C, middle and left). Different clones with integrations at 25 or 75 kb showed consistent behavior after backbone excision (Figure 1E, middle and left). We did not observe a fraction of *mCherry*-negative cells at 25 or 75 kb, and expression levels remained stable over several passages. Next, we compared mCherry expression across different e*Nanog* integration distances (Figures 1F and 1G). In a previous study, reporter gene fluorescence was linearly correlated with the number of mRNAs.[3] We performed Flow-fluorescence *in situ* hybridization (Flow-FISH)[42] for mCherry mRNA in cell lines with *eNanog* integrated at 1.5, 25, or 75 kb and the cell line without enhancers (Figure S1E). Both mRNA and protein levels decreased with increasing enhancer-promoter distance. We calculated the mean mCherry fluorescence, subtracting the autofluorescence of unmodified WT cells (Figure 1B), and used this as a direct measure of transcriptional activity. We normalized fluorescence at each LP to the 1.5 kb level to compare activity across distances (Figures 1G and S1F). mCherry expression was slightly reduced upon integrating e*Nanog* at 25 kb compared with 1.5 kb. However, when the enhancer is moved to 75 kb, mCherry expression strongly decreased across the population, though it remained higher than the control (Figures 1G and S1F). In summary, our reporter system allows for the efficient integration of enhancer sequences at three different distances from a reporter gene and confirms the critical role of genomic enhancer-promoter distance in transcriptional regulation.[3,4]

Next, we tested whether increased mCherry expression is due to the integration of active enhancers. *Fgf5 E1* is only activated during the differentiation of mESCs into so-called Epiblast-like cells (EpiLCs), is not actively repressed in mESCs, and has low intrinsic strength in luciferase assays.[18] As expected, integration of the *Fgf5 E1* enhancer at 1.5 kb did not increase mCherry expression beyond the no-enhancer control (Figures 1H, 1I, and S1G).

The β-globin locus is inactive in mESCs; however, individual promoters may be activated by integrating active enhancers. We performed ATAC-seq on selected cell lines with different enhancers and aligned the sequencing data against the relevant custom genomes. First, we analyzed the open chromatin landscape at the *β-globin* and *Nanog* loci, comparing cell lines with and without enhancers to the parental cell lines (Figures S2A and S2B). The profiles at the *Nanog* locus and other naive pluripotency genes were unchanged between v6.5 and the reporter cell lines (Figure S2B). At the *β-globin* locus, integration of the reporter and enhancers caused minimal changes in chromatin accessibility: peak calling identified open chromatin surrounding the integrated reporter and at only one additional peak about 14 kb 3′ of the reporter gene present in all cell lines with the synthetic locus independent of enhancer integration (Figure S2A). Importantly, the reporter

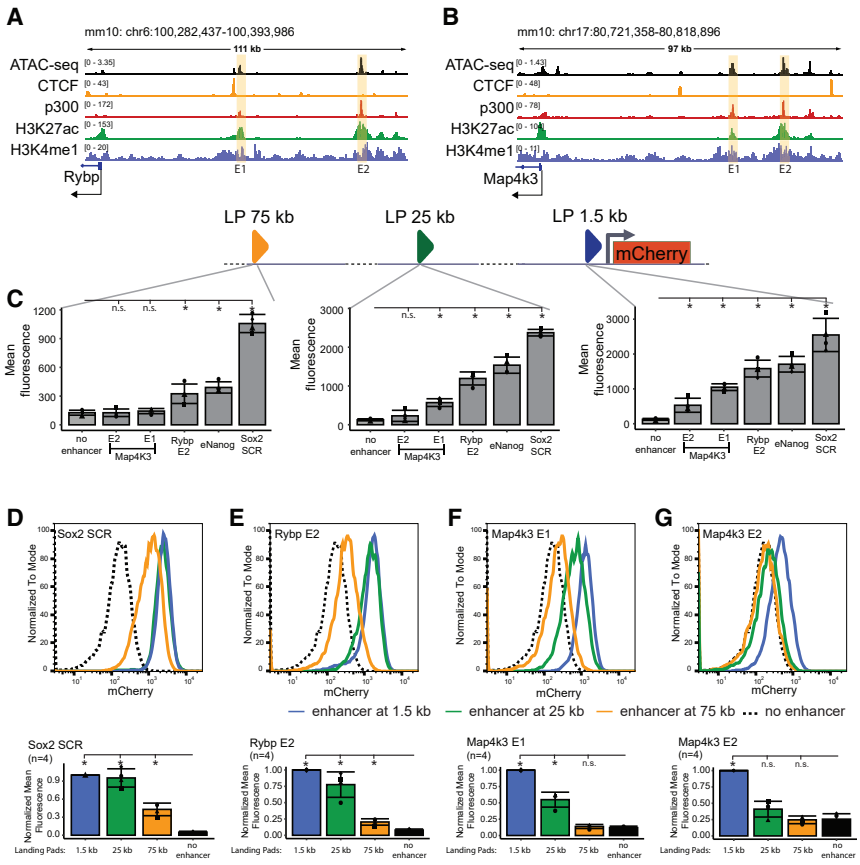


Figure 2. Enhancer strength determines dependency on enhancer-promoter distance

(A) IGV browser track depicting ATAC-seq,[33] CTCF,[35] p300,[34] H3K27ac,[34] and H3K4me1[34] ChIP-seq tracks for the *Rybp* locus.

(B) IGV browser track depicting ATAC-seq,[33] CTCF,[35] p300,[34] H3K27ac,[34] and H3K4me1[34] ChIP-seq tracks for the *Map4k3* locus.

(C) Quantification of mCherry expression of the five indicated enhancers compared with the no-enhancer control at the three different LPs. Left: integration at 75 kb; middle: integration at 25 kb; right: integration at 1.5 kb. Statistically significant signals are marked by stars compared with the no-enhancer control ($p < 0.05$, one-sided paired t tests). n.s.: not significant, $p>0.05$.

(D–G) Quantifying mean mCherry expression of indicated enhancers integrated at the indicated LPs. Top: individual examples; bottom: normalized mCherry expression. All mean expressions were normalized to the 1.5 kb integration. Statistically significant different signals ($p < 0.05$, one-sided paired t tests) are marked.

See also Figure S4.

gene gained accessibility while the endogenous genes remained unaffected. We aligned all sequencing results to custom genomes reflecting the specific enhancer integrations (Figures S2C, S2D, S3C, and S3D). All integrated enhancers gained accessibility at their respective integration sites. Finally, we analyzed the expression of *Hbb-γ* and *Olfr67* in cell lines with a strong enhancer integrated at 1.5, 25, and 75 kb before and after backbone removal (Figures S3A and S3B). While *Olfr67* is very lowly expressed and showed slight increase in expression upon integration of a strong enhancer at 25 kb, *Hbb-γ* levels increased with integration of a strong enhancer, correlating with activation of the reporter gene. However, the results were not statistically significant due to low and variable expression of *Hbb-γ*. In summary, while we observed minor differences after integrating the reporter and enhancers, the manipulations did not alter cell identity, and the minimal TK promoter is the main promoter activated at the synthetic locus.

Taken together, we have generated a highly versatile synthetic locus that allows for the reproducible interrogation of enhancer activity at defined distances to a reporter gene. The locus itself is mostly inert, and reporter gene expression is only activated by active enhancers.

Enhancer strength determines dependency on enhancer-promoter distance

Recently, both the *Sox2* control region (SCR) from the *Sox2* locus in mESCs and the human β-globin micro-LCR in erythroleukemia K562 cells have been demonstrated to depend on the genomic distance to the promoter in model loci.[3,4] Both studies used highly active enhancers, so it remains unclear if all enhancers respond similarly to genomic distance. The high efficiency of enhancer integration into our reporter system enabled us to systematically compare how different enhancers are affected by genomic distance. We selected enhancers from multi-enhancer clusters with varying strengths.[43] We included e*Nanog* (Figure 1[44]) the E2 enhancer element located 90 kb upstream of the *Rybp* locus (Figure 2A), and two enhancers (*E1* and *E2*) from the *Map4k3* locus (60 and 75 kb upstream of the transcription start site [TSS], Figure 2B). All enhancers activated luciferase in a plasmid-based assay (Figure S4A), albeit to different degrees: e*Nanog* and *Rybp E2* were the strongest activators, followed by *Map4k3 E2* and *Map4k3 E1*. We also included the well-studied SCR, a 6 kb multi-enhancer cluster roughly 100 kb upstream of the *Sox2* gene[3,45] (see also Figure S5A).

Next, we integrated each enhancer into each LP and measured mCherry levels using FACS (Figures 2C and S4B). At 1.5 kb, all enhancers increased *mCherry* expression, with the SCR being the strongest activator, followed by e*Nanog, Rybp* E2, *Map4k3* E1, and *Map4k3* E2 (Figures 2C, right; and S4B, right). The expression levels of mCherry were reproducible between different clones and replicates, though we observed more variability in lower-expressing cell lines with *Map4k3* E1 or *Map4k3* E2 integration. Despite higher activity in luciferase assays, *mCherry* levels were lower with *Map4k3* E2 integration compared to *Map4k3* E1 (Figure 2C). We then assessed mCherry expression when the same enhancers were integrated at either 25 kb (Figures 2C, middle; and S4B, middle) or 75 kb (Figures 2C, left; and S4B, left). At all distances, the order of relative strength of the enhancers remained consistent:

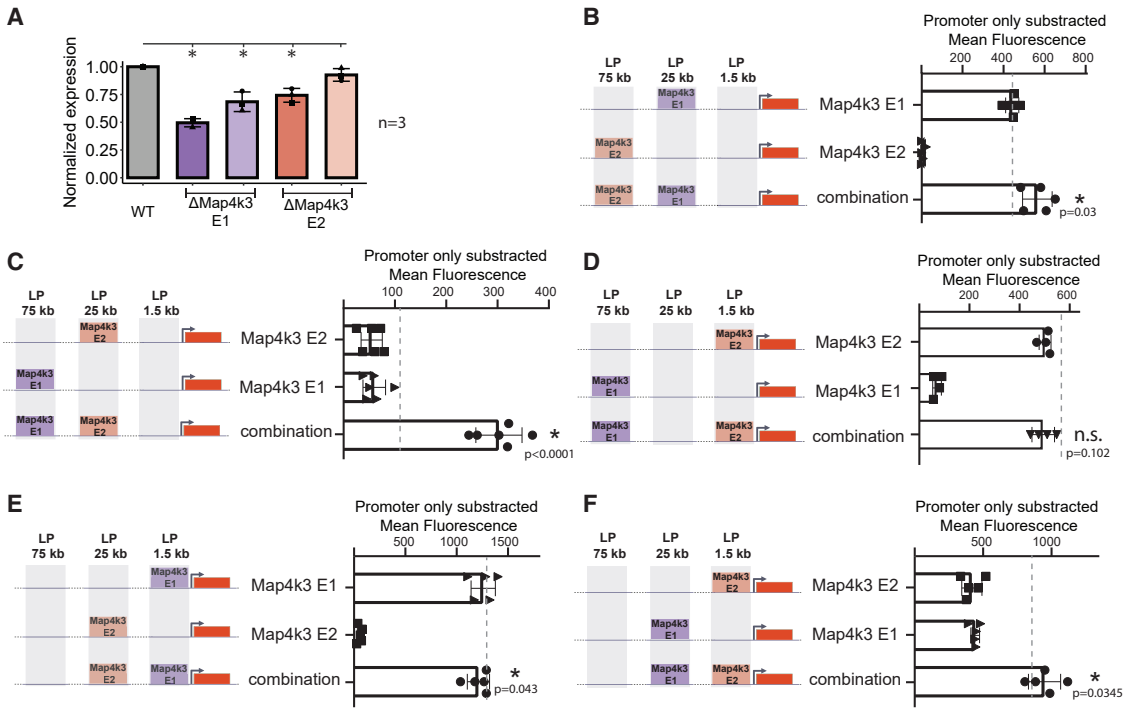


Figure 3. Weak enhancers activate transcription synergistically from a distance

(A) Quantification of *Map4k3* expression after deletion of the two different individual enhancers compared with the WT expression.

(B–F) Comparison of combinations of enhancers to the individual integrations and the expected additive behavior. Left: schematic of individual and combinations of integrations at the different LPs. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. The gray dashed line indicates the expected values under an additive model. Statistically significant signals compared with the expected additive model ($p < 0.05$, one-sided paired t tests) are marked by stars or n.s. (not significant). (B) *Map4k3 E2* at 75 kb combined with *Map4k3 E1* at 25 kb. (C) *Map4k3 E1* at 75 kb combined with *Map4k3 E2* at 25 kb. (D) *Map4k3 E1* at 75 kb combined with *Map4k3 E2* at 1.5 kb. (E) *Map4k3 E2* at 25 kb combined with *Map4k3 E1* at 1.5 kb. (F) *Map4k3 E1* at 25 kb combined with *Map4k3 E2* at 1.5 kb.

See also Figure S4.

SCR was the strongest, followed by e*Nanog*, *Rybp* E2, *Map4k3* E1, and finally *Map4k3* E2.

While the order in enhancer strength was preserved, mCherry expression decreased with increasing distances (Figures 2C and S4B). The strong enhancers *SCR* (Figures 2D and S4C), *eNanog* (Figures 1G and S1D), and *Rybp E2* (Figures 2E and S4D) showed strong activation at 1.5 and 25 kb, with no (*SCR*) or only slight loss (*eNanog* and *Rybp E2*) of activity at the intermediate distance 25 kb. In comparison, at 75 kb, mCherry expression is substantially reduced but still detectable. For the weaker enhancers, *Map4k3 E1* (Figures 2F and S4E) and *Map4k3 E2* (Figures 2G and S4F), activation from 25 kb was greatly reduced and almost undetectable at 75 kb. Of note, neither *Map4k3 E1* at 75 kb (Figures 2C and 2F) nor *Map4k3 E2* at 25 kb (Figures 2C and 2G) in clone 1 reached statistical significance. However, with increasing replicates (see Figures 3 and 4), both enhancers reached statistical significance at these loci. Nevertheless, the measured activity was very low, and the relative loss of activity compared with 1.5 kb was more pronounced for the weaker enhancers (Figures 2C and S4B). In summary, increasing distance to the minimal promoter decreases the ability of an enhancer to activate that promoter. Still, stronger enhancers can activate this promoter from larger distances.

The cooperation of weak enhancers can activate transcription from a distance

Map4k3 E2 could not elicit considerable mCherry expression from 25 kb. At its endogenous locus, this element is located about 75 kb from its potential target gene, *Map4k3*. We tested whether *Map4k3 E2* contributed to *Map4k3* expression. We deleted both *Map4k3* enhancers individually at the endogenous locus using CRISPR-Cas9, selected two clones each, and analyzed *Map4k3* expression using qPCR. Deletion of *E1* clearly reduced *Map4k3* expression in both analyzed clones (Figure 3A). Deletion of the *E2* enhancer also reduced the expression of the target gene (Figure 3A), albeit not as strongly as the *E1* enhancer and only reaching significance in one clone.

As the weak enhancer *Map4k3 E2* affects expression levels at its endogenous locus from a distance where it is inactive in our synthetic locus, we hypothesized that the presence of additional enhancers might influence how a given enhancer is affected by genomic distance. Therefore, we integrated both *Map4k3 E1* and *E2* elements step-wise at different LPs into our synthetic locus: first integrating one enhancer, selecting individual clones, and then integrating the second enhancer, followed by backbone removal at both integration sites. We used clone 1 for all dual enhancer experiments but validated

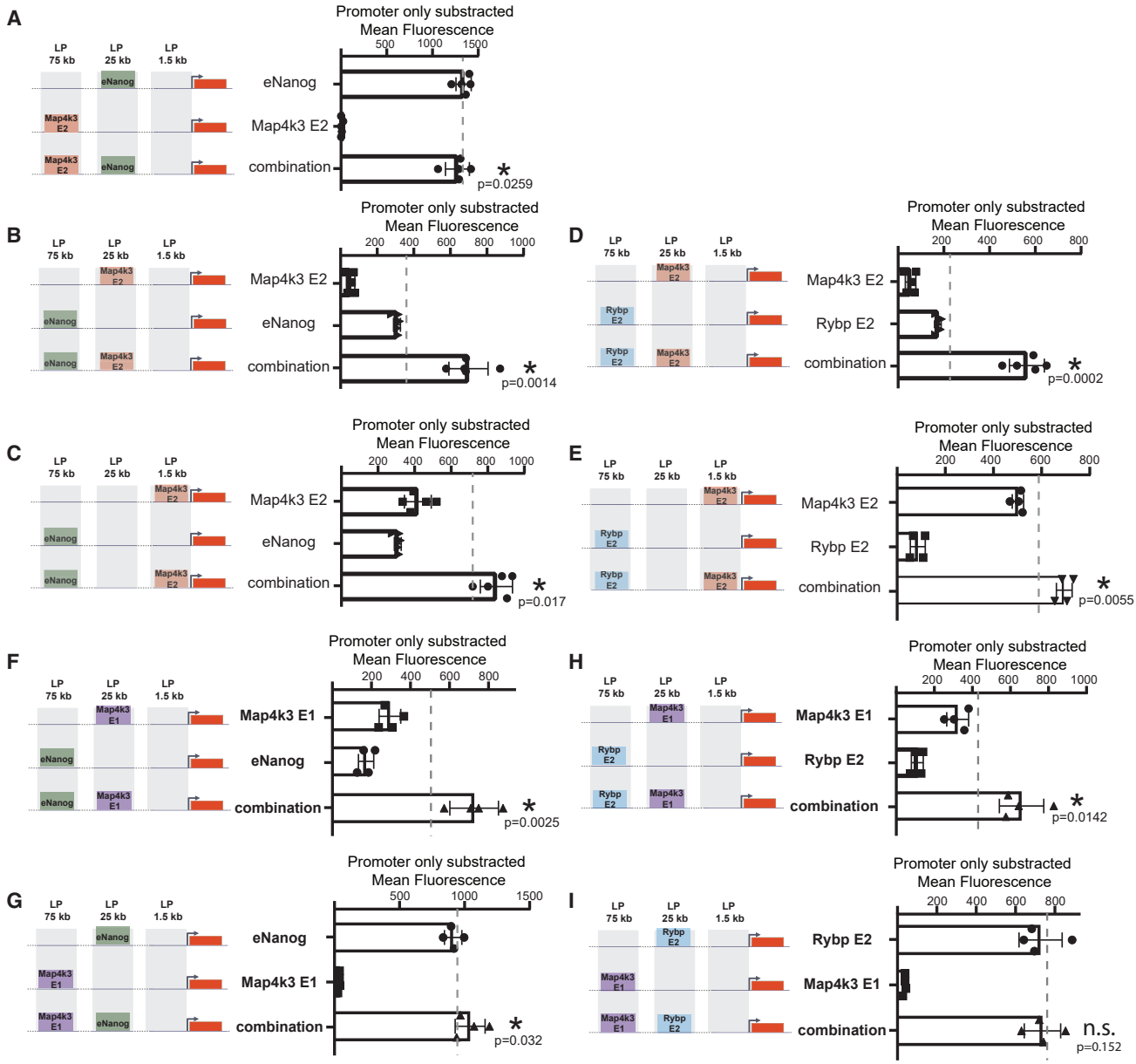


Figure 4. *Map4k3 E1 and E2* can cooperate with strong enhancers to activate transcription from a distance

Comparison of combinations of enhancers (third row) to the individual integrations (top two rows) and the expected additive behavior. Left: schematic of individual and combinations of integrations at the different LPs. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. The gray dashed line indicates the expected values under an additive model. Statistically significant signals compared with the expected additive model ($p < 0.05$, one-sided paired t tests) are marked by stars or n.s. (not significant). (A) *Map4k3 E2* at 75 kb combined with *eNanog* at 25 kb. (B) *eNanog* at 75 kb combined with *Map4k3 E2* at 25 kb. (C) *eNanog* at 75 kb combined with *Map4k3 E2* at 1.5 kb. (D) *Rybp E2* at 75 kb combined with *Map4k3 E2* at 25 kb. (E) *Rybp E2* at 75 kb combined with *Map4k3 E2* at 1.5 kb. (F) *eNanog* at 75 kb combined with *Map4K3 E1* at 25 kb. (G) *Map4k3 E1* at 75 kb combined with *eNanog* at 25 kb. (H) *Rybp E2* at 75 kb combined with *Map4k3 E1* at 25 kb. (I) *Map4K3 E1* at 75 kb combined with *Rybp E2* at 25 kb.

See also Figure S4.

selected combinations in clone 2. To compare dual to individual integrations to assess enhancer cooperativity, we calculated the expected activity for enhancer combinations based on an additive model (gray dotted lines), summing the individual enhancer contributions to predict expression. We then tested if the observed expression was significantly higher than the expected value. We subtracted the baseline *mCherry* expression from the no-enhancer control to avoid counting basic promoter activity twice when calculating the expected activity for dual integrations.

First, we tested whether *Map4k3 E2* at 75 kb increased expression levels when combined with *Map4k3 E1* at 25 kb. Both elements together led to increased mCherry expression compared with the individual integrations (Figure 3B). Next, we swapped their positions: placing *Map4k3 E1* at 75 kb, and *Map4k3 E2* at 25 kb. While both enhancers showed minimal activity at these distances, their combination increased *mCherry* expression super-additively (Figures 3C and S4G). Hence, weak enhancers can synergize to activate transcription from distances where they are individually hardly active. Next, we placed *Map4k3 E2* closer to the promoter. Integrating *Map4k3* E2 at 1.5 kb and *Map4k3* E1 at 75 kb, super-additivity was not observed, and *mCherry* expression was similar to the level of *Map4k3* E2 alone at 1.5 kb (Figure 3D).

Next, we moved both elements closer to the promoter. Placing the stronger *Map4k3* E1 at 1.5 kb and the weaker *Map4k3* E2 at 25 kb did not further increase *mCherry* expression (Figure 3E). When we reversed the positions, with *Map4k3* E2 at 1.5 kb and *Map4k3* E1 at 25 kb, their combination slightly exceeded the expected additive behavior (Figure 3F).

Cooperation between strong and weak enhancers depends on their relative position

We next explored whether *Map4k3* E2 could cooperate with enhancers from other loci and combined *Map4k3 E2* with e*Nanog*. When *eNanog* was integrated at 25 kb, *Map4k3 E2* at 75 kb did not increase the expression of mCherry further (Figure 4A). However, placing e*Nanog* at 75 kb and *Map4k3 E2* at 25 kb resulted in strong super-additive mCherry expression (Figures 4B and S4H). Moving *Map4k3 E2* to 1.5 kb with e*Nanog* at 75 kb increased mCherry expression slightly above the combined individual enhancer levels (Figure 4C). The cooperative effect of *Map4k3 E2* was not limited to e*Nanog*: combining *Map4k3* E2 at 25 kb with *Rybp* E2 at 75 kb also produced super-additive *mCherry* expression (Figures 4D and S4I). Furthermore, placing *Map4k3* E2 at 1.5 kb and *Rybp* E2 at 75 kb yielded only minor increases above additive levels (Figure 4E).

Next, we combined *Map4k3 E1* at 25 kb with a stronger enhancer and integrated either *eNanog* or *Rybp E2* at 75 kb. In both cases, the enhancer combinations yielded stronger mCherry expression than the sum of their individual effects (Figures 4F and 4H). Similar to *Map4k3 E2*, the order of enhancers mattered: placing the weaker *Map4k3* E1 at 75 kb and the stronger enhancer at 25 kb did not yield a strong synergistic effect (Figures 4G and 4I). In summary, weak enhancers such as *Map4k3 E2 or Map4k3 E1* can synergize with strong enhancers, partially alleviating the distance-dependent drop in activation of the strong enhancer. This behavior somewhat depends on the weaker enhancer's position: closer integrations to the promoter produce additive-like expressions, while more distal combinations yield synergistic levels.

Map4k3 E1 and *E2* exemplify two weak enhancers with strong synergistic interactions, both with each other and with other enhancers. However, it remains unclear if all weak enhancers can cooperate with strong enhancers from a distance. Therefore, we selected four additional weak enhancers: *Rybp E1* (Figure 2A) and three well-described enhancers from the *Sox2* locus: *Sox2 E15*, *Sox2 E19*, and *Sox2 E20* (Figure S5A). *Sox2 E15* is a weak enhancer with a described role in transcriptional bursting of *Sox2*.[46] *Sox2 E19* and *E20* are potential facilitator elements, lacking intrinsic enhancer strength but capable of supporting *Sox2* expression.[19] First, we tested each enhancer's activity at 1.5 kb and at 25 kb (Figure 5A). *Sox2 E19* showed very low activity even at 1.5 kb. All other elements showed reduced mCherry expression at 25 kb compared with 1.5 kb. Notably, *Rybp E1* and *Sox2 E20* showed interesting behavior: both enhancers show lower mCherry expression at 1.5 kb than *Map4k3 E2* (Figure 5A), but the mCherry expression at 25 kb only drops by about 25% in the case of *Rybp E1* (Figure 5B), whereas for *Map4k3 E2* the drop is around 75% on average (Figure 2G). Therefore, the decrease in expression for *Rybp E1* and *Sox2 E20* from 25 kb compared with 1.5 kb was much shallower than expected for their respective enhancer strength at 1.5 kb (Figure 5B). The distance-dependent decrease in expression may be influenced not only by intrinsic enhancer strength but also by specific characteristics of each genomic element.

We next tested the cooperative behavior of these additional weak enhancers by integrating each element at 25 kb in cell lines with *eNanog* at 75 kb (Figures 5C–5F). In all these cell lines, we observed synergistic expression levels. Finally, we examined whether synergism is limited to weak enhancers by combining *eNanog* at 75 kb with either the *Rybp E2* or a second e*Nanog* at 25 kb (Figures S5B and S5C). *eNanog* at 25 kb alone produced higher levels of mCherry expression than *Rybp E2* at the same position (compare Figures S5B, S5C, and 2C, middle). Interestingly, the *mCherry* expression from *Rybp* E2 and *eNanog* combined exceeded the expected additive levels (Figure S5D), whereas combining two e*Nanog* at 75 and 25 kb did not produce a synergistic effect (Figure S5C). To test the limits of our minimal TK promoter, we combined *eNanog* and *Rybp E2* at 1.5 and 25 kb. The resulting gene expression was sub-additive, regardless of the order of the enhancers (Figures S5D and S5E). Although this combination produced stronger expression than two distal e*Nanog*, it was still lower than the *SCR* at 1.5 or 25 kb (Figure 2C). Therefore, promoter saturation may not function as a simple fixed threshold when combining strong enhancers (see discussion).

In summary, all tested weak enhancers can synergize with a distal strong enhancer, increasing expression at the target promoter beyond the additive levels of individual enhancer activities.

Two recent publications examined enhancer synergies using episomal plasmids but reached different conclusions[28,29]: in mESCs, enhancers worked predominantly additive,[29] while in *Drosophila* S2 cells most developmental enhancers were super-additive following a simple multiplicative model, with the caveat that strong enhancers might saturate limited promoter capacity.[28] We compared these models of enhancer cooperativity to explain our experimental results.

Multiple independently derived clones of individual and combinations of enhancers exhibited consistent expression changes. To assess variability, we first compared $\log_2$ fold changes to the no-enhancer control across clones derived after integration (step 1) and after backbone removal (step 2). Clones with the same enhancer integration showed exceptional reproducibility (Figure S5F), allowing us to pool data from all individual

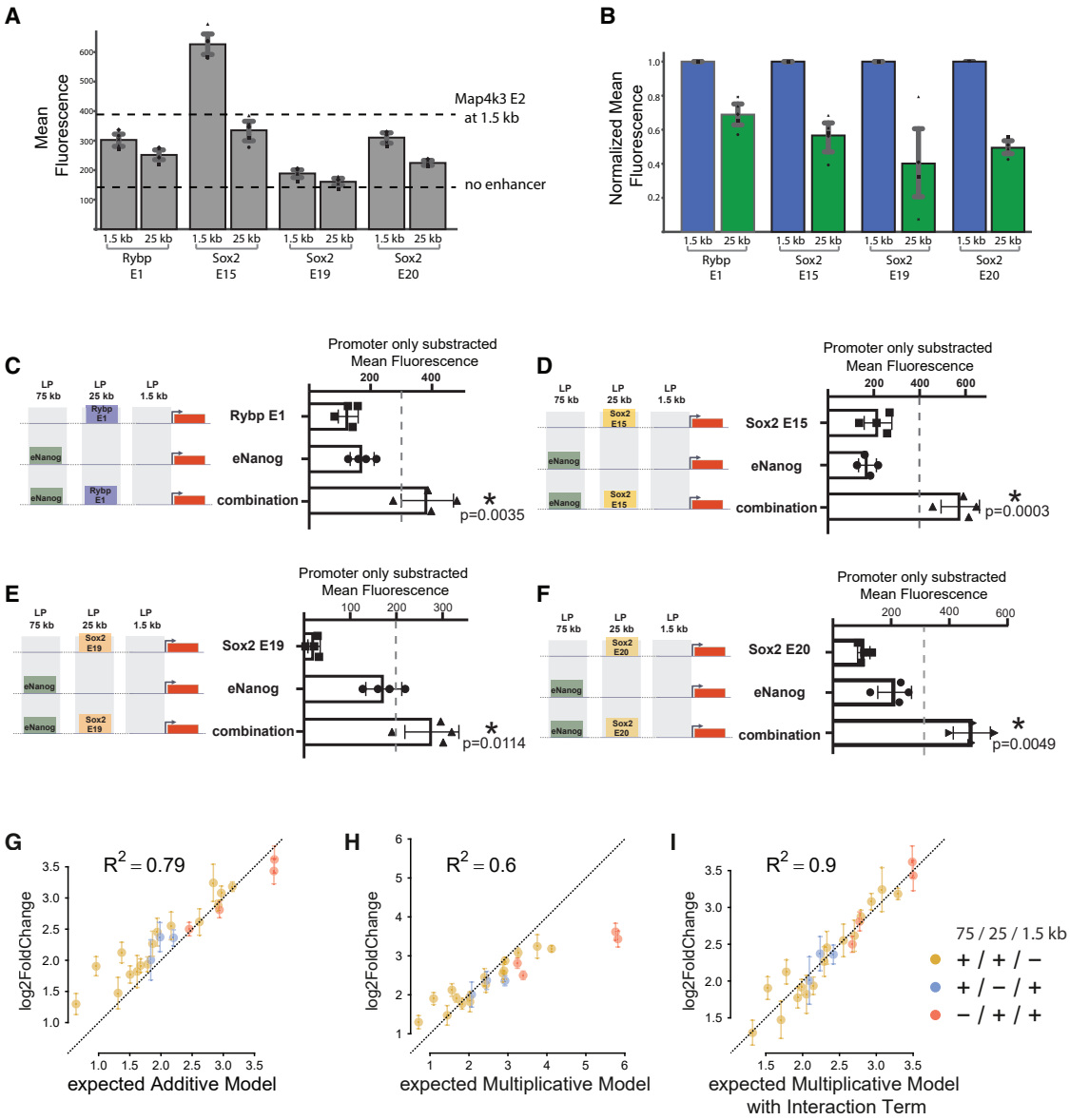


Figure 5. Weak enhancers cooperate with strong enhancers at distal positions

(A) Quantification of mCherry expression of the four indicated enhancers at the two indicated different LPs 1.5 and 25 kb. Both no enhancer and *Map4k3 E2* at 1.5 kb mCherry levels are indicated with dashed gray lines.

(B) Normalized mean mCherry expression of indicated enhancers integrated at the indicated LPs. All mean expressions were normalized to the 1.5 kb integration.

(C–F) Comparison of combinations of enhancers (third row) to the individual integrations (top two rows) and the expected additive behavior. Left: schematic of individual and combinations of integrations at the different LPs. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. The gray dashed line indicates the expected values under an additive model. Statistically significant signals compared with the expected additive model ($p < 0.05$, one-sided paired t tests) are marked by stars or n.s. (not significant).

(C) *eNanog* at 75 kb combined with *Rybp E1* at 25 kb. (D) *eNanog* at 75 kb combined with *Sox2 E15* at 25 kb. (E) *eNanog* at 75 kb combined with *Sox2 E19* at 25 kb. (F) *eNanog* at 75 kb combined with *Sox2 E20* at 25 kb. (G) Additive model of enhancer activity. (H) Multiplicative model of enhancer activity. (I) Multiplicative model with interaction term to predict enhancer activity.

See also Figure S5.

clones. We then tested if additive or multiplicative models could explain our results. The simple additive model predicts combined activity by summing individual enhancer activities (in linear gene expression space) and accurately described combinations when one enhancer is positioned proximal to the promoter (Figure 5G). However, many distal enhancer combinations deviated from this pattern, showing super-additive behavior (Figures 5G, 3, 4, and 5C–5F). A simple multiplicative model revealed a poor

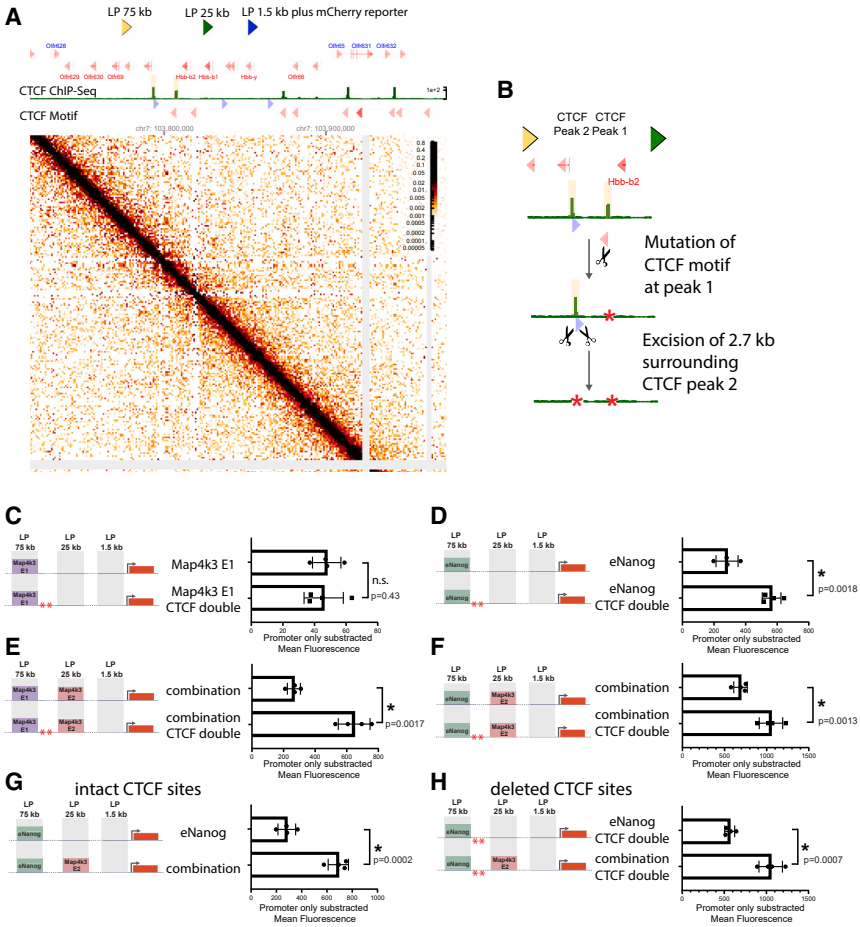


Figure 6. CTCF sites between the most distal LP and the promoter reduce activity of strong enhancers and enhancer combinations

(A) Top: positions of the LPs at the β-globin locus. Middle: ChIP-seq against CTCF[35] and direction of CTCF motifs; Bottom: MicroC data[48] across the β-globin locus.

(B) Strategy to eliminate both CTCF sites. Guide RNA against peak 1 binds in the CTCF motif of peak 1. In a second step, peak 2 was deleted with two guide RNAs binding about 3 kb apart from each other.

(C–H) Comparison of mCherry expression before and after deletion of both CTCF sites. Left: schematic of individual or combinations of integrations at the different LPs and the deletion of the CTCF sites. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. Statistically significant signals ($p < 0.05$, one-sided paired t tests) are marked by stars or n.s. (not significant). (C) *Map4k3 E1* integrated at 75 kb with and without CTCF motifs. (D) *eNanog* integrated at 75 kb with and without CTCF motifs. (E) *Map4k3 E1* integrated at 75 kb and *Map4k3 E2* integrated at 25 kb with and without CTCF motifs.(F) *eNanog* integrated at 75 kb and *Map4k3 E2* integrated at 25 kb with and without CTCF motifs. (G) Comparison of mCherry expression with *eNanog* at 75 kb without and with *Map4k3 E2* at 25 kb. (H) Comparison of mCherry expression with *eNanog* at 75 kb without and with *Map4k3 E2* at 25 kb after deletion of the CTCF motifs.

See also Figure S6.

fit: distal enhancer combinations showed higher activity than expected for multiplicative behavior, while combinations with a proximal enhancer yielded lower than predicted activity (Figure 5H). Neither the simple additive nor the simple multiplicative models alone accurately predicted combined activities. However, incorporating interaction terms into the multiplicative model significantly improved accuracy (R^2 = 0.90, Figure 5I). In this refined model, the individual activity of the two integrated enhancers is weighted with an interaction term, which considers how the activity of one enhancer modifies or scales the activity of the other. This model highlighted a linear relationship between individual and combined activities, allowing predictions based on individual activities at different integration sites. While each integration site contributed positively to the overall activity, the model revealed a significant negative interaction for enhancers placed closest to the promoter (−0.126), indicating that placing an enhancer at these proximal sites can reduce the combined effect (see Table S6).

In conclusion, how two enhancers cooperate depends strongly on their genomic distance to the promoter and their individual activity from that location, but strong individual activity close to the promoter is detrimental for strong synergistic behavior, suggesting that promoter saturation may limit stronger super-additive outcomes.[28]

CTCF sites between the most distal LP and the promoter reduce activity of strong enhancers and enhancer combinations

CTCF sites between enhancer and promoter can reduce target gene expression.[3,10,47] At the *β-globin* locus, two CTCF sites are positioned between the *Hbb-bt* and the *Olfr67* gene, located near the 75 and the 25 kb LPs (Figure 6A). In mESCs, this locus is inactive with no notable 3D genome structure surrounding the integration sites in the WT cells (Figure 6A, compare with the *Nanog* locus, Figure S6A). However, these CTCF sites might influence enhancer-driven gene expression at the locus. We devised a two-step strategy to remove the CTCF sites from existing enhancer cell lines (Figure 6B). We used a single guide RNA to target the CTCF motif near the 25 kb LP (peak 1). The second CTCF site (peak 2), located in a highly repetitive region, required deletion of a 3 kb genomic region. First, we examined whether removing CTCF sites would restore the activity of a weak enhancer, *Map4k3* E1, at 75 kb. Neither deletion of peak 1 alone (Figure S6B) nor removal of both peaks increased *mCherry* expression through *Map4k3 E1* (Figure 6C). We then tested the strong *eNanog* at 75 kb, observing a slight increase in *mCherry* expression upon single CTCF deletion (Figure S6C), which increased after removing both peaks (Figure 6D). We next examined enhancer combinations: both *Map4k3 E1* or

eNanog at 75 kb showed synergy with *Map4k3 E2* at 25 kb. In both cases, overall expression of mCherry is slightly increased by mutation of peak 1 (Figures S6D and S6E), but much stronger increased upon removal of both CTCF peaks (Figures 6E and 6F). Finally, we compared the synergy between *eNanog* and *Map4k3 E2* with intact and disrupted CTCF sites. In both cases, the integration of the *Map4k3 E2* enhancer in combination with *eNanog* at 75 kb leads to increased mCherry expression. Finally, we compared the synergy between *eNanog* and *Map4k3 E2* with intact and disrupted CTCF sites. In both cases, the integration of the *Map4k3 E2* enhancer in combination with *eNanog* at 75 kb leads to increased mCherry expression. Taken together, the CTCF peak 1 has a minor negative effect on gene expression; however, removal of both CTCF peaks leads to context-dependent increases in gene expression at the promoter.

DISCUSSION

We developed a platform to study distance-dependent enhancer-enhancer cooperativity in the mammalian genome. Our flexible platform can accommodate different sequences at distinct positions, testing enhancers in their native chromatinized setting. As the locus contains no active regulatory elements besides the ones introduced, it provides a controlled environment for building and analyzing complex regulatory landscapes *de novo*.

Our system uncovered key principles of enhancer cooperativity and the impact of genomic distance on individual enhancers and their combinations. We found that increased genomic distance does not equally affect all enhancers: stronger enhancers generally retain more activation potential at greater distances. However, intrinsic strength alone cannot fully predict the activity drop when enhancers are integrated distally, suggesting additional mechanisms at play. Our findings also illuminate how genomic distance influences enhancer cooperativity. Combining a strong distal enhancer with a weaker enhancer at 25 kb consistently led to super-additive activation, while placing the strong enhancer closer to the promoter rarely produced this effect. Notably, *Map4k3 E1* can act as either a strong distal enhancer with *Map4k3* E2 or as a weaker enhancer at 25 kb with *Nanog* or *Rybp* E2, highlighting that relative strength between enhancers might matter more than absolute activity. Moving a weak enhancer closer to the promoter diminishes or eliminates super-additivity. This indicates that enhancer cooperation depends on both intrinsic strength and relative activity at specific distances, which may explain the variability in enhancer cooperativity modes at endogenous loci reported previously. Finally, how enhancers are affected by CTCF sites depends on the enhancer strength, as well as on the overall enhancer landscape.

Our results align with and expand on previous studies of distance-dependent regulation.[3,4] While increasing the enhancer-promoter distance generally reduces activation, many enhancers are positioned at distances where they should not activate a promoter individually. Interrogating multiple weak enhancers from three different genomic loci (*Map4k3*, *Rybp*, and *Sox2*) suggests that addition of weak enhancers can partially overcome decreased enhancer activity from increased genomic distances. Individually, weak enhancers cannot activate transcription strongly even from intermediate distances. However, placed between a strong enhancer and a promoter, they significantly boost promoter activation. This might explain recent findings from the *α-globin*[20] and the *Sox2* loci,[19] where weak enhancers help strong enhancers communicate with the promoter. However, we cannot rule out that these facilitators apply distinct mechanisms of enhancer cooperation.

Why can stronger enhancers bridge more considerable genomic distance than weaker ones, and how do weak enhancers synergize with distal strong enhancers? Increased genomic distance reduces enhancer-promoter contact frequency,[49,50] but contacts do not directly correlate with transcriptional activation,[3] and it remains unclear whether and how close enhancers need to come to their promoter to activate transcription.[51–53] Regulatory elements form large "transcription hubs," aggregating TFs, co-factors, and RNA polymerase II (RNA Pol II), which may bridge enhancers to promoters.[54,55] Actual 3D distance might matter less if the transcription hub still "touches" the promoter.[56] Stronger enhancers may form more substantial hubs, allowing them to activate from larger distances and making them less sensitive to small increases in 3D distance. This model could explain why the SCR maintains high expression from both 1.5 and 25 kb, unlike weaker enhancers. At large distances, strong enhancers may only intermittently contact the promoter via their hub, reducing gene expression. Combining strong distal enhancers with weak ones at intermediate distances could create larger hubs that increase promoter contact frequency, leading to super-additivity. Conversely, moving the weak enhancer closer to the promoter may allow its hub to contact the promoter independently, reducing the benefit of the strong distal enhancer and lowering super-additivity. However, intrinsic strength alone does not fully explain distance-dependent activation loss. Our comparison of weak enhancers suggests that distance sensitivity is characteristic of each enhancer, likely influenced by sequence composition and associated factor recruitment.

An alternative explanation could be provided by how promoters integrate enhancer input. Previous studies suggest that expression levels follow a sigmoidal curve relative to enhancer contact frequency[3,57]: at low contact frequency and expression levels, even large increases in contact produce only small increases in expression, whereas at intermediate contact frequencies, the same increase results in large jumps in expression levels. At high frequencies, further contact increases yield minimal expression gains. This behavior has been attributed to multiple transitions that a promoter has to undergo while in contact with the enhancer in order to be activated.[3,57] Sigmoidal integration of enhancer input into expression levels by promoters might be a general feature of transcriptional regulation that can explain our findings[58]: stronger enhancers might increase the rate of promoter transitions more than weaker enhancers. This could explain why strong enhancers are less affected by increased genomic distance than weaker ones. At the same time, introducing a weak enhancer alone at intermediate distances would have no effect since the promoter would be in the regime of the curve where big changes are required for increasing gene expression, but

adding a weak enhancer to a strong distal enhancer—where the promoter is in a more responsive, intermediate regime due to the presence of the distal enhancer—could lead to strong activation of gene expression. Such a model describes our results reasonably well and might explain why the strongest cooperativity arises when combining enhancers at distances where they have rather low individual activity. The sigmoidal model also suggests that high expression levels require increasing amounts of enhancer input, which might explain reduced or sub-additive effects observed for close enhancer combinations or for strong enhancers at 1.5 kb, similar to observations in the fly.[11] The response curve might vary by promoter; hence, some of the enhancer combinations might show different behavior when combined with a stronger promoter. Here, the promoter might be rate limiting or "saturated," even though drastically increasing enhancer input (e.g., by taking the strong SCR) can still push expression levels further. To fully assess whether a sigmoidal response curve can explain distance-dependent enhancer interactions, further studies with diverse enhancers, promoters, precise contact measurements, and mathematical modeling will be needed.

Gene expression regulation at the target locus depends on the interplay of promoters, enhancers, CTCF sites, and other regulatory elements. While enhancer-promoter distance remains the primary factor, multiple CTCF sites at our locus reduce gene expression without completely blocking it. CTCF may act as a boundary or insulator, hindering communication with the target gene. Interestingly, these CTCF sites do not form a detectable structure in WT lacking integrated reporters or enhancers, suggesting minimal loop extrusion in this region without active regulatory elements. Future studies on 3D contact changes will help clarify the emergence of potential new loops.

Previously, we demonstrated that elements lacking intrinsic activity can still contribute to gene expression.[18] Here, we expand on this by showing that weak enhancers positioned between a strong enhancer and the promoter can significantly enhance target gene expression. A recent MPRA study in human cell lines categorized *cis*-regulatory elements as classical enhancers or chromatin-dependent enhancers,[59] the latter marked by strong epigenomic signatures but low intrinsic activity. Both types often coexist within enhancer clusters, where they may serve distinct roles. Multiple weak enhancers could provide a buffering effect against the loss of individual elements. Additionally, combining weak and strong enhancers may increase the genomic distance range at which enhancer clusters can activate their target genes.[19,20] Moreover, we speculate that enhancer cooperativity could improve efficiency in target promoter selection: promoters associated with an additional weak enhancer may exhibit an increased responsiveness to a distal strong enhancer compared with other promoters without such an element.

Limitations

While mCherry levels were similar between independently generated cell lines, protein levels are not a direct readout of transcription. For selected enhancers the change in mRNA levels corresponded well with protein levels, however, other *cis*-regulatory elements could in principle also affect mRNA processing or half-life, in which case protein and transcription levels will not correlate well.

We have built the synthetic locus around the minimal TK promoter that might be rate limiting: most strong enhancers and combinations at close distances showed similar mean expression levels. In addition, strong activity from close distances penalized against a strong synergistic behavior in our linear model. The incorporation of different promoters in the future will test the limitations of the system and expand our toolbox. Similarly, we used the β-globin locus as our neutral environment and limited the tested distance to 75 kb.

In this proof-of-concept study, we analyzed a limited set of individual enhancers that showed very similar behavior; therefore, it is tempting to speculate that the super-additive cooperation of different enhancer elements is a general feature in transcriptional regulation. However, other, thus far untested elements could not follow this same trend, might antagonize each other, or simply not be compatible with each other.

Finally, we are currently limited to mere speculation regarding the molecular nature of enhancer cooperativity.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Christa Buecker (christa.buecker@maxperutzlabs.ac.at).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

- All ATAC-seq datasets generated in this study have been deposited at ENA with the accession number PRJEB80727. FACS data have been deposited to Mendeley Data, and the DOI is reported in the key resources table. They are publicly available as of the date of publication.
- This study does not report any original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We would like to thank all members of the Buecker lab, Martin Leeb, and the members of his lab. In addition, we would like to thank Patricia Rothe and Mustafa Alaboo for technical help and Anton Goloborodko and Barak Cohen for insightful discussions. The BioOptics-FACS facility at the Max Perutz labs was instrumental in the success of this project. Cas9-seq and ATAC-seq were performed by the Next Generation Sequencing facility at the Vienna BioCenter Core Facilities (VBCF), a member of the Vienna BioCenter (VBC). The research was supported by the Austrian Science Fund FWF (P30599, P34123, and PAT9017923 to C.B., SMICH). Research at the IMP is supported by Boehringer Ingelheim GmbH and the Austrian Research Promotion Agency (FFG, FO999902549).

AUTHOR CONTRIBUTIONS

H.T. and C.B. designed experiments, analyzed results, and wrote the manuscript with the help from all authors. H.T. and S.F. performed the experimental work and analyzed results with the support of M.H. and D.V. F.H. generated and analyzed additional weak enhancer cell lines. M.G.G. performed

Flow-FISH. M.P. performed Cas9-seq analysis. V.L. and A.S. performed the modeling.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - General culture conditions of mouse embryonic stem cells
  - Generation of reporter cell line
- METHOD DETAILS
  - Design of plasmids for enhancer integration
  - Integration of enhancer sequences
  - Cas9-targeted sequencing
  - FACS analysis of mCherry expression
  - FACS sorting of mCherry-populations
  - Luciferase assays
  - Enhancer KO and RT-qPCR
  - CTCF binding sites KO
  - RNA Flow-FISH
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantification and statistical analysis of FACS data
  - Modelling of expected combined activities
  - RT-qPCR of endogenous genes in the synthetic locus
  - ATAC-seq analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.11.008.

Received: January 2, 2024
Revised: October 1, 2024
Accepted: November 7, 2024
Published: December 2, 2024

REFERENCES

1. Kim, S., and Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. Mol. Cell *83*, 373–392. https://doi.org/10.1016/j.molcel.2022.12.032.
2. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell *167*, 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018.
3. Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., et al. (2022). Nonlinear control of transcription through enhancer–promoter interactions. Nature *604*, 571–577. https://doi.org/10.1038/s41586-022-04570-y.
4. Rinzema, N.J., Sofiadis, K., Tjalsma, S.J.D., Verstegen, M.J.A.M., Oz, Y., Valdes-Quezada, C., Felder, A.-K., Filipovska, T., van der Elst, S., de Andrade dos Ramos, Z., et al. (2022). Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. Nat. Struct. Mol. Biol. *29*, 563–574. https://doi.org/10.1038/s41594-022-00787-7.
5. Jensen, C.L., Chen, L.-F., Swigut, T., Crocker, O.J., Yao, D., Bassik, M.C., Ferrell, J.E., Boettiger, A.N., and Wysocka, J. (2024). Long range regulation of transcription scales with genomic distance in a gene specific manner. Preprint at bioRxiv. https://doi.org/10.1101/2024.07.19.604327.
6. Catarino, R.R., and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. Genes Dev. *32*, 202–223. https://doi.org/10.1101/gad.310367.117.
7. Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN beta gene expression requires the assembly of an enhanceosome. Cell *83*, 1091–1100. https://doi.org/10.1016/0092-8674(95)90136-1.
8. Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328. https://doi.org/10.1126/science.aac6948.
9. Goto, T., Macdonald, P., and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. Cell *57*, 413–422. https://doi.org/10.1016/0092-8674(89)90916-1.
10. Pachano, T., Haro, E., and Rada-Iglesias, A. (2022). Enhancer-gene specificity in development and disease. Development *149*, dev186536. https://doi.org/10.1242/dev.186536.
11. Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. eLife *4*, e07956. https://doi.org/10.7554/eLife.07956.
12. Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L.L.P., Kassouf, M.T., Marieke Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903. https://doi.org/10.1038/ng.3605.
13. Carleton, J.B., Berrett, K.C., and Gertz, J. (2017). Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor α-bound enhancers. Cell Syst. *5*, 333–344.e5. https://doi.org/10.1016/j.cels.2017.08.011.
14. Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.-C., et al. (2016). Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. Dev. Cell *36*, 9–23. https://doi.org/10.1016/j.devcel.2015.12.014.
15. Shin, H.Y., Willi, M., HyunYoo, K.H., Zeng, X., Wang, C., Metser, G., and Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat. Genet. *48*, 904–911. https://doi.org/10.1038/ng.3606.
16. Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., and Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol. Cell *58*, 362–370. https://doi.org/10.1016/j.molcel.2015.02.014.
17. Moorthy, S.D., Davidson, S., Shchuka, V.M., Singh, G., Malek-Gilani, N., Langroudi, L., Martchenko, A., So, V., Macpherson, N.N., and Mitchell, J.A. (2017). Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res. *27*, 246–258. https://doi.org/10.1101/gr.210930.116.
18. Thomas, H.F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., et al. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. Mol. Cell *81*, 969–982.e13. https://doi.org/10.1016/j.molcel.2020.12.047.
19. Brosh, R., Coelho, C., Ribeiro-dos-Santos, A.M., Ellis, G., Hogan, M.S., Ashe, H.J., Somogyi, N., Ordoñez, R., Luther, R.D., Huang, E., et al. (2023). Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. Mol. Cell *83*, 1140–1152.e7. https://doi.org/10.1016/j.molcel.2023.02.027.
20. Blayney, J.W., Francis, H., Rampasekova, A., Camellato, B., Mitchell, L., Stolper, R., Cornell, L., Babbs, C., Boeke, J.D., Higgs, D.R., et al. (2023). Super-enhancers include classical enhancers and facilitators to fully activate gene expression. Cell *186*, 5826–5839.e18. https://doi.org/10.1016/j.cell.2023.11.030.

21. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310. https://doi.org/10.1038/s41576-019-0209-0.

22. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077. https://doi.org/10.1126/science.1232542.

23. Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat. Genet. *45*, 1021–1028. https://doi.org/10.1038/ng.2713.

24. Muerdter, F., Boryń, Ł.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., et al. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. Nat. Methods *15*, 141–149. https://doi.org/10.1038/nmeth.4534.

25. Gschwind, A.R., Mualim, K.S., Karbalayghareh, A., Sheth, M.U., Dey, K.K., Jagoda, E., Nurtdinov, R.N., Xi, W., Tan, A.S., Jones, H., et al. (2023). An encyclopedia of enhancer-gene regulatory interactions in the human genome. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.09.563812.

26. Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., and Levine, M.S. (2022). Genome organization controls transcriptional dynamics during development. Science *375*, 566–570. https://doi.org/10.1126/science.abi7178.

27. Chen, L.-F., Long, H.K., Park, M., Swigut, T., Boettiger, A.N., and Wysocka, J. (2023). Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus. Mol. Cell *83*, 1446–1461.e6. https://doi.org/10.1016/j.molcel.2023.03.009.

28. Loubiere, V., de Almeida, B.P., Pagani, M., and Stark, A. (2023). Developmental and housekeeping transcriptional programs display distinct modes of enhancer-enhancer cooperativity in Drosophila. Nat. Commun. *15*, 8584. https://doi.org/10.1038/s41467-024-52921-2.

29. Martinez-Ara, M., Comoglio, F., and van Steensel, B. (2023). Large-scale analysis of the integration of enhancer-enhancer signals by promoters. Elife *12*, RP91994. https://doi.org/10.7554/eLife.91994.

30. Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097. https://doi.org/10.1038/ng.946.

31. Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M., et al. (2019). Multi-omic profiling reveals dynamics of the phased progression of pluripotency. Cell Syst. *8*, 427–445.e10. https://doi.org/10.1016/j.cels.2019.03.012.

32. Spencley, A.L., Bar, S., Swigut, T., Flynn, R.A., Lee, C.H., Chen, L.-F., Bassik, M.C., and Wysocka, J. (2023). Co-transcriptional genome surveillance by HUSH is coupled to termination machinery. Mol. Cell *83*, 1623–1639.e8. https://doi.org/10.1016/j.molcel.2023.04.014.

33. Romeike, M., Spach, S., Huber, M., Feng, S., Vainorius, G., Elling, U., Versteeg, G.A., and Buecker, C. (2022). Transient upregulation of IRF1 during exit from naive pluripotency confers viral protection. EMBO Rep. *23*, e55375. https://doi.org/10.15252/embr.202255375.

34. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853. https://doi.org/10.1016/j.stem.2014.04.003.

35. Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. Mol. Cell *76*, 395–411.e13. https://doi.org/10.1016/j.molcel.2019.07.039.

36. Rideout, W.M., Wakayama, T., Wutz, A., Eggan, K., Jackson-Grusby, L., Dausman, J., Yanagimachi, R., and Jaenisch, R. (2000). Generation of mice from wild-type and targeted ES cells by nuclear cloning. Nat. Genet. *24*, 109–110. https://doi.org/10.1038/72753.

37. Greenshpan, Y., Sharabi, O., Yegodayev, K.M., Novoplansky, O., Elkabets, M., Gazit, R., and Porgador, A. (2022). The contribution of the minimal promoter element to the activity of synthetic promoters mediating CAR expression in the tumor microenvironment. Int. J. Mol. Sci. *23*, 7431. https://doi.org/10.3390/ijms23137431.

38. Arcot, S.S., Flemington, E.K., and Deininger, P.L. (1989). The human thymidine kinase gene promoter: Deletion analysis and specific protein binding. J. Biol. Chem. *264*, 2343–2349. https://doi.org/10.1016/S0021-9258(18)94182-7.

39. Araki, K., Araki, M., and Yamamura, K. (1997). Targeted integration of DNA using mutant lox sites in embryonic stem cells. Nucleic Acids Res. *25*, 868–872. https://doi.org/10.1093/nar/25.4.868.

40. Branda, C.S., and Dymecki, S.M. (2004). Talking about a revolution: the impact of site-specific recombinases on genetic analyses in mice. Dev. Cell *6*, 7–28. https://doi.org/10.1016/S1534-5807(03)00399-X.

41. Gilpatrick, T., Lee, I., Graham, J.E., Raimondeau, E., Bowen, R., Heron, A., Downs, B., Sukumar, S., Sedlazeck, F.J., and Timp, W. (2020). Targeted nanopore sequencing with Cas9-guided adapter ligation. Nat. Biotechnol. *38*, 433–438. https://doi.org/10.1038/s41587-020-0407-5.

42. Porichis, F., Hart, M.G., Griesbeck, M., Everett, H.L., Hassan, M., Baxter, A.E., Lindqvist, M., Miller, S.M., Soghoian, D.Z., Kavanagh, D.G., et al. (2014). High throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. Nat. Commun. *5*, 5641. https://doi.org/10.1038/ncomms6641.

43. Peng, T., Zhai, Y., Atlasi, Y., ter Huurne, M., Marks, H., Stunnenberg, H.G., and Megchelenbrink, W. (2020). STARR-seq identifies active, chromatin-masked, and dormant enhancers in pluripotent mouse embryonic stem cells. Genome Biol. *21*, 243. https://doi.org/10.1186/s13059-020-02156-3.

44. Agrawal, P., Blinka, S., Pulakanti, K., Reimer, M.H., Stelloh, C., Meyer, A.E., and Rao, S. (2021). Genome editing demonstrates that the −5 kb Nanog enhancer regulates Nanog expression by modulating RNAPII initiation and/or recruitment. J. Biol. Chem. *296*, 100189. https://doi.org/10.1074/jbc.RA120.015152.

45. Taylor, T., Sikorska, N., Shchuka, V.M., Chahar, S., Ji, C., Macpherson, N.N., Moorthy, S.D., de Kort, M.A.C., Mullany, S., Khader, N., et al. (2022). Transcriptional regulation and chromatin architecture maintenance are decoupled functions at the Sox2 locus. Genes Dev. *36*, 699–717. https://doi.org/10.1101/gad.349489.122.

46. Du, M., Stitzinger, S.H., Spille, J.-H., Cho, W.-K., Lee, C., Hijaz, M., Quintana, A., and Cissé, I.I. (2024). Direct observation of a condensate effect on super-enhancer controlled gene bursting. Cell *187*, 331–344.e17. https://doi.org/10.1016/j.cell.2023.12.005.

47. de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. Mol. Cell *60*, 676–684. https://doi.org/10.1016/j.molcel.2015.09.023.

48. Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Mol. Cell *78*, 539–553.e8. https://doi.org/10.1016/j.molcel.2020.03.002.

49. 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. https://doi.org/10.1038/nature11082.

50. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

51. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer–promoter topology and gene activity. Nat. Genet. *50*, 1296–1303. https://doi.org/10.1038/s41588-018-0175-z.

52. Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. Mol. Cell *76*, 473–484.e7. https://doi.org/10.1016/j.molcel.2019.07.038.

53. Alexander, J.M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., and Weiner, O.D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. eLife *8*, e41769. https://doi.org/10.7554/eLife.41769.

54. Li, J., Dong, A., Saydaminova, K., Chang, H., Wang, G., Ochiai, H., Yamamoto, T., and Pertsinidis, A. (2019). Single-molecule nanoscopy elucidates RNA polymerase II transcription at single genes in live cells. Cell *178*, 491–506.e28. https://doi.org/10.1016/j.cell.2019.05.029.

55. Li, J., Hsu, A., Hua, Y., Wang, G., Cheng, L., Ochiai, H., Yamamoto, T., and Pertsinidis, A. (2020). Single-gene imaging links genome topology, promoter-enhancer communication and transcription control. Nat. Struct. Mol. Biol. *27*, 1032–1040. https://doi.org/10.1038/s41594-020-0493-6.

56. Heist, T., Fukaya, T., and Levine, M. (2019). Large distances separate coregulated genes in living Drosophila embryos. Proc. Natl. Acad. Sci. USA *116*, 15062–15067. https://doi.org/10.1073/pnas.1908962116.

57. Xiao, J.Y., Hafner, A., and Boettiger, A.N. (2021). How subtle changes in 3D structure can create large changes in transcription. eLife *10*, e64320. https://doi.org/10.7554/eLife.64320.

58. Veitia, R.A. (2003). A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. Biol. Rev. Camb. Philos. Soc. *78*, 149–170. https://doi.org/10.1017/s1464793102006036.

59. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., et al. (2021). Sequence determinants of human gene regulatory elements. Nature Genetics *54*, 283–294. https://doi.org/10.1101/2021.03.18.435942.

60. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proc. Natl. Acad. Sci. USA *90*, 8424–8428. https://doi.org/10.1073/pnas.90.18.8424.

61. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. Nat. Biotechnol. *38*, 276–278.

62. Robinson, J.T., Thorvaldsdottir, H., Turner, D., and Mesirov, J.P. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). Bioinformatics *39*, btac830. https://doi.org/10.1093/bioinformatics/btac830.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Stbl3 bacteria	propagated from the initial Invitrogen stock	N/A
Chemicals, peptides, and recombinant proteins		
NcoI-HF	NEB	Cat#R3193S
XhoI	NEB	Cat#R0146S
SphI-HF	NEB	Cat#R3182S
BamHI-HF	NEB	Cat#R3136S
NotI-HF	NEB	Cat#R3189S
BsrGI-HF	NEB	Cat#R3575S
Puromycin	InvivoGen	Cat#ant-pr-1
Ganciclovir	InvivoGen	Cat#sud-gcv
Neomycin	Sigma-Aldrich	Cat#G8168
Hygromycin B	Sigma-Aldrich	Cat#10843555001
Blasticidin	InvivoGen	Cat#ant-bl-1
Gibson Assembly mix	homemade	N/A
HyClone DMEM/F12 medium without HEPES	Cytiva (Formerly GE Healthcare Life Sciences)	Cat#SH30271.01
AlbuMAXTM II Lipid-Rich Bovine Serum Albumin	GIBCO	Cat#11021-029
serum-free B- 27 Supplement (50 x)	GIBCO	Cat#17504044
N2 supplement	homemade	N/A
MEM NEAA (100 x)	GIBCO	Cat#11140035
Penicillin-Streptomycin (5,000 U/mL)	GIBCO	Cat#15070063
Sodium Pyruvate (100 mM)	GIBCO	Cat#11360039
2-Mercaptoethanol	GIBCO	Cat#21985023
CHIR-99021	Selleckchem	Cat#S1263
PD0325901	Selleckchem	Cat#S1036
hLIF	VBCF Protein Technologies Facility	N/A
Poly-L-ornithine hydrobromide	Sigma-Aldrich	Cat#P4638
Laminin from Engelbreth-Holm-Swarm	Sigma-Aldrich	Cat#L2020
Trypsin-EDTA solution	Sigma-Aldrich	Cat#T3924
Fetal Bovine Serum	Sigma-Aldrich	Cat#F7524
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat#11668019
Critical commercial assays		
Invitrogen PrimeFlow RNA Assay Kit	Thermo Fisher Scientific	Cat#88-18005-204
Dual-Glo Luciferase Assay	Promega	Cat#E2920
Deposited data		
FACS data	This Paper	Mendeley Data: https://doi.org/10.17632/2j75sff6jw.2
ATAC Seq Data	This Paper	ENA, accession number PRJEB80727
Experimental models: Cell lines		
Mouse: v6.5 ES cells	Rideout et al.[36]	N/A
Mouse: v6.5 3xlox clone 1	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: v6.5 3xlox clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3* E2 (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3* E2 (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + eNanog (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + SCR-rev (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + SCR-rev (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Fgf5*-PE (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Fgf5*-PE (1.5 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E1* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E2* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + Rybp *E2* (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + eNanog (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + SCR (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + SCR (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + SCR-rev (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + SCR-rev (25 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E1* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E2* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + Rybp *E2* (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + eNanog (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + SCR (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + SCR (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + SCR-rev (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 2 + SCR-rev (75 kb)	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E1* (1.5 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E1* (25 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E15* (1.5 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E15* (25 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E19* (1.5 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E19* (25 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E20* (1.5 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E20* (25 kb) - clone 1 - clone 3	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (25 kb) + *Map4k3 E2* (75 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (25 kb) + eNanog (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3* E2 (25 kb) + *Map4k3 E1* (75 kb) clone 1 - clone 2	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (25 kb) + Rybp *E2* (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (25 kb) + *Map4k3 E2* (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (1.5 kb) + *Map4k3 E2* (25 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (1.5 kb) + eNanog (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (25 kb) + Rybp *E2* (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (25 kb) + *Map4k3 E1* (75 kb) clone 1 - clone 5	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (25 kb) + *Map4k3 E1* (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E1* (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E2* (1.5 kb) + eNanog (25 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + eNanog (1.5 kb) + Rybp E2 (25 kb) clone 1 - clone 4	This paper	N/A
Mouse: v6.5 3xlox clone 1 + Rybp *E1* (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E15* (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E19* (25 kb) + eNanog (75 kb) clone 1 - clone 6	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Sox2 E20* (25 kb) + eNanog (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (1.5 kb) + *Map4k3 E1* (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 1 + *Map4k3 E2* (1.5 kb) + Rybp *E2* (75 kb) clone 1 - clone 2	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E2* (25 kb) + eNanog (75 kb) clone 1 - clone6	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3* E2 (25 kb) + Rybp E2 (75 kb) clone 1 - clone6	This paper	N/A
Mouse: v6.5 3xlox clone 2 + *Map4k3 E2* (25 kb) + *Map4k3 E1* (75 kb) clone 1 - clone6	This paper	N/A
Mouse: R1 mouse ESC	Nagy et al.[60]	N/A
Mouse: R1 Δ*Rybp E2* clone 1	This paper	N/A
Mouse: R1 Δ*Rybp* E2 clone 2	This paper	N/A
Mouse: R1 Δ*Map4k3 E1* clone 1	This paper	N/A
Mouse: R1 Δ*Map4k3 E1* clone 2	This paper	N/A
Mouse: R1 Δ*Map4k3 E2* clone 1	This paper	N/A
Mouse: R1 Δ*Map4k3 E2* clone 2	This paper	N/A
Recombinant DNA		
px330-U6-Chimeric_BB-CBh-hSpCas9	Addgene	Cat#42230
Fgf5-HAL-PE-pdTK-HAR	Thomas et al., [18]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BamHI - β-globin HAR - NotI	BioCat GmbH	N/A
targeting vector (pGemT-HAL-pdTK-spacer-TK-mCherry-pA-HAR)	This Paper	N/A
pGemT (circularised)	Promega	Cat#A362A
pGemT-FRT-lox66-FRT-blasticidin-deltaTK(7 additional versions of this plasmid with integration of Map4k3 E1, Map4k3 E2, Rybp E2, eNanog, SCR, SCR-rev and Fgf5-PE were also generated)	This Paper	N/A
pGemT-FRT3-loxm2/66-FRT3-puromycin-deltaTK (6 additional versions of this plasmid with integration of Map4k3 E1, Map4k3 E2, Rybp E2, eNanog, SCR and SCR-rev were also generated)	This Paper	N/A
pGemT-FRT5-lox2272/66-FRT5-hygromycin-deltaTK (6 additional versions of this plasmid with integration of Map4k3 E1, Map4k3 E2, Rybp E2, eNanog, SCR and SCR-rev were also generated)	This Paper	N/A
FlpO-expressing plasmid	Provided by the lab of Stefan Ameres	N/A
pGL3-SV40 promoter-luciferase-Map4k3 E1	This Paper	N/A
pGL3-SV40 promoter-luciferase-Map4k3 E2	This Paper	N/A
pGL3-SV40 promoter-luciferase-Rybp E2	This Paper	N/A
pGL3-SV40 promoter-luciferase-eNanog	This Paper	N/A
pGL3-SV40 promoter-luciferase-SCR	This Paper	N/A
pGL3-SV40 promoter-luciferase-SCR-rev	This Paper	N/A
Software and algorithms		
FlowJo (version 10.5.3)	BD Life Sciences - Biosciences	N/A
nf-core atacseq pipeline (version 2.1.2)	Ewels et al.[61]	N/A
Reform tool	https://github.com/gencorefacility/reform	N/A
Genrich (v0.6.1)	https://github.com/jsh58/Genrich	N/A
Other		
BD FACSAria III Cell Sorter	BD Life Sciences - Biosciences	N/A
Mus Musculus Castaneus DNA	Provided by the lab of Kikue Tachibana	N/A
HEK-293 DNA	Provided by the lab of Dea Slade	N/A

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

General culture conditions of mouse embryonic stem cells

For all work in this study we used the male v6.5 mouse embryonic stem cell line cultured as described in Thomas et al.[18]. For maintenance, cells were grown on 6/12-wells coated with first poly-L-ornithine hydrobromide (6 μg/ml in PBS, 1h at 37 °C) and then laminin (1.2 μg/ml in PBS, 1 hour at 37°C).

Cells were cultured in base medium DMEM/F12 without Hepes with 4 mg/mL Lipid-Rich Bovine Serum Albumin, 1x MACS NeuroBrew-21 with Vitamin A, 1x NEAA, 50 U/mL Penicillin-Streptomycin, 1 mM Sodium Pyruvate and 1x 2-Mercaptoethanol. The base medium was supplemented with 3.3 mM CHIR-99021, 0.8 mM PD0325901 and 10 ng/mL hLIF (provided by the VBCF Protein Technologies Facility, https://www.viennabiocenter.org/facilities/).

For splitting, cells were treated with 1x Trypsin-EDTA solution at 37 °C until cells detached. Trypsinization was stopped with 2i+LIF medium with 10% FSC, cells were harvested by centrifugation at 300g for 3 min, resuspended in 2i+LIF and seeded in appropriate ratios.

Generation of reporter cell line

For KI of the reporter gene into the β-globin locus, we first generated a gRNA-expressing plasmid and a targeting vector. Therefore, a single gRNA was designed to target the locus roughly 1.5 kb downstream of the *Hbb-y* gene. As described before, the gRNA was cloned into Addgene plasmid #42230.[18] Due to a SNP in the underlying sequence, this gRNA recognises the C57BL/6 but not the 129/sv allele of v6.5 mouse ESCs, allowing us to introduce the reporter gene specifically into the β-globin-locus on the C57BL/6 allele.

We used Gibson assembly to generate the targeting vector to insert the reporter gene into the *β-globin*-locus unless noted otherwise. Homology arms were designed to disrupt the gRNA recognition sequence upon successful KI. All guide RNAs used in this study are listed in Table S1, primer sequences used for cloning in this study are listed in the Table S2, all single strand DNA oligos are listed in Table S4.

We first amplified a 77 bp long minimal TK promoter followed by the mCherry coding sequence by PCR and inserted the fragment into an NcoI-HF-digested (NEB) pGemT-vector (Promega). We amplified the left homology arm (637 bp) targeting the β-globin locus roughly 1.5 kb downstream of the *Hbb-y* gene from R1 ESC DNA by PCR. We inserted it upstream of the TK-promoter into the SphI-HF-digested (NEB) TK-mCherry plasmid, leaving the SphI-site intact. We used this SphI site to introduce a loxP-flanked puromycin-deltaTK cassette and a 1.5 kb spacer from inert human DNA between the left homology arm and the TK promoter. The loxP-flanked puromycin-deltaTK cassette was PCR-amplified from a previously generated targeting vector.[18] Using a forward primer that mapped to the loxP site upstream of the puromycin-deltaTK cassette but contained mutations, we turned the upstream loxP site into a lox71 site. The 1.5 kb spacer was amplified from HEK-293 genomic DNA, kindly provided by the lab of Dea Slade. We then inserted a 232 bp long BGH polyA sequence downstream of the mCherry coding sequence into the NotI-HF-digested (NEB) plasmid, leaving the NotI-site intact and introducing an additional BamHI-site. As we did not manage to PCR-amplify the right homology arm, we ordered the synthesised fragment (832 bp) integrated into a plasmid and flanked by a BamHI- (upstream) and a NotI-site (downstream) from BioCat GmbH. We obtained the right homology arm fragment by BamHI-HF- and NotI-HF-digestion (both NEB) and ligated it into the likewise BamHI-HF- and NotI-HF-digested HAL-pdTK-spacer-TK-mCherry-pA plasmid, downstream of the polyA site.

For KI of the reporter gene, v6.5 mouse ESCs[38] were cultured and transfected by lipofection with 400 ng of circular targeting vector, 400 ng of gRNA-containing plasmid and 4 μL of Lipofectamine 2000 Transfection Reagent (Invitrogen), as described before.[18] In brief, cells were transferred to a 10 cm dish the day after transfection and selection with Puromycin (2 μg/mL, InvivoGen) was started within 48 h of transfection. After a week of selection, single colonies were picked. Integration was validated by PCR with primers "KI validation 1 forward+reverse" (mapping upstream of the left homology arm in the genome and downstream of the left homology arm in the insert) and "KI validation 2 forward+reverse" (mapping upstream of the right homology arm in the insert and downstream of the right homology arm in the genome). All analysed clones were heterozygous, with the reporter gene being inserted on the C57BL/6 allele as judged by SNPs in Sanger-sequenced PCR products (Microsynth AG). Clones were expanded and subsequently transfected with a plasmid expressing Cre-recombinase to remove the puromycin-deltaTK cassette and leave a single lox71 site (in reverse orientation) behind. After one week of selection with Ganciclovir (500 ng/mL, InvivoGen), single clones were picked, and removal of the puromycin-deltaTK cassette was confirmed with primers "KI validation after Cre forward+reverse" (mapping up- and downstream of the excised cassette). Overlapping PCR products (primers "reporter validation 1-4 forward+reverse") spanning the entire insert, the homology arms and the genome up- and downstream of the insertion were submitted for Sanger sequencing to confirm the intactness of the β-globin-locus, the homology arms and the inserted sequence. We selected two clones with reporter gene integration for subsequent integration of the remaining LPs.

To introduce lox2272/71 and loxm2/71 sites roughly 25 and 75 kb upstream of the mCherry reporter gene, we designed gRNAs and cloned them into Addgene plasmid #42230. We also ordered single-stranded (ss) DNA oligos from Microsynth AG, containing the respective lox site (34 bp; in reverse orientation) and left and right homology arm. The homology arm at the 5' end of the respective oligo was 66 bp long, and the homology arm at the 3' end was 50 bp (150 bp total, including the lox site). Homology arms were designed to disrupt the gRNA recognition sequence upon successful KI.

We used two independent clones with reporter gene integration to sequentially knock-in the two lox sites. Therefore, we transfected 500 ng of gRNA-expressing plasmid, 1000 ng of ssDNA-oligo and 100 ng of plasmid expressing a fluorescent marker with 10 μl of Lipofectamine 2000 Transfection Reagent (Invitrogen). Transfected cells were processed as described before,[18] i.e., single fluorescent cells were sorted into 96-well plates. Successful insertion of the lox sites into the genome was initially confirmed by PCR with forward primers overlapping the integrated lox sites (and thus only binding to the modified alleles) and reverse primers downstream of the integration site ("lox KI 25/75 kb initial forward+reverse"). Selected clones were tested further with a more upstream forward primer binding to both alleles ("lox KI 25/75 kb validation forward"), and resulting PCR products were subcloned into a pGEM-T vector (Promega, pGEM-T Vector Systems) for subsequent Sanger sequencing of both alleles. The recognition sequence of the gRNA used for the lox site at 25 kb contains SNPs on the 129/sv allele, allowing us to integrate the lox2272/71 site on the C57BL/6 allele specifically. For integration at 75 kb, we did not design a gRNA overlapping a SNP. Instead, we chose heterozygous clones with the integration of the loxm2/71 site on the C57BL/6 allele, as judged by SNPs surrounding the integration site.

METHOD DETAILS

Design of plasmids for enhancer integration

We generated three plasmids to integrate enhancers at 1.5, 25 and 75 kb distance upstream of the reporter gene by targeting the lox71, loxm2/71 and lox2272/71 sites. We first introduced a single FRT site into a pGemT-vector (Promega). Therefore, forward and reverse DNA oligonucleotides - containing the FRT sequence as well as the overhangs required for cloning - were ordered from Microsynth AG, annealed and cloned into a NcoI-HF- and SphI-HF- digested pGemT plasmid, as described before.[18] We generated three plasmids containing FRT, FRT3 and FRT5 sites, respectively. The original restriction sites were disrupted during this process, but an additional NcoI-site was introduced downstream of the FRT sites as part of the inserted oligo. We then ordered additional DNA oligos and introduced lox66 sites downstream of the FRT sites of the resulting NcoI-HF- and NotI-HF-digested plasmids (both NEB). For the plasmid containing an FRT site, we introduced a lox66 site; for the plasmid containing the FRT3 site, we introduced a loxm2/66 site; and for the plasmid with the FRT5 site, a lox2272/66 site.

We digested the resulting plasmids by NotI-HF (NEB) and inserted cassettes expressing fusions of antibiotic resistance genes with deltaTK downstream of the lox66 sites by Gibson assembly. We inserted blasticidin-deltaTK into the FRT-lox66-plasmid, puromycin-deltaTK into the FRT3-loxm2/66-plasmid and hygromycin-deltaTK into the FRT5-lox2272/66-plasmid. The forward primer used for amplification of the cassettes included an additional FRT site in the case of blasticidin-deltaTK, an additional FRT3 site in the case of hygromycin-deltaTK and an additional FRT5 site for puromycin-deltaTK. The NotI site between the lox and the newly inserted FRT site was restored and used to integrate enhancer sequences.

All in all, we generated the following three plasmids:

- pGemT-FRT-lox66-FRT-blasticidin-deltaTK
- pGemT-FRT3-loxm2/66-FRT3-puromycin-deltaTK and
- pGemT-FRT5-lox2272/66-FRT5-hygromycin-deltaTK.

These plasmids were digested by NotI-HF, and enhancer sequences amplified from *castaneus* mouse strain DNA (kindly provided by the lab of Kikuë Tachibana) were inserted by Gibson assembly. Primers for amplifying enhancer sequences were designed to encompass the central p300 peak (ChIP-seq from Buecker et al.[34]). The resulting plasmids were transfected together with a Cre-recombinase expressing plasmid into the reporter cell line for insertion of enhancers at the lox71 sites of the reporter locus.

Integration of enhancer sequences

To integrate enhancer sequences, 200,000 cells of the reporter cell line were plated and transfected by lipofection on the following day. Therefore, 250 ng of enhancer-containing plasmid, 1750 ng of Cre-recombinase expressing plasmid and 10 μL of Lipofectamine 2000 Transfection Reagent (Invitrogen) were used. Cells were selected with Blasticidin (10 μg/mL, InvivoGen), Hygromycin B (400 μg/mL, Sigma-Aldrich) or Puromycin (2 μg/mL, InvivoGen), depending on the selection cassette present in the enhancer plasmid. Single colonies were picked after a week of selection. Plasmid integration was validated by PCR with forward primers mapping to the integrated plasmid backbone (“integration forward”) and reverse primers mapping downstream of the respective lox site (“integration 1.5/75 reverse; for 25 kb: “lox KI 25 kb initial reverse” that was used before for validating KI of the lox site). Colonies with integration of enhancer plasmid were expanded and transfected with plasmid expressing FlpO-recombinase (a more active, codon-optimized version of Flp kindly provided by the lab of Stefan Ameres) to remove the plasmid backbone including the selection cassette (200,000 cells, 2 μg of FlpO-recombinase expressing plasmid, 10 μL of Lipofectamine 2000 Transfection Reagent (Invitrogen)). Cells were passaged, seeded at low density the day after transfection and selected with Ganciclovir (5 μg/mL, InvivoGen) for one week. Single colonies were picked. PCR confirmed the removal of the selection cassette with an enhancer-specific forward primer combined with a primer downstream of the respective lox71 site (“integration 1.5/75 reverse”; for 25 kb: “lox KI 25 kb initial reverse”). In addition, the PCR with primers “integration forward” and “integration 1.5/75 reverse” or “lox KI 25 kb initial reverse” was repeated. As the forward primer maps to the plasmid backbone, the absence of a band in that PCR confirms the complete removal of the plasmid backbone in the entire cell population.

Allele-specific primers up and downstream of the lox sites were designed. In the case of the lox71 site at 1.5 kb, a forward primer upstream of the lox71 site (“1.5 forward”, recognising both alleles) was combined with a reverse primer mapping to the spacer sequence that is only integrated on the C57BL/6 allele as part of the reporter gene and has been used before for confirming plasmid and enhancer integration (“integration 1.5 reverse”). For the other two lox sites, forward primers mapping just upstream of and partially overlapping the lox sites (“25/75 forward”) were combined with reverse primers that had been used before for confirming plasmid and enhancer integration and recognise both alleles (“integration 75 reverse”/“lox KI 25 kb initial reverse”). These primer combinations gave rise to PCR products spanning the insert and the surrounding genome. Thus, PCR products having the expected size confirmed the intactness of the inserted sequence and the surrounding genome. Selected PCR products were submitted for Sanger sequencing to further confirm the absence of mutations in the inserted enhancer sequences. The size of the SCR and SCR-rev prevented the amplification of a single PCR product spanning the entire insert. Instead, we performed two PCRs giving rise to partially overlapping PCR products that together span the entire insert and the surrounding genome (“1.5/25/75 forward+SCR-reverse” and “SCR-forward+integration 1.5/25/75 reverse” for SCR, “1.5/25/75 forward+SCR-forward” and

"SCR-reverse+integration 1.5/25/75 reverse" for SCR-rev). We introduced an initial set of enhancers (*Sox2 SCR*, *eNanog*, *Rybp E2*, *Map4k3 E1*, *Map4k3 E2*) into the two clones we derived from the initial knock-in of the reporter gene (see above). For each of the two initial reporter cell lines, we generated a single clone with the enhancer integration, treated with FlpO and selected a single clone after FlpO-treatment (resulting in a total of 2 clones for each enhancer integration after FlpO treatment, with each of them being derived from a different parental reporter cell line clone). Since the results in the two clones of the reporter cell line were highly reproducible, we integrated the remaining individual enhancers (*Rybp E1*, *Sox2 E15*, *Sox2 E19* and *Sox2 E20*) only into clone 1 of the initial reporter cell line. After Cre-mediated insertion of the enhancer-containing plasmid, two clones were independently expanded and transfected with FlpO. After selection, two clones per enhancer integration were expanded and verified, resulting in a total of four clones per single enhancer.

To integrate the dual enhancers, we selected 1-2 individual clonal cell lines before removing the plasmid backbone, i.e., before FlpO treatment. We integrated the second enhancer by transfecting the cells with Cre-recombinase and the second enhancer plasmid. We included both selections to ensure that both enhancers are integrated and selected single clones. We identified positive clones via PCR with the primers described above and removed both backbones by FlpO treatment afterwards. For all cell lines, extensive PCR validation ensured both enhancers were integrated. Multiple independently derived clones were selected for further analysis.

Cas9-targeted sequencing

To verify correct integrations and that the whole locus is not rearranged during the construction of the synthetic locus, we performed targeted nanopore sequencing with Cas9-guided adapter ligation for selected cell lines, both before and after enhancer integration. We designed two small pools of gRNA libraries targeting the modified reporter gene locus. The distance between each gRNA is 5 kb. Cas9 protein, tracRNA and crRNA were ordered from IDT. We cooperated with the Vienna BioCenter NGS core facility and followed the protocol for Cas9-targeted sequencing using the Cas9 sequencing Kit(SQK-CS9109) from Oxford Nanopore Technologies.

FACS analysis of mCherry expression

To assess reporter gene activation upon enhancer integration, mCherry levels were measured by FACS. Therefore, 200,000 cells were plated on a 6-well plate in 2i/LIF medium. For each experiment, the two parental reporter cell line clones without enhancer integration and an untransfected v6.5 control were included. After two days, cells were collected by adding trypsin-EDTA (Sigma-Aldrich). Trypsinisation was stopped after incubation at 37 °C for 7 minutes by adding 2i/LIF medium containing 10% serum. Cells were resuspended and centrifuged for 3 minutes at 300 g, the supernatant was removed, and the cell pellet was resuspended in 2i/LIF medium. mCherry-fluorescence was measured with a BD LSRFortessa Flow Cytometer (BD Life Sciences - Biosciences).

FACS sorting of mCherry-populations

Upon integration of the *Nanog* enhancer at 1.5 kb, we observed some clones with a fraction of mCherry-negative cells. We sorted mCherry-positive and -negative cells with a BD FACSAria III Cell Sorter (BD Life Sciences - Biosciences). Sorting gates were determined by measuring mCherry fluorescence of a clone without a fraction of mCherry-negative cells: All cells with mCherry signal lower than the lowest-expressing cells of that clone were sorted as negative, and all remaining cells were sorted as positive. Sorted cells were kept in culture, and mCherry levels were analysed with a BD LSRFortessa Flow Cytometer (BD Life Sciences - Biosciences) every few passages.

Luciferase assays

For luciferase assays, we used a pGL3 plasmid with the Firefly luciferase coding sequence under the control of the SV40 promoter followed by a poly-adenylation signal (Promega). The same enhancer sequences we integrated into the reporter locus were amplified by PCR from *castaneus* mouse strain DNA and inserted downstream of the poly-adenylation signal by Gibson assembly. The primer sequences used to generate these plasmids are indicated in Table S2. Luciferase assays were slightly adapted but otherwise performed as described previously.[18] 10,000 cells from the v6.5 cell line were plated on a 96-well plate and immediately transfected with 120 ng of enhancer-luciferase plasmid, 4 ng of Renilla control plasmid (Promega) and 0.62 μL of Lipofectamine 2000 Transfection Reagent (Invitrogen). The medium was removed 5-7 h after transfection, and fresh 2i/LIF medium was added. Firefly and Renilla luciferase activity were measured 48 h after transfection. Background-subtracted Firefly measurements were normalised to background-subtracted Renilla values. The resulting values were normalised to the control plasmid without enhancer integration. To detect statistically significant increases in luciferase activity compared to the no-enhancer control, we performed one-sided Welch one-sample t-tests on the resulting normalised values to assess whether they are significantly higher than 1 (the value of the no-enhancer control).

Enhancer KO and RT-qPCR

Enhancers at the *Map4k3* locus were deleted in male R1 WT ESCs[60] as described.[18] RNA was extracted from confluent 6-well plates, and resulting *Map4k3* expression levels were measured by RT-qPCR as described before.[18] All qPCR primers used in this study are listed in Table S3. Expression levels of each replicate were normalised first to Rpl13a and then to WT. To assess statistically significant decreases in expression upon enhancer deletion, we performed one-sided Welch one-sample t-tests on the resulting

WT-normalized values to assess whether they are significantly lower than 1 (as all values are normalised to WT, a value of 1 corresponds to WT expression levels).

CTCF binding sites KO

Two CTCF binding sites are found between LP 25 kb and LP 75 kb. To delete the CTCF binding site, we designed one gRNA targeting the motif of CTCF peak 1, and two gRNAs surrounding the motif of CTCF peak 2. We first deleted the CTCF peak 1 in four cell lines (*eNanog* at 75 kb, *ME1* at 75 kb, *eNanog* at 75 kb with *ME2* at 25 kb and *ME1* at 75 kb with *ME2* at 25 kb). We choose two clones of CTCF peak 1 deletion from each cell line for further KO of CTCF peak 2. Several double KO clones of CTCF binding sites from each cell line are verified by PCR and Sanger sequencing.

RNA Flow-FISH

To validate the correlation between the levels of reporter gene fluorescence and the number of mRNAs, mCherry mRNA levels were quantified performing RNA Flow-FISH.

For this purpose, PrimeFlow RNA Assay Kit was purchased from Thermo Fisher Scientific (88–18005-204) along with Alexa Fluor 647 mCherry target probes. The assay was performed four times following the recommended protocol. Shortly, the cells were first fixed and then permeabilized using the PrimeFlow RNA Fixation/Permeabilization Buffers provided in the kit. Next, cells were incubated with mCherry target probe for the hybridization step, followed by signal amplification via incubation with corresponding pre-amplifiers, amplifiers and fluorescent label probes. Cells were then analyzed on a BD LSRFortessa Flow Cytometer (BD Life Sciences - Biosciences). The subsequent gating and data analysis was done in FlowJo software.

As negative controls, we also performed the entire protocol with samples containing cells without our synthetic locus, therefore lacking mCherry reporter gene; and samples lacking the mCherry target probes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification and statistical analysis of FACS data

In each FACS experiment, we always included the no enhancer control cell line to account for small changes in laser settings and where indicated the v6.5 WT control. In addition, we only compared experimental replicates that were analyzed in the same experiments.

The resulting FCS files were analysed with the FlowJo software (BD Life Sciences – Biosciences, version 10.5.3). We visualised cell populations as histograms normalised to the maximum cell count for each cell line to analyse the fluorescence distribution in a cell population. We calculated the mean of mCherry fluorescence with the FlowJo software to compare different cell lines and replicates. To account for background fluorescence, we subtracted the mean fluorescence of an untransfected v6.5 control from all values and used the resulting values for the subsequent analysis described below.

We calculated the average of mean fluorescence and plotted the resulting values for each measurement in bar graphs (see Figures). To assess statistically significant increases in mCherry fluorescence upon enhancer integration, we performed one-sided paired t-tests on cell lines with enhancer integration compared to their respective parental reporter cell line clone without enhancers.

To compare the distance dependency of different enhancers, we normalised the mean fluorescence of each enhancer construct as well as the "no enhancer" control to the clone carrying the enhancer at 1.5 kb. We performed one-sided paired t-tests on the normalised values to identify statistically significant increases compared to the respective "no enhancer" control.

To compare combinations of enhancers, we calculated the mean mCherry expression of all cell lines, either with one or the combination of two enhancers, and subtracted the mean mCherry expression of the ‘‘no enhancer’’ control from the mean expression of each cell line. We calculated the expected additive expression by adding the resulting (‘‘no enhancer’’ control subtracted) values. Normalized values from independently derived enhancer constructs were pooled and the mean expression was calculated for each replicate. The expected additive expression was calculated by adding the mean over all replicates. A one-sided paired t-test was performed to assess whether the dual enhancer cell line exceeds the expected additive expression.

Modelling of expected combined activities

The individual activity of enhancers in the 75 kb, 25 kb and 1.5 kb pads were scaled using the negative control conditions where the three pads contain inactive control sequences (-/-/-). Therefore, they correspond to fold-changes normalized to the basal activity of the core promoter expressed in log2. Hence, for a given combination C, $\log_2$ additive and multiplicative predicted values were computed using the following formulas:

$$C_{additive} = log\,2(2^{i_{75}} + 2^{i_{25}} + 2^{i_{1.5}} - 2)$$

$$C_{multiplicative} = i_{75} + i_{25} + i_{1.5}$$

Where i_{75}, i_{25} *and* $i_{1.5}$ correspond to the $\log_2$ individual enhancer activities inserted 75 kb, 25 kb and 1.5 kb upstream of the core promoter, respectively. Finally, we fitted a multiplicative model with interaction term using $\log_2$ activity values and the *lm* function in R:

$$C_{activity} = \beta_0 + \beta_1 i_{75} + \beta_2 i_{25} + \beta_3 i_{1.5} + \beta_{12} i_{75} i_{25} + \beta_{13} i_{75} i_{1.5} + \beta_{23} i_{25} i_{1.5} + \beta_{123} i_{75} i_{25} i_{1.5} + \epsilon$$

Where β_0 is the intercept, and β_1, β_2 and β_3 coefficients represent the individual contributions of each enhancer's $\log_2$ activity to the overall activity contribution. β_{12}, β_{13} and β_{23} are the coefficients for the two-way interactions, and capture the combined effects when two enhancers are active simultaneously, beyond what would be expected from their individual effects alone. β_{123} is the coefficient of the three-way interaction, indicating how the simultaneous presence of enhancer activities at 75 kb, 25 kb, and 1.5 kb influences the core promoter activity. However, since no such combination was present in our dataset, this coefficient was automatically set to NA. Finally, ϵ is the error term.

The performance of each model was assessed using R-squared (R^2) computed with the following formula:

$$R^2 = \frac{SS_{regression}}{SS_{total}}$$

Where, $SS_{regression}$ and SS_{total} correspond to the Sum of Squares due to regression and the total Sum of Squares, respectively. For fitted models, R-squared values were adjusted (Adj. R^2) to account for the number of predictors.

RT-qPCR of endogenous genes in the synthetic locus

Cell lines with the SCR enhancer integrated at 1.5 kb, 25 kb, and 75 kb before and after FlpO treatment were selected to investigate the endogenous gene expression in the reporterlocus. We also included the v6.5 ES cell line and parental clone 1 cell line without any enhancers integrated as negative controls. RNA of each cell line was extracted from confluent 6-well plates, and was further reverse transcribed to cDNA. Two genes in the synthetic locus, *Olfr67* and *Hbb-γ*, were selected to analyse their expression level by RT-qPCR as described above. To assess statistically significant increases in expression upon enhancer integration, we performed one-sided Welch one-sample t-tests on the resulting v6.5-normalized values to assess whether they are significantly higher than 1 (as all values are normalised to v6.5, a value of 1 corresponds to v6.5 expression levels).

ATAC-seq analysis

For the ATAC-seq analysis performed by the Next Generation Sequencing facility at the Vienna BioCenter Core Facilities (VBCF) we seeded 100.000 cells and harvested 48h later, resuspending in Dulbecco's Phosphate Buffered Saline (PBS, Sigma-Aldrich).

We used cell lines with the SCR enhancer at 1.5 kb and 75 kb as well as *Map4k3 E2* at the same distances. V6.5 ES cell line and parental clone 2 without enhancer cell lines were used as controls. Over 100k cells with over 85% viability were required by the facility and the biological replicates were delivered on consecutive days. Samples were run on NovaSeq S4 PE150 XP.

We used the reform tool (https://github.com/gencorefacility/reform) to generate custom-made mm10 genome and annotation files for the modified cell lines with integration of the reporter gene, the landing pads and in some cases the different enhancers. Fastq files were analyzed with version 2.1.2 of the nf-core atacseq pipeline,[61] aligning to the custom genomes generated above. For the files aligned to the genome that only contains integration of the reporter gene and the landing pads, but not of the respective enhancers, we performed peak calling with Genrich (v0.6.1) and identified consensus peaks that overlap at least 50% between replicates with the samtools intersect tool. Peaks overlapping blacklisted regions (from http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were removed with substractBed, and peaks as well as aligned reads were displayed in IGV[62] using the 'Group Autoscale' function.

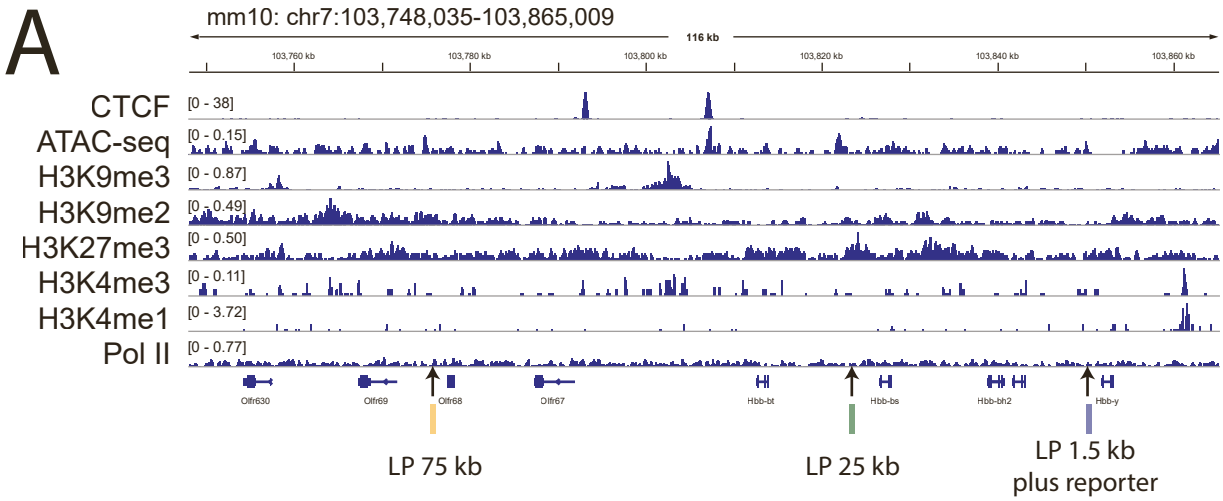
Molecular Cell, Volume *85*

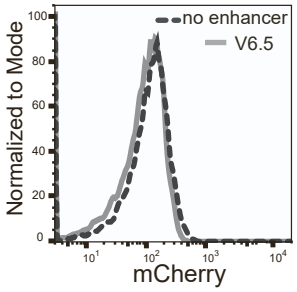
Supplemental information

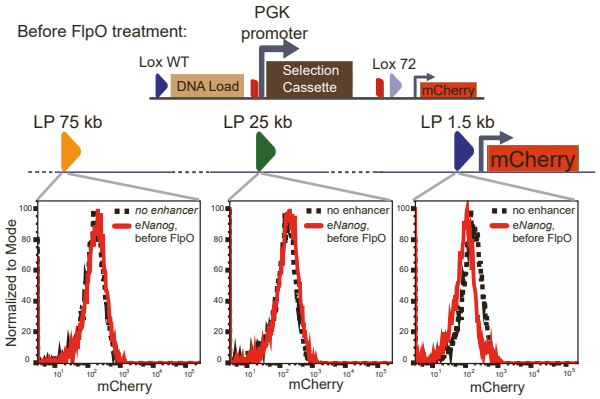
Enhancer cooperativity can compensate for loss of activity over large genomic distances

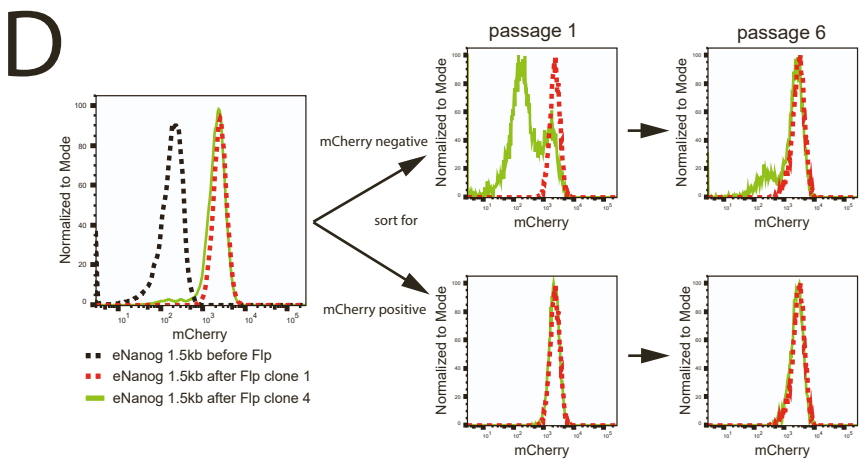
Henry F. Thomas, Songjie Feng, Felix Haslhofer, Marie Huber, María García Gallardo, Vincent Loubiere, Daria Vanina, Mattia Pitasi, Alexander Stark, and Christa Buecker

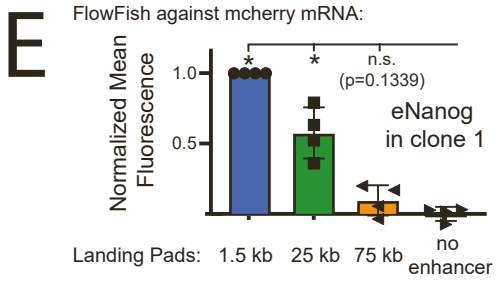
Figure S1

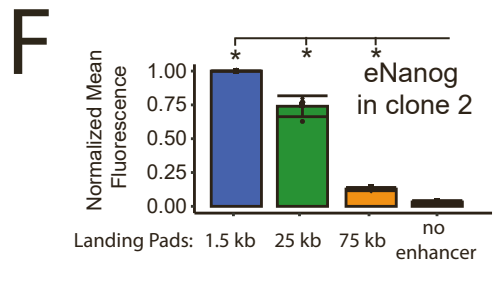

A
mm10: chr7:103,748,035-103,865,009
116 kb
103,760 kb
103,780 kb
103,800 kb
103,820 kb
103,840 kb
103,860 kb
CTCF
[0 - 38]
ATAC-seq
[0 - 0.15]
H3K9me3
[0 - 0.87]
H3K9me2
[0 - 0.49]
H3K27me3
[0 - 0.50]
H3K4me3
[0 - 0.11]
H3K4me1
[0 - 3.72]
Pol II
[0 - 0.77]
Olfr630
Olfr69
Olfr68
Olfr67
Hbb-bt
Hbb-bs
Hbb-bh2
Hbb-y
LP 75 kb
LP 25 kb
LP 1.5 kb
plus reporter


B
no enhancer
V6.5
Normalized to Mode
mCherry


C
Before FlpO treatment:
PGK
promoter
Lox WT
DNA Load
Selection
Cassette
Lox 72
mCherry
LP 75 kb
LP 25 kb
LP 1.5 kb
mCherry
no enhancer
eNanog,
before FlpO
Normalized to Mode
mCherry


D
passage 1
passage 6
mCherry negative
sort for
mCherry positive
Normalized to Mode
mCherry
eNanog 1.5kb before Flp
eNanog 1.5kb after Flp clone 1
eNanog 1.5kb after Flp clone 4


E
FlowFish against mcherry mRNA:
Normalized Mean
Fluorescence
1.0
0.5
*
*
n.s.
(p=0.1339)
eNanog
in clone 1
Landing Pads:
1.5 kb
25 kb
75 kb
no
enhancer


F
Normalized Mean
Fluorescence
1.00
0.75
0.50
0.25
0.00
*
*
*
eNanog
in clone 2
Landing Pads:
1.5 kb
25 kb
75 kb
no
enhancer

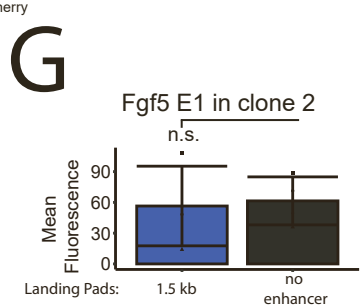

G
Fgf5 E1 in clone 2
n.s.
Mean
Fluorescence
90
60
30
0
Landing Pads:
1.5 kb
no
enhancer

Figure S1: Establishment and validation of the synthetic locus, related to Figure 1

A: IGV browser shot depicting ATAC-seq [1] and ChIP-seq against CTCF [2], H3K9me3 [3], H3K9me2 [4], H3K27me3 [3], H3K4me3 [4], H3K4me1 [5] and Polymerase II (Pol II) [4] across the b-globin locus and neighboring genes

B: FACS plot showing a selected example of clone 2 no enhancer and the parental cell line v6.5

C: Top: schematic of the synthetic locus organisation before removal of the backbone plasmid, Bottom: representative FACS plots of different *eNanog* integrated clones at the indicated LPs before removal of the plasmid backbone

D: Quantification of mean mCherry expression of *eNanog* integrated at the indicated LPs in clone 2. All mean expressions were normalised to the 1.5 kb integration. Statistically significant different signals ($p < 0.05$, one-sided paired t-tests) are marked by stars.

E: Quantification of mean mCherry mRNA levels of *eNanog* integrated at the indicated LPs in clone 1 using FlowFISH. All mean expressions were normalised to the 1.5 kb integration. Statistically significant different signals ($p < 0.05$, one-sided paired t-tests) are marked by stars.

F: FACS plots of mCherry expression in selected clones after backbone plasmid removal; mCherry negative (top) and positive (bottom) cells were sorted and analysed after 1 (left) or 6 (right) additional passages

G: Representative FACS plots showing an individual clonal cell line with the *Fgf5 E1* enhancer integrated at the 1.5 kb LP in clone 2.

Figure S2

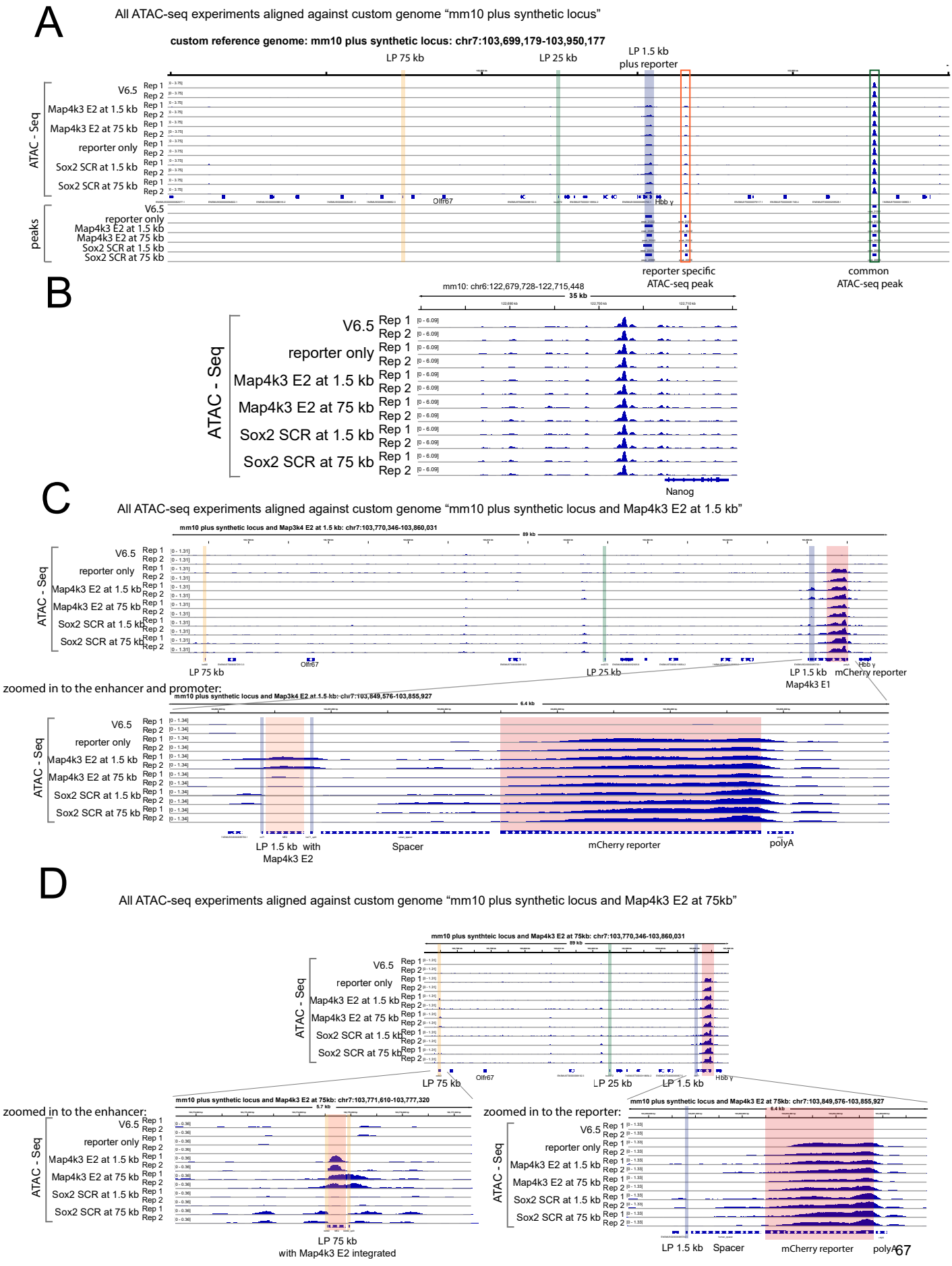

A
All ATAC-seq experiments aligned against custom genome “mm10 plus synthetic locus”
custom reference genome: mm10 plus synthetic locus: chr7:103,699,179-103,950,177
LP 75 kb
LP 25 kb
LP 1.5 kb plus reporter
ATAC - Seq
V6.5
Map4k3 E2 at 1.5 kb
Map4k3 E2 at 75 kb
reporter only
Sox2 SCR at 1.5 kb
Sox2 SCR at 75 kb
Rep 1
Rep 2
Olfr67
Hbb y
peaks
reporter specific ATAC-seq peak
common ATAC-seq peak
B
mm10: chr6:122,679,728-122,715,448
35 kb
Nanog
C
All ATAC-seq experiments aligned against custom genome “mm10 plus synthetic locus and Map4k3 E2 at 1.5 kb”
mm10 plus synthetic locus and Map3k4 E2 at 1.5 kb: chr7:103,770,346-103,860,031
89 kb
LP 1.5 kb Map4k3 E1
mCherry reporter
zoomed in to the enhancer and promoter:
mm10 plus synthetic locus and Map3k4 E2 at 1.5 kb: chr7:103,849,576-103,855,927
6.4 kb
LP 1.5 kb with Map4k3 E2
Spacer
mCherry reporter
polyA
D
All ATAC-seq experiments aligned against custom genome “mm10 plus synthetic locus and Map4k3 E2 at 75kb”
mm10 plus synthteic locus and Map4k3 E2 at 75kb: chr7:103,770,346-103,860,031
89 kb
LP 1.5 kb
zoomed in to the enhancer:
mm10 plus synthetic locus and Map4k3 E2 at 75kb: chr7:103,771,610-103,777,320
5.7 kb
LP 75 kb with Map4k3 E2 integrated
zoomed in to the reporter:
mm10 plus synthetic locus and Map4k3 E2 at 75kb: chr7:103,849,576-103,855,927
6.4 kb
LP 1.5 kb
Spacer
mCherry reporter
polyA

Figure S2: ATAC seq of cell lines with the synthetic locus, related to Figure 1

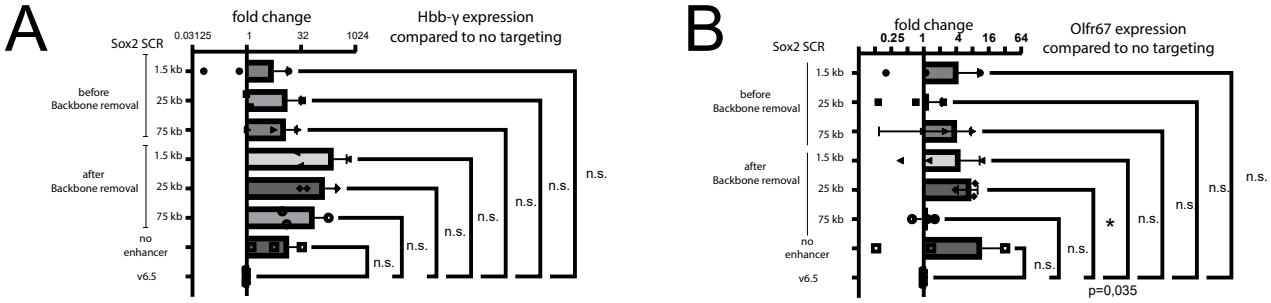
A: IGV browser track showing ATAC seq signal at the β-globin locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs. Sequencing reads were aligned against a custom genome containing the synthetic locus. Bottom: Significant ATAC peaks are indicated.

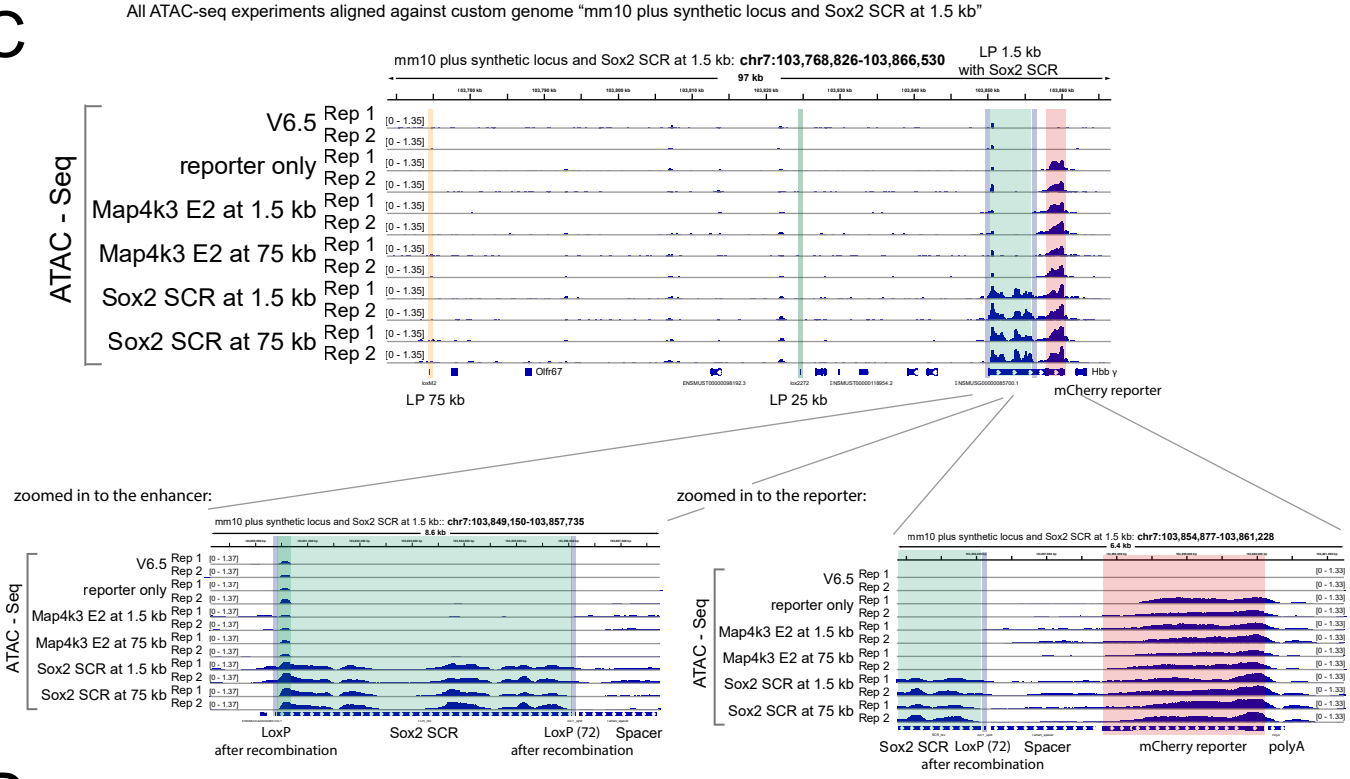
B: IGV browser track showing ATAC seq signal at the *Nanog* locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs.

C: IGV browser track showing ATAC seq signal at the β-globin locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs. Sequencing reads were aligned against a custom genome containing the synthetic locus and the *Map4k3 E2* enhancer at 1.5 kb. Bottom: zoom into the integrated enhancer and the mCherry reporter.

D: IGV browser track showing ATAC seq signal at the β-globin locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs. Sequencing reads were aligned against a custom genome containing the synthetic locus and the *Map4k3 E2* enhancer at 75 kb. Bottom: zoom into the integrated enhancer (left) and the mCherry reporter (right).

Figure S3


A
fold change
Hbb-γ expression compared to no targeting
Sox2 SCR
0.03125
1
32
1024
before Backbone removal
after Backbone removal
no enhancer
v6.5
1.5 kb
25 kb
75 kb
n.s.
B
fold change
Olfr67 expression compared to no targeting
0.25
1
4
16
64
*
p=0,035


C
All ATAC-seq experiments aligned against custom genome “mm10 plus synthetic locus and Sox2 SCR at 1.5 kb”
mm10 plus synthetic locus and Sox2 SCR at 1.5 kb: chr7:103,768,826-103,866,530
LP 1.5 kb with Sox2 SCR
97 kb
ATAC - Seq
V6.5
reporter only
Map4k3 E2 at 1.5 kb
Map4k3 E2 at 75 kb
Sox2 SCR at 1.5 kb
Sox2 SCR at 75 kb
Rep 1
Rep 2
Olfr67
Hbb y
LP 75 kb
LP 25 kb
mCherry reporter
zoomed in to the enhancer:
zoomed in to the reporter:
mm10 plus synthetic locus and Sox2 SCR at 1.5 kb:: chr7:103,849,150-103,857,735
LoxP after recombination
Sox2 SCR
LoxP (72) after recombination
Spacer
mm10 plus synthetic locus and Sox2 SCR at 1.5 kb: chr7:103,854,877-103,861,228
Sox2 SCR after recombination
LoxP (72)
Spacer
mCherry reporter
polyA

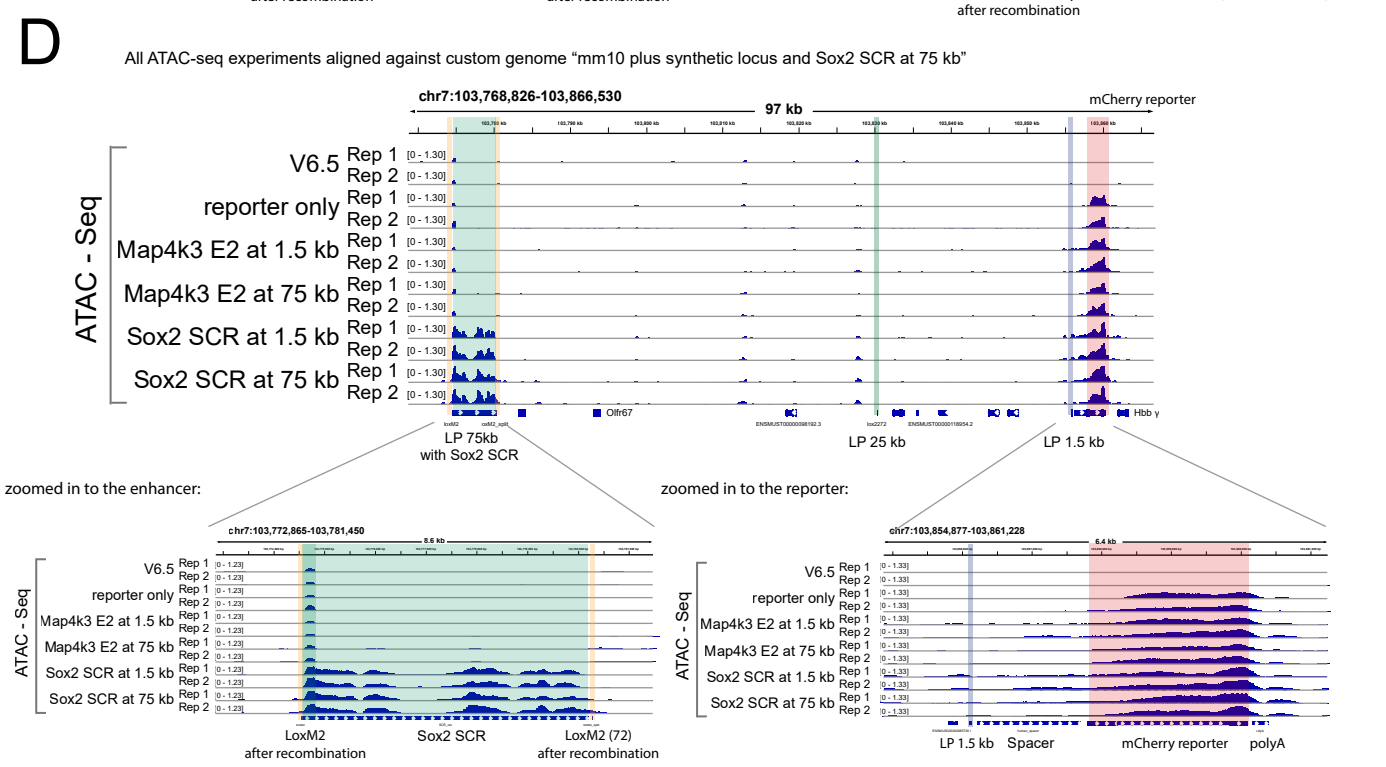

D
All ATAC-seq experiments aligned against custom genome “mm10 plus synthetic locus and Sox2 SCR at 75 kb”
chr7:103,768,826-103,866,530
97 kb
mCherry reporter
ATAC - Seq
V6.5
reporter only
Map4k3 E2 at 1.5 kb
Map4k3 E2 at 75 kb
Sox2 SCR at 1.5 kb
Sox2 SCR at 75 kb
Rep 1
Rep 2
Olfr67
Hbb y
LP 75kb with Sox2 SCR
LP 25 kb
LP 1.5 kb
zoomed in to the enhancer:
zoomed in to the reporter:
chr7:103,772,865-103,781,450
LoxM2 after recombination
Sox2 SCR
LoxM2 (72) after recombination
chr7:103,854,877-103,861,228
LP 1.5 kb
Spacer
mCherry reporter
polyA

Figure S3: Expression analysis and ATAC seq in cell lines with the Sox2 SCR, related to Figure 1

A: Quantification of *Hbb-y* expression using qPCR in cell lines with the strong *Sox2 SCR* enhancer at the indicated LPs before (top) or after backbone removal (bottom).

B: Quantification of *Olfr67* expression using qPCR in cell lines with the strong *Sox2 SCR* enhancer at the indicated LPs before (top) or after backbone removal (bottom).

C: IGV browser track showing ATAC seq signal at the β-globin locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs. Sequencing reads were aligned against a custom genome containing the synthetic locus and the *Sox2 SCR* enhancer at 1.5 kb. Bottom: zoom into the integrated enhancer (left) and the mCherry reporter (right).

D: IGV browser track showing ATAC seq signal at the β-globin locus of the parental cell lines v6.5, the no enhancer control and different enhancers integrated at different LPs. Sequencing reads were aligned against a custom genome containing the synthetic locus and the *Sox2 SCR* enhancer at 75 kb. Bottom: zoom into the integrated enhancer (left) and the mCherry reporter (right).

Figure S4

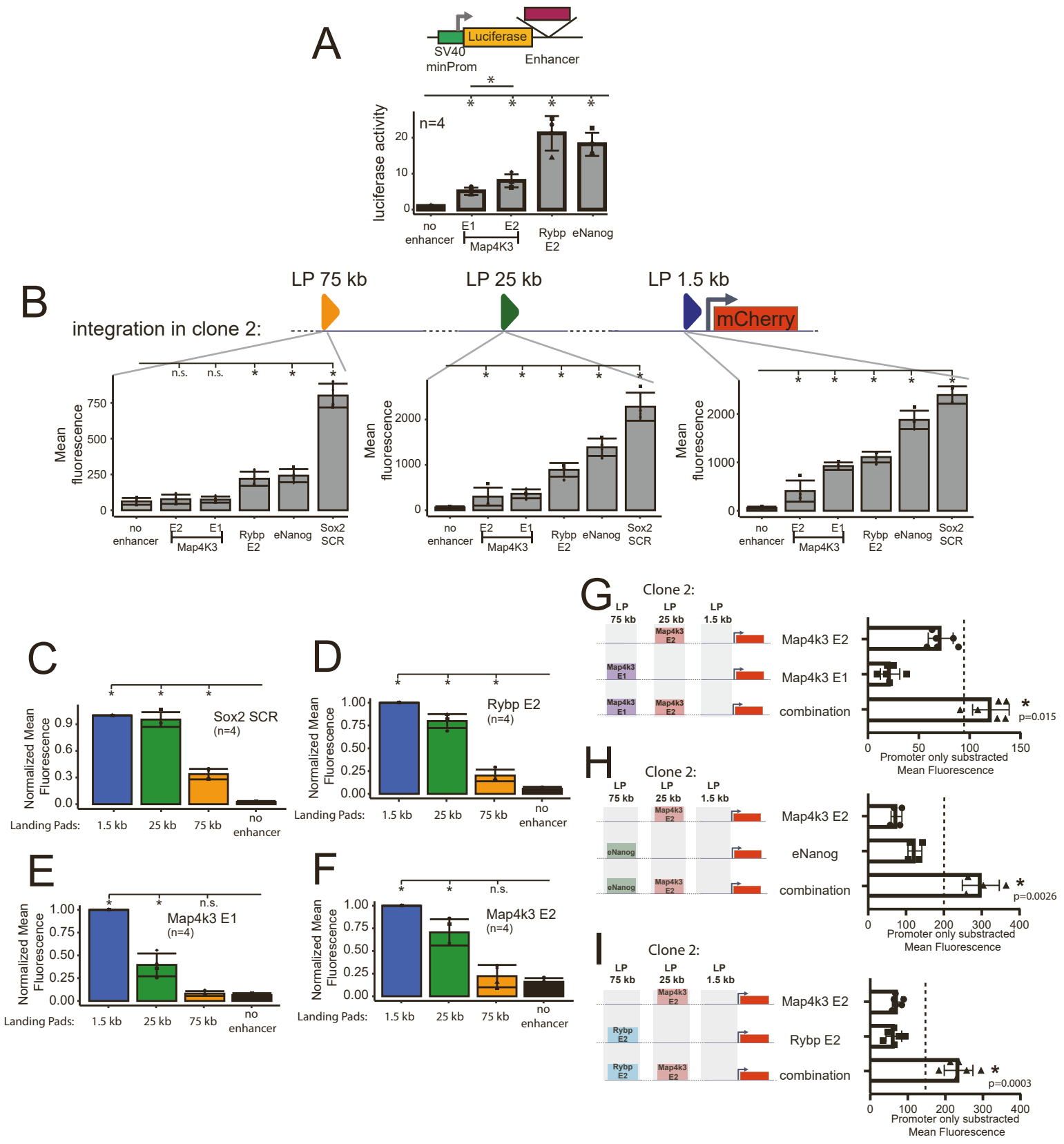

A
Luciferase
SV40 minProm
Enhancer
luciferase activity
n=4
no enhancer
E1
E2
Map4K3
Rybp E2
eNanog
B
integration in clone 2:
LP 75 kb
LP 25 kb
LP 1.5 kb
mCherry
Mean fluorescence
n.s.
no enhancer
E2
E1
Map4K3
Rybp E2
eNanog
Sox2 SCR
C
Sox2 SCR
(n=4)
Normalized Mean Fluorescence
Landing Pads:
1.5 kb
25 kb
75 kb
no enhancer
D
Rybp E2
(n=4)
E
Map4k3 E1
(n=4)
F
Map4k3 E2
(n=4)
G
Clone 2:
LP 75 kb
LP 25 kb
LP 1.5 kb
Map4k3 E2
Map4k3 E1
combination
p=0.015
Promoter only substracted Mean Fluorescence
H
eNanog
p=0.0026
I
Rybp E2
p=0.0003

Figure S4: Validation of enhancer integration in clone 2, related to Figures 2, 3 and 4

A: Luciferase Assay of selected short enhancers. Top: schematic depicting the integration of the enhancers. Statistical significance was tested by one-sided Welch one-sample t-tests on the resulting normalised values to assess whether they were significantly higher than 1 (the value of the no-enhancer control).

B: Quantification of mCherry expression of the five indicated enhancers compared to the no enhancer control at three different LPs in the clone 2 cell line. Left: integration at 75 kb, middle: integration at 25 kb, Right: integration at 1.5 kb. Statistically significant signals are marked by stars compared to the no enhancer control ($p < 0.05$, one-sided paired t-tests). n.s.: not significant, $p>0.05$.

C-F: Quantifying mean mCherry expression of indicated enhancers integrated at the indicated LPs in the clone 2 cell line. All mean expressions were normalised to the 1.5 kb integration. Statistically significant different signals ($p < 0.05$, one-sided paired t-tests) are marked by stars.

G-I: Comparison of combinations of enhancers (third row) to the individual integrations (top two rows) and the expected additive behaviour in the independently derived clone 2 cell line. Left: schematic of individual and combinations of integrations at the different LPs. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. The grey dashed line indicates the expected values under an additive model. Statistically significant signals compared to the expected additive modell ($p < 0.05$, one-sided paired t-tests) are marked by stars or n.s. (not significant).

G: *Map4k3 E1* at 75 kb combined with *Map4k3 E2* at 25 kb

H: *eNanog* at 75 kb combined with *Map4k3 E2* at 25 kb

I: *Rybp E2* at 75 kb combined with *Map4k3 E2* at 25 kb

Figure S5

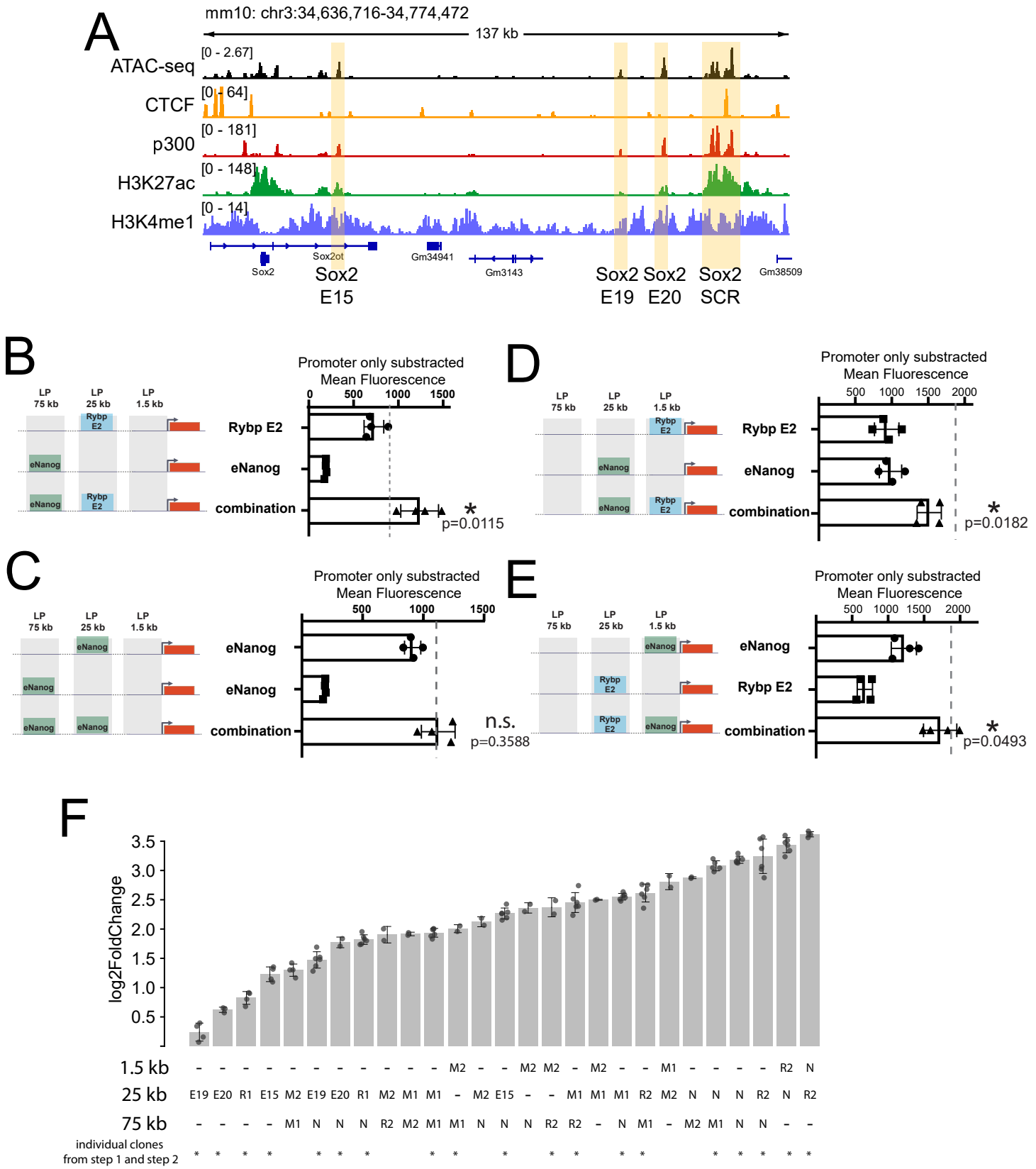

A
mm10: chr3:34,636,716-34,774,472
137 kb
ATAC-seq
[0 - 2.67]
CTCF
[0 - 64]
p300
[0 - 181]
H3K27ac
[0 - 148]
H3K4me1
[0 - 14]
Sox2
Sox2ot
Gm34941
Gm3143
Gm38509
Sox2 E15
Sox2 E19
Sox2 E20
Sox2 SCR
B
LP 75 kb
LP 25 kb
LP 1.5 kb
Rybp E2
eNanog
Promoter only substracted Mean Fluorescence
Rybp E2
eNanog
combination
p=0.0115
C
eNanog
eNanog
combination
p=0.3588
D
p=0.0182
E
p=0.0493
F
log2FoldChange
1.5 kb
25 kb
75 kb
individual clones from step 1 and step 2

Figure S5: Analysis of additional clonal enhancer cell lines, related to Figure 5

A: IGV browser track depicting ATAC-seq [1], CTCF [2], p300 [5], H3K27ac [5], and H3K4me1 [5] ChIP-seq tracks for the *Sox2* locus, the selected enhancers are indicated by a yellow box.

B-E: Comparison of combinations of enhancers (third row) to the individual integrations (top two rows) and the expected additive behaviour. Left: schematic of individual and combinations of integrations at the different LPs. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. The grey dashed line indicates the expected values under an additive model. Statistically significant signals compared to the expected additive modell ($p < 0.05$, one-sided paired t-tests) are marked by stars or n.s. (not significant).

B: *eNanog* at 25 kb combined with *Rybp E2* at 1.5 kb

C: *Rybp E2* at 25 kb combined with *eNanog* at 1.5 kb

B: *eNanog* at 75 kb combined with *Rybp E2* at 25 kb

B: *eNanog* at 75 kb combined with *eNanog* at 25 kb

F: log2 fold change of indicated individual clones were compared for the indicated combinations of enhancers. Indicated clones were obtained at two different steps of the cell line generation (after the integration of the indicated enhancer and then again after backbone removal), the clonal cell lines without a star were derived by picking clones after backbone removal.

Figure S6

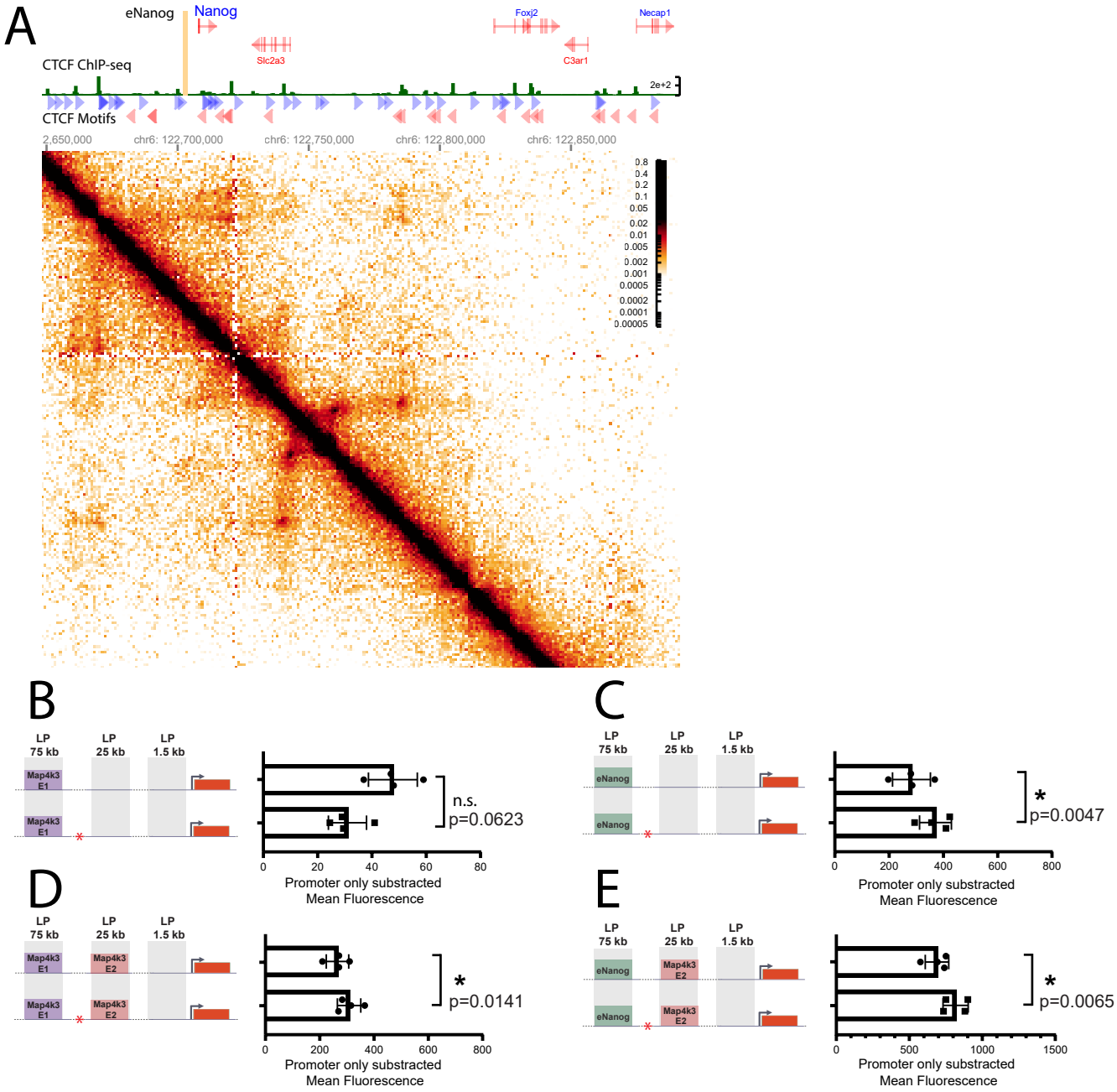

A
eNanog
Nanog
Foxj2
Necap1
Slc2a3
C3ar1
CTCF ChIP-seq
2e+2
CTCF Motifs
2,650,000
chr6: 122,700,000
chr6: 122,750,000
chr6: 122,800,000
chr6: 122,850,000
0.8
0.4
0.2
0.1
0.05
0.02
0.01
0.005
0.002
0.001
0.0005
0.0002
0.0001
0.00005
B
LP 75 kb
LP 25 kb
LP 1.5 kb
Map4k3 E1
Map4k3 E1
n.s.
p=0.0623
0
20
40
60
80
Promoter only substracted Mean Fluorescence
C
LP 75 kb
LP 25 kb
LP 1.5 kb
eNanog
eNanog
*
p=0.0047
0
200
400
600
800
Promoter only substracted Mean Fluorescence
D
LP 75 kb
LP 25 kb
LP 1.5 kb
Map4k3 E1
Map4k3 E2
Map4k3 E1
Map4k3 E2
*
p=0.0141
0
200
400
600
800
Promoter only substracted Mean Fluorescence
E
LP 75 kb
LP 25 kb
LP 1.5 kb
eNanog
Map4k3 E2
eNanog
Map4k3 E2
*
p=0.0065
0
500
1000
1500
Promoter only substracted Mean Fluorescence

Figure S6: Individual mutation of CTCF peak 1 has minor effects in gene expression, related to Figure 6

A: Top: Positions of the *Nanog* enhancer; Middle: ChIP-seq against CTCF [2] and direction of CTCF motifs; Bottom: MicroC data [6] across the *Nanog* locus

B-E: Comparison of mCherry expression before and after deletion of the peak 1 CTCF site. Left: schematic of individual or combinations of integrations at the different LPs and the deletion of the CTCF sites. The mean mCherry expression was calculated, and then the mean expression of the no-enhancer control was subtracted from all individual and combination mean mCherry expressions. Statistically significant signals ($p < 0.05$, one-sided paired t-tests) are marked by stars or n.s. (not significant).

B: *Map4k3 E1* integrated at 75 kb with and without peak 1 CTCF site

C: *eNanog* integrated at 75 kb with and without peak 1 CTCF site

D: *Map4k3 E1* integrated at 75 kb and *Map4k3 E2* integrated at 25 kb with and without peak 1 CTCF site

E: *eNanog* integrated at 75 kb and *Map4k3 E2* integrated at 25 kb with and without peak 1 CTCF site

Table S1: gRNA sequences used in this study, related to STAR Methods

integration of mCherry reporter gene forward	CACCGGGCCCTCCTCGACCCCATT
integration of mCherry reporter gene reverse	AAACAATGGGGTCGAGGAGGGCCC
lox KI 25 kb forward	CACCGTGTGTACCTAACAAGATATG
lox KI 25 kb reverse	AAACCATATCTTGTTAGGTACACAC
lox KI 75 kb forward	CACCGTTCACTATGAGCAGGGCATC
lox KI 75 kb reverse	AAACGATGCCCTGCTCATAGTGAAC
Map4k3 E1 gRNA forward 1	CACCGCAAGGGACAGTATTGCTGCG
Map4k3 E1 gRNA reverse 1	AAACCGCAGCAATACTGTCCCTTGC
Map4k3 E1 gRNA forward 2	CACCGACACAGTAGAGCTTGAATGC
Map4k3 E1 gRNA reverse 2	AAACGCATTCAAGCTCTACTGTGTC

Map4k3 E2 gRNA forward 1	CACCGACTCCTGCCTAAAACCGCGT
Map4k3 E2 gRNA reverse 1	AAACACGCGGTTTTAGGCAGGAGTC
Map4k3 E2 gRNA forward 2	CACCGAAGTCTAAACACCGTGAGG
Map4k3 E2 gRNA reverse 2	AAACCCTCACGGTGTTTAGACTTC
Rybp E2 gRNA forward 1	CACCGTACCCCTTAATGAGATGGGT
Rybp E2 gRNA reverse 1	AAACACCCATCTCATTAAGGGGTAC
Rybp E2 gRNA forward 2	CACCGTAGAGCTAAATAAAAGGTGC
Rybp E2 gRNA reverse 2	AAACGCACCTTTTATTTAGCTCTAC
CTCF-peak1-KO gRNA forward	CACCGTGTTCTCAGTCAACCTCAAG
CTCF-peak1-KO gRNA reverse	AAACCTTGAGGTTGACTGAGAACAC
CTCF-peak2-KO gRNA forward1	CACCGCTCTTGGGCTAAACGTCTA
CTCF-peak2-KO gRNA reverse1	AAACTAGACGTTTAGCCCAAGAGC

CTCF-peak2-KO gRNA forward1	CACCGCTAGGTCCGAGGGTAGGTC
CTCF-peak2-KO gRNA reverse1	AAACGACCTACCCTCGGACCTAGC

Table S3: RT-qPCR primers used in this study, related to STAR Methods

Rybp forward	AGACCAGCGAAACAAACCAC
Rybp reverse	GGAGGAGGAGCGAGTCTTTT
Map4k3 forward	GGTGTGGTCATATTACAA
Map4k3 reverse	CAGGATTTATGTCCATATTAC
Hbb g for	TCA TCA ATG GCC TGT GGA GT
Hbb g rev	GGG TCC ATG GGT ACA CAA CA
Olfr67 FOR	CAA AAC AGT TCC TCG CCA CA
Olfr67 rev	GAT GGG GTT AAG CAT TGG GG

Table S4: ssDNA oligos used in this study, related to STAR Methods

ssDNA KI 25 kb	GGGCTTGTAAACAAAAAAGCTAAACTACAC TGTTAAGTCCCTTCCTGTTTGTGTACCTAA CAAGATTACCGTTCGTATAGGATACTTTATA CGAAGTTATATGTGGTACACTTATCAATAGA GTATTGAGATGAATATAATTGAATGCTA
ssDNA KI 75 kb	AAAATGAAGTCTGAAAGCAAAATGGAGTTT GTAATGTTAAACTGGGCTCTTCACTATGAG CAGGGCATAACTTCGTATAAGAAACCATATA CGAACGGTAATCTGGATCTGTTTTCACATT GTTCCTTTCTCTTTGCCTACAATTTCAGG
FRT forward	CATGAGAAGTTCCTATTCTCTAGAAAGTATA GGAACTTCCCATGG
FRT reverse	CTAGCCATGGGAAGTTCCTATACTTTCTAG AGAATAGGAACTTCT
FRT3 forward	CATGAGAAGTTCCTATTCTTCAAATAGTATA GGAACTTCCCATGG
FRT3 reverse	CTAGCCATGGGAAGTTCCTATACTATTTGA AGAATAGGAACTTCT
FRT5 forward	CATGAGAAGTTCCTATTCTTCAAAAGGTATA GGAACTTCCCATGG

FRT5 reverse	CTAGCCATGGGAAGTTCCTATACCTTTTGAAGAATAGGAACTTCT
lox66 forward	CATGATAACTTCGTATAGCATACATTATACGAACGGTAGC
lox66 reverse	GGCCGCTACCGTTCGTATAATGTATGCTATACGAAGTTAT
lox2272/66 forward	CATGATAACTTCGTATAGGATACTTTATACGAACGGTAGC
lox2272/66 reverse	GGCCGCTACCGTTCGTATAAAGTATCCTATACGAAGTTAT
loxm2/66 forward	CATGATAACTTCGTATATGGTTTCTTATACGAACGGTAGC
loxm2/66 reverse	GGCCGCTACCGTTCGTATAAGAAACCATATACGAAGTTAT

Table S5: Summary of Landing Pads and integration efficiencies, related to Figure 1

	lox71 / lox66	**FRT**	**Selection Cassette**	**percentage of positive clones**
Landing Pad 1.5 kb	loxP	FRT	Blasticidine-deltaTK	25%
Landing Pad 25 kb	Lox2272	FRT5	Hygromycin-delta TK	19%
Landing Pad 75 kb	loxm2	FRT3	Puromycin-delta TK	24.30%

Explanations on the lox sites: The lox66 and lox71 sites are listed together since the landing pad carries the lox71 mutation, but the targeting vector carries the lox66 mutation, both are build on the same incompatible lox sites mentioned in the table. Once the lox66 cites recombines with the lox71 site, two new lox sites are formed, one WT site and one with both mutations (lox72).

Table S6: Detailed Results from the Modelling, related to Figure 5

Residuals:

```
    Min      1Q  Median      3Q     Max
-0.23323 -0.09984 -0.02701  0.12613  0.37757
```

Coefficients: (1 not defined because of singularities)

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.90099    0.22990   3.919  0.00122 **
indA            0.55119    0.23300   2.366  0.03096 *
indB            0.64666    0.10600   6.100 1.53e-05 ***
indC            0.61416    0.11471   5.354 6.46e-05 ***
indA:indB      -0.04428    0.11639  -0.380  0.70861
indA:indC      -0.09821    0.14464  -0.679  0.50683
indB:indC      -0.12623    0.05119  -2.466  0.02535 *
indA:indB:indC       NA         NA      NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.1967 on 16 degrees of freedom

Multiple R-squared: 0.9276, Adjusted R-squared: 0.9004

F-statistic: 34.16 on 6 and 16 DF, p-value: 2.977e-08

Supplemental References

1. Romeike, M., Spach, S., Huber, M., Feng, S., Vainorius, G., Elling, U., Versteeg, G.A., and Buecker, C. (2022). Transient upregulation of IRF1 during exit from naive pluripotency confers viral protection. EMBO Rep. *23*, e55375. https://doi.org/10.15252/embr.202255375.
2. Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Saldaña-Meyer, R., Reinberg, D., Darzacq, X., and Tjian, R. (2019). Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. Mol. Cell *76*, 395-411.e13. https://doi.org/10.1016/j.molcel.2019.07.039.
3. Spencley, A.L., Bar, S., Swigut, T., Flynn, R.A., Lee, C.H., Chen, L.-F., Bassik, M.C., and Wysocka, J. (2023). Co-transcriptional genome surveillance by HUSH is coupled to termination machinery. Mol. Cell *83*, 1623-1639.e8. https://doi.org/10.1016/j.molcel.2023.04.014.
4. Yang, P., Humphrey, S.J., Cinghu, S., Pathania, R., Oldfield, A.J., Kumar, D., Perera, D., Yang, J.Y.H., James, D.E., Mann, M., et al. (2019). Multi-omic Profiling Reveals Dynamics of the Phased Progression of Pluripotency. Cell Syst. *8*, 427-445.e10. https://doi.org/10.1016/j.cels.2019.03.012.
5. Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. Cell Stem Cell *14*, 838–853. https://doi.org/10.1016/j.stem.2014.04.003.
6. Hsieh, T.-H.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R., and Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. Mol. Cell *78*, 539-553.e8. https://doi.org/10.1016/j.molcel.2020.03.002.

Chapter II: Transient upregulation of IRF1 during the exit from naïve pluripotency confers viral protection

Romeike, Merrit, Stephanie Spach, Marie Huber, Songjie Feng, Gintautas Vainorius, Ulrich Elling, Gjis A Versteeg, Christa Buecker

Status: Published

Contribution:

Pluripotent stem cells intrinsically express a subset of interferon-stimulated genes (ISGs), independent of external stimuli such as viral infection. However, the mechanism underlying the regulation of ISGs is largely unknown. Here, we found that formative pluripotency gene regulatory network regulates the expression of a key ISG named *Irf1*, which is transiently upregulated during the transition from naive to formative pluripotency. We further identify *Irf1* as a regulator of an enhancer controlling Oct6, a marker of formative pluripotency. Loss of *Irf1* compromises the expression of a specific subset of ISGs and renders cells exiting naive pluripotency more vulnerable to viral infection. I contributed to quantifying the function of an EpiLC-specific enhancer regulating *Irf1*. To assess its loss-of-function effect, I first generated several independent cell lines in which this enhancer was knocked out. The knockout cell lines were verified by PCR genotyping followed by Sanger sequencing. The KO cell lines showed reduced mRNA and protein expression of IRF1 validated by qPCR and Western Blot. The results indicate that IRF1 expression is regulated by an enhancer that is controlled by the formative pluripotency gene regulatory network.

Article

TRANSPARENT PROCESS OPEN ACCESS

EMBO reports

Transient upregulation of IRF1 during exit from naive pluripotency confers viral protection

Merrit Romeike[1,2], Stephanie Spach[1], Marie Huber[1], Songjie Feng[1,2], Gintautas Vainorius[2,3], Ulrich Elling[3], Gjis A Versteeg[1] & Christa Buecker[1,*]

Abstract

Stem cells intrinsically express a subset of genes which are normally associated with interferon stimulation and the innate immune response. However, the expression of these interferon-stimulated genes (ISG) in stem cells is independent from external stimuli such as viral infection. Here, we show that the interferon regulatory factor 1, *Irf1*, is directly controlled by the murine formative pluripotency gene regulatory network and transiently upregulated during the transition from naive to formative pluripotency. IRF1 binds to regulatory regions of a conserved set of ISGs and is required for their faithful expression upon exit from naive pluripotency. We show that in the absence of IRF1, cells exiting the naive pluripotent stem cell state are more susceptible to viral infection. *Irf1* therefore acts as a link between the formative pluripotency network, regulation of innate immunity genes, and defense against viral infections during formative pluripotency.

Keywords pluripotency; viral infection; cell fate transition; ISG; gene regulatory networks

Subject Categories Microbiology, Virology & Host Pathogen Interaction; Signal Transduction; Stem Cells & Regenerative Medicine

DOI 10.15252/embr.202255375 | Received 9 May 2022 | Revised 14 June 2022 | Accepted 23 June 2022 | Published online 19 July 2022

EMBO Reports (2022) 23: e55375

Introduction

Mouse embryonic stem cells (ESCs) are self-renewing in a naive pluripotent cell state *in vitro*. The essential gene regulatory network required to maintain naive pluripotency has been the subject of numerous studies and is very well-defined (Betschinger *et al*, 2013; Dunn *et al*, 2014; Leeb *et al*, 2014). Development although is characterized by tightly regulated cell fate transitions, which are required to establish the remarkable complexity of multicellular organisms. After exiting naive pluripotency, cells enter a transient state termed formative pluripotency (Smith, 2017). This transition is characterized by global reorganization of the enhancer landscape and upregulation of the formative gene regulatory network (Buecker *et al*, 2014a). However, factors involved in this formative pluripotency gene regulatory network are less well-understood: Genetic screens interrogating this transition have mostly focused on maintenance of (Li *et al*, 2018; Seruggia *et al*, 2019) or exit from naive pluripotency (Betschinger *et al*, 2013; Leeb *et al*, 2014; Li *et al*, 2018; Lackner *et al*, 2021). Formative pluripotency was only considered in the context of a multistep differentiation into primordial germ cell fate (Hackett *et al*, 2018).

Pluripotent and multipotent stem cells intrinsically express subsets of genes which are referred to as interferon-stimulated genes (ISGs; Wu *et al*, 2018). The subsets of expressed ISGs are distinct between different stem cell types and change in differentiation (Wu *et al*, 2018). How these ISGs are regulated is unknown, as mouse ESCs for example do not respond to interferon (IFN) stimulation (Burke *et al*, 1978; Harada *et al*, 1990; Wang *et al*, 2013; Chen *et al*, 2020). Strikingly, pluripotent stem cells are more resistant against viral infection compared with their differentiated descendants (Swartzendruber & Lehman, 1975; Wolf & Goff, 2009), and this is at least in part due to ISG expression (Wu *et al*, 2018). However, whether ISGs and pluripotency gene expression are functionally connected has remained unknown.

Here, we identify the ISG *Irf1* in a CRISPR-KO screen as a regulator of an enhancer that controls the expression of *Oct6*, a formative marker gene. IRF1 has a conserved role as regulator of other ISGs. Interestingly, many ISGs are differentially expressed between naive and formative pluripotency, and we show that IRF1 is directly responsible for the expression of a subset of these genes. Furthermore, IRF1 confers viral protection to cells during the exit from naive pluripotency. Finally, transient *Irf1* expression in formative pluripotent cells is directly regulated by the formative pluripotency gene regulatory network through a stem cell-specific enhancer. Pluripotency and interferon responsive gene regulatory networks are therefore connected through *Irf1* during the transition from naive to formative pluripotency.

1 Max Perutz Labs Vienna, Vienna Biocenter (VBC), University of Vienna, Vienna, Austria
2 Vienna Biocenter PhD Program, A Doctoral School of the University of Vienna and Medical University of Vienna, Vienna, Austria
3 Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Vienna, Austria
*Corresponding author. Tel: +43 1 4277 74646; E-mail: christa.buecker@univie.ac.at

Results

Monitoring exit from naive and entry into formative pluripotency with a dual fluorescent reporter

We designed a fluorescent reporter system which simultaneously monitors exit from naive and entry into formative pluripotency. Murine ESCs are cultured under defined 2i + LIF conditions, which preserve the naive state of pluripotency. Upon withdrawal of 2i + LIF and stimulation with FGF2, the cells differentiate within 48 h irreversibly from the ESC state into epiblast-like cells (EpiLC; Hayashi *et al*, 2011; Buecker *et al*, 2014a), often referred to as the formative state of pluripotency (Smith, 2017). Two differentially active enhancer elements (Appendix Fig S1A) were cloned into reporter constructs to activate different fluorescent markers. Thereby, we could follow the exit from naive pluripotency and the entry into formative pluripotency simultaneously: a naive-specific enhancer close to *Tbx3* (*eTbx3*) controls GFP expression and a formative-specific enhancer close to *Oct6* (also known as *Pou3f1*, *eOct6*) drives mCherry expression (Buecker *et al*, 2014a). Both constructs and *Cas9* cDNA were introduced into ESC, and two independent clonal cell lines were selected which showed clear separation of ESC and EpiLC populations after 48 h of differentiation by FACS analysis (Appendix Fig S1B–D). We tested the screening set-up through deletion of *Tcf7l1* (also known as *Tcf3*): Loss of the transcription factor *Tcf7l1* causes a strong delay in the exit from naive pluripotency (Wray *et al*, 2011). As expected, *Tcf7l1* knockout (KO) caused a shift in dual fluorescent marker expression in the EpiLC population (Appendix Fig S1E), with mCherry showing lower expression than the WT cells and GFP showing higher expression in the *Tcf7l1* KO cells 48 h after initiation of differentiation. In summary, the dual fluorescent reporter enables faithful monitoring of the transition from naive into formative pluripotency.

Screening for factors regulating entry into formative pluripotency

Next, we performed a pooled CRISPR KO screen with the dual reporter cell line to identify factors regulating the entry into formative pluripotency. We transduced the reporter cell lines (100 million cells) with a lentiviral library for the expression of 22,781 unique sgRNAs targeting 2,524 nuclear localized genes, including 159 genes considered ISGs (Figs 1A and EV1A, Dataset EV1, Dataset EV3, Dataset EV4). We expected a candidate regulator of formative pluripotency to modify transcription and such a small, selected library increases statistical robustness. We also infected wild-type ESC (R1 cells) lacking CAS9 expression to control for sgRNA abundance in the lentiviral library. We selected infected cells with Neomycin for 48 h. After recovery from selection for an additional 48 h, we harvested 60 million cells to identify ESC essential factors: We compared the abundance of sgRNA coding sequences in genomic DNA from CAS9 expressing or non-expressing cells. As expected, we observed enrichment of *Trp53* sgRNA, as cells lacking *p53* proliferate faster and enrich in populations of mixed clones (Sabapathy, 1997; Fig 1B). Conversely, ESCs lacking core pluripotency factors including OCT4 (also known as POU5F1) and NANOG depleted from this population, as these factors are required for naive pluripotency (Fig 1B). Moreover, larger sets of known essential factors showed depletion (Figs 1B and EV1B; Hart *et al*, 2017; Li *et al*, 2018). We conclude that the overall screening set-up could identify known essential regulators of the naive pluripotent stem cell state.

Next, we applied our screening platform to identify potential activators of the formative pluripotency network. We differentiated ESCs for 48 h into EpiLC and binary scored EpiLCs as differentiation "impaired" or "unimpaired": Cells with reporter activity overlapping the behavior of no sgRNA controls were FACS-sorted as unimpaired. Conversely, cells with higher *eTbx3* controlled GFP and/or lower *eOct6* controlled mCherry were sorted as impaired (see Figs 1A, and EV1C and D for gating strategy). In addition, we also collected unsorted EpiLCs as baseline. This strategy allowed several comparisons of sgRNA abundance: KO of a candidate driving the formative state should increase the probability of impaired differentiation; therefore, the sgRNAs should be enriched in the impaired populations. This was the case for a total of 72 factors (Fig 1C). Conversely, a candidate KO should be incompatible with unimpaired differentiation, and therefore, sgRNAs drop out from the unimpaired population. Two hundred and eighty-five factors showed such a behavior (Fig 1D). We identified an overlapping set of 17 hits, which were found in both comparisons and are therefore the most stringent candidates involved in this cell fate transition (Figs 1E, and EV1E and F). Among these candidates were members of pathways with known involvement in stem cell differentiation: The FGF receptor *Fgfr1*, *Tead1* which plays a role in the Hippo pathway (Molotkov *et al*, 2017) and *Zic3*, a known regulator of the exit from naive pluripotency and entry into primed pluripotency (Yang *et al*, 2019).

In addition, we identified several members of the SWI/SNF chromatin remodeling family, also called BAF complexes (*Arid1a*, *Smarce1*, and *Smarcad1*). Specific subunits of BAF complexes are connected to diverse phenotypes in ESCs (reviewed in Ye *et al*,

Figure 1. Screening for factors required for activation of formative pluripotency.

A Outline of CRISPR-KO screening strategy. Reporter ESCs were transduced with a pooled CRISPR-KO library. Cells selected for library integration with and without CAS9 were compared, assaying genes essential for the ESC state. Reporter cell lines were differentiated and scored by flow cytometry sorting as differentiation impaired (blue box) or unimpaired (red box). See also Fig EV1C and D for gating strategy.

B–D Volcano plots of indicated comparisons in CRISPR-KO screen analysis. The *x*-axis represents fold change of sgRNA presence between indicated conditions, the *y*-axis shows the binomial *P*-value. Non-targeting control sgRNA are indicated in black. All analyses are based on two replicates in two independent cell lines. (B) sgRNA representation in ESCs with CAS9 vs. ESCs without CAS9. Indicated are known ESC essential genes (Hart *et al*, 2017) and selected factors. See also Fig EV1B. (C) sgRNA representation in EpiLC sorted as impaired vs. non-sorted EpiLC. Indicated in blue are factors enriched with FDR < 0.1. See also Fig EV1E. (D) sgRNA representation in EpiLC sorted as unimpaired vs. non-sorted ESCs. Indicated in red are factors depleted with FDR < 0.1. See also Fig EV1F.

E Overlap of all factors enriched in impaired and/or depleted from unimpaired populations, candidate factors found under both conditions are indicated.

F MA plot of gene expression changes in EpiLC vs. ESC. The candidate factors which were identified as shared between both screening conditions (Fig 1E) are indicated in red. *P*-value for *Irf1* calculated with DESeq2. $n = 2$ biological replicates.

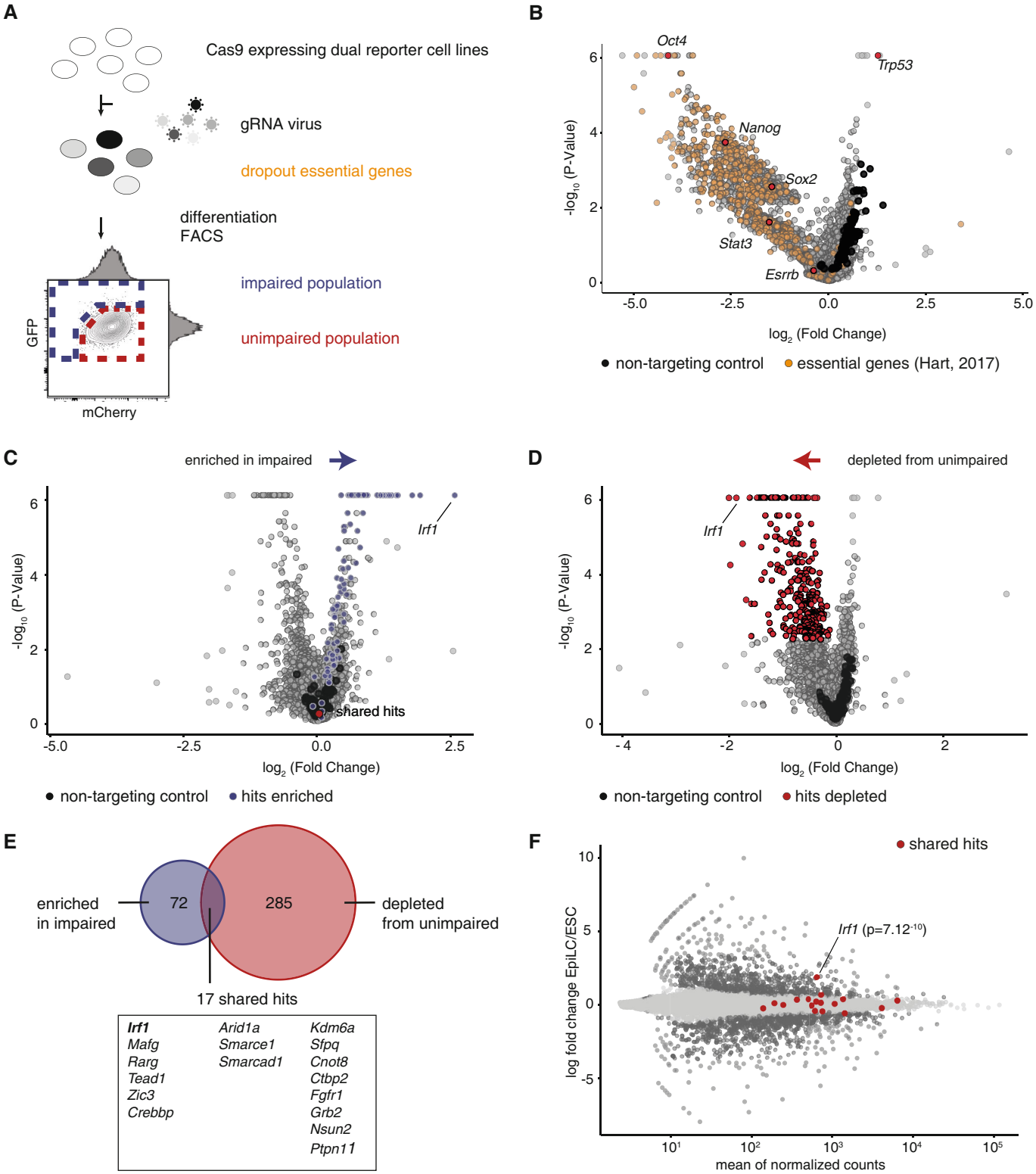


Figure 1.

2021). SMACRCE1/BAF57 is a canonical BAF subunit, and ARID1A/BAF250A is a component of the ESC-specific esBAF, with *Arid1a*-deficient mice showing an arrest in early embryonic development (Gao *et al*, 2008). SMARCAD1 has been shown to silence endogenous retrovirus in ESCs (Sachs *et al*, 2019). In sum, we were able to identify known regulators of the exit from naive pluripotency, demonstrating that our screening approach can identify factors involved in the exit from naive pluripotency.

One of the top screening hits was the interferon regulatory factor 1, *Irf1*, a transcription factor and a member of ISGs. *Irf1* was depleted from the unimpaired cell population and enriched in the impaired population for every single analyzed condition. We validated gene expression levels for all hits in ESC and EpiLC cells (Fig 1F). All hits are expressed, but only *Irf1* shows significant upregulation in differentiation. Therefore, we focused on *Irf1* and explored its function during the exit from naive pluripotency.

Irf1 is activated by the formative pluripotency gene regulatory network

As *Irf1* is upregulated during exit from naive pluripotency, we analyzed the chromatin environment surrounding the *Irf1* gene to identify potential regulatory regions that might control *Irf1* expression in EpiLCs. We identified a putative enhancer region 10 kb upstream of *Irf1*. In ESCs, this element is not strongly marked by active enhancer marks. However, in EpiLCs, this element is marked by OCT4, P300, the formative transcription factor OTX2 and the active histone modifications H3K27ac and H3K4me1 (Fig 2A). KO of *Otx2* has limited effect on the exit from naive pluripotency, but *Otx2* KO EpiLCs have significantly lower expression of *Irf1*, suggesting that the pluripotency network controls the expression of *Irf1* (Buecker *et al*, 2014a). Importantly, this putative enhancer is specific for the ESC to EpiLC transition as it was not marked by H3K27ac in unstimulated and IFN-ɣ exposed bone marrow-derived macrophages (BMDMs). In contrast, a region 6 kb upstream of *Irf1* is marked by increasing H3K27ac in BMDMs after IFN-ɣ stimulation. This suggests a cell type-specific regulation mechanism for *Irf1* during the exit from naive pluripotency.

We deleted the identified putative enhancer region to analyze whether the element indeed controls the expression of *Irf1*. We generated four independent enhancer KO cell lines and validated the absence of the enhancer region by genotyping PCRs and sequencing of the PCR products (Fig EV2A and B). We then tested whether *Irf1* expression is perturbed during differentiation into EpiLCs upon enhancer loss. Indeed, enhancer KO lines showed significantly reduced *Irf1* mRNA levels (Fig 2B). We also confirmed reductions in IRF1 protein levels by Western blot (Fig EV2C). The formative marker OTX2 is not influenced by the enhancer KO, and lack of IRF1 expression is therefore not caused by lack of transition to the EpiLC state. We conclude that *Irf1* expression in EpiLCs is directly connected to the formative network by enhancer regulation.

In the IFN response, JAK/STAT signaling is a known regulator of *Irf1* expression and JAK/STAT is also active in naive and formative pluripotency. We therefore tested whether JAK/STAT signaling activates *Irf1* expression in formative pluripotency. We treated cells with the JAK inhibitor ruxolitinib at the onset of differentiation. Treatment drastically reduced phosphorylation levels of STAT3 (Fig EV2D) but did not interfere with differentiation, as *Tbx3* and *Otx2* as naive and formative markers, respectively, were not misregulated (Fig EV2E and F). *Irf1* mRNA levels were not significantly reduced by ruxolitinib treatment in WT cells (Fig 2C), and ruxolitinib did not decrease *Irf1* expression further in enhancer KOs (Fig 2C). Together, these results demonstrate that direct action of the formative pluripotency network, and not IFN signaling, is the main driver of *Irf1* expression during the exit from naive pluripotency.

Transient IRF1 expression in the epiblast is conserved

We wondered whether upregulation of *Irf1* during the exit from naive pluripotency is limited to the *in vitro* 2D cell culture differentiation system. Therefore, we analyzed publicly available RNA-seq data collected from peri- and early postimplantation murine embryos at days E5.0, E5.5, and E6.5 (Fig 2D; Kinoshita *et al*, 2021a; Data ref: Kinoshita *et al*, 2021b). *Irf1* expression is transiently increased during implantation at E5.0, but rapidly decreases again within 1 day of development. *Irf1* is specific to the epiblast, as E4.5 primitive endoderm cells are not showing *Irf1* expression (Fig EV2G; Boroviak *et al*, 2015a; Data ref: Boroviak *et al*, 2015b). Accordingly, *Irf1* expression in a single cell dataset spanning mouse gastrulation is highest in epiblast and primitive streak metacells and downregulated in later states (Mittnenzweig *et al*, 2021). We conclude that the expression of *Irf1* is transiently upregulated in implantation stage embryos; therefore, the expression of *Irf1* in EpiLCs reflects the *in vivo* expression patterns observed in murine peri-implantation epiblasts.

We further asked whether *Irf1* expression is conserved in humans. Due to the limited availability of human embryonic data, we used a published dataset which integrates preimplantation single-cell expression data with data collected from blastocysts derived at 5 d.p.f. and cultured until 9 and 11 d.p.f., representing the pre- and postimplantation epiblast (Molè *et al*, 2021a; Data ref: Molè *et al*, 2021b). Cells showed an increase in *Irf1* expression in peri- and early postimplantation epiblasts (Fig 2E). The transient expression of *Irf1* is a conserved feature of implantation between murine and human embryogenesis.

IRF1 is a regulator of interferon-stimulated genes during pluripotency

Next, we wanted to understand the function of IRF1 in the exit from naive pluripotency and generated *Irf1* KO cell lines. We simultaneously transfected two CRISPR sgRNAs to delete an exon–intron boundary around intron 3 of *Irf1* into our dual reporter cell lines and R1 WT cells using lipofection. We identified individual clones through genotyping PCR and validated the absence of full-length protein by Western blot (Figs 3A and EV3A). Control ESCs showed very low IRF1 levels, but in *Irf1* KO cells, the protein is absent. IRF1 is localized to the nucleus, with homogenous faint expression in ESCs and robust homogenous signals in EpiLCs (Fig 3B). This signal in EpiLCs is lost upon deletion of *Irf1* (Fig 3B).

Next, we asked whether IRF1 affects gene expression in differentiation from ESC to EpiLCs. We performed RNA-seq using the Quantseq protocol for WT and two independent *Irf1* KO cell lines under ESC and EpiLC conditions (Dataset EV5). Principal component analysis captured differentiation along PC1 and genotype-specific effects along PC2 (Fig 3C). These results indicate that *Irf1* KO cells do not show drastically changed cell states in differentiation. To corroborate these findings, we focused on well-established pluripotency markers, expressed either in ESCs (such as *Klf4*, *Esrrb*, and *Nanog*), or in EpiLCs (for example, *Fgf5* and *Otx2*) and found little change in expression. With the exception of *Oct6*, which was reduced (Fig 3D; all shown factors *P*-value control vs. *Irf1* KO > 0.05, besides Tbx3 (*P*-value = 0.031)). Previously, a list of 496 genes which are either positively or negatively associated with the core pluripotency markers was described (naive-associated genes,

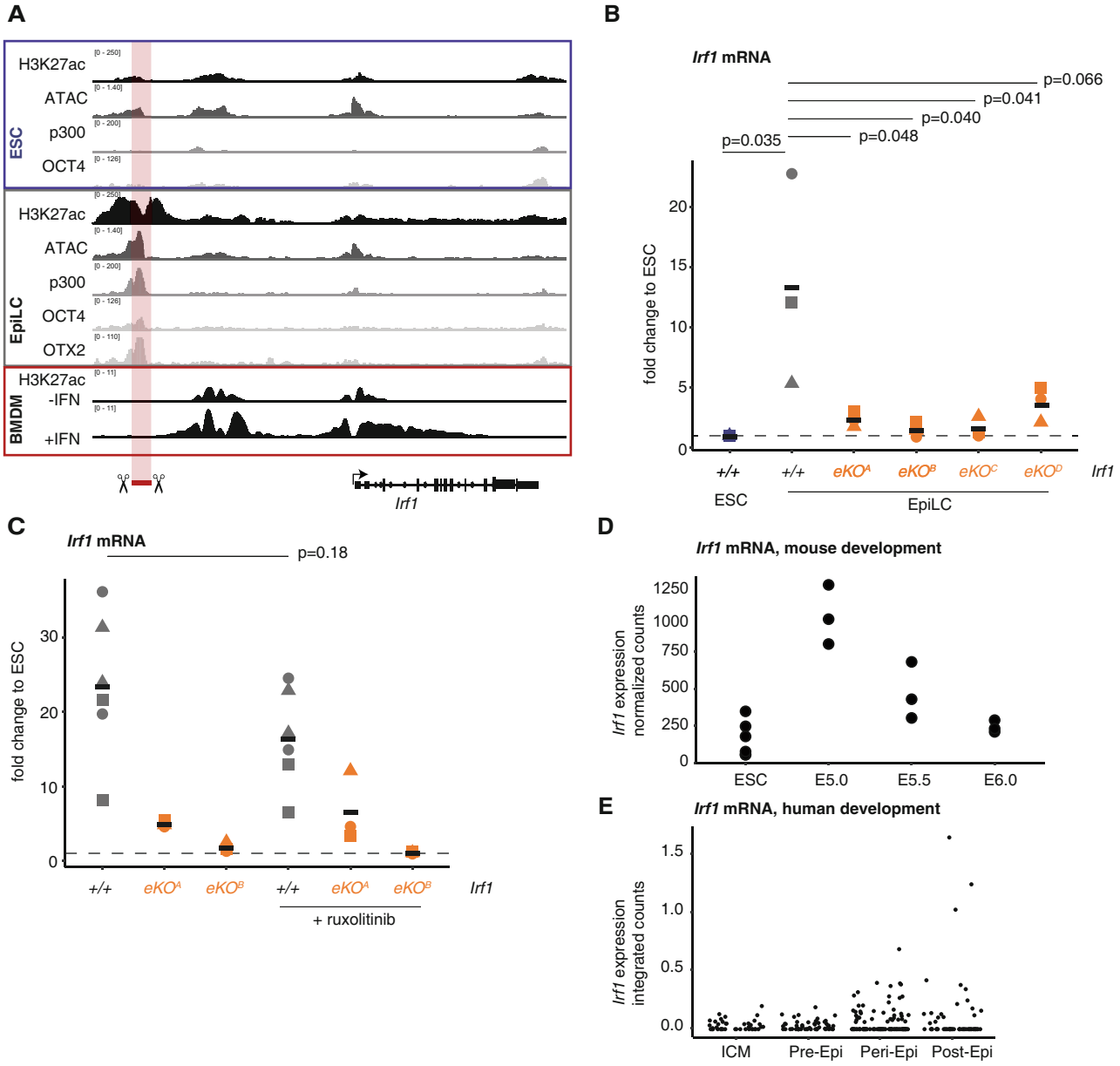


Figure 2. Formative pluripotency drives transient *Irf1* expression.

A Chromatin context of *Irf1* in ESCs, EpiLC and BMDMs with and without IFNγ stimulation. Shown are ChIP-seq tracks of histone mark H3K27ac, transcription factors OCT4, OTX2, and p300 and ATAC seq profiles (data were generated by Data ref: Buecker *et al*, 2014b; BMDM Data ref: Langlais *et al*, 2016b). The putative enhancer that was knocked out is indicated.

B RT-qPCR analysis of *Irf1* mRNA in $Irf1^{+/+}$ and *Irf1* enhancer KO in differentiation. Fold change is normalized against *Rpl13a* housekeeping mRNA expression and is calculated against ESC for each indicated cell line. This is shown as baseline with the dashed line at fold change = 1. Statistical tests were performed as homoscedastic one-sided *t*-tests. Shown are three biological replicates, datapoints from the same replicate are indicated with the same symbol. Black horizontal lines show the mean of the data.

C RT-qPCR analysis of *Irf1* mRNA in $Irf1^{+/+}$ and *Irf1* enhancer KO in differentiation, treated with ruxolitinib. Fold change is normalized against *Rpl13a* housekeeping mRNA expression and is calculated against ESC for each indicated cell line. This is shown as baseline with the dashed line at fold change = 1. Statistical test was performed as homoscedastic one-sided *t*-tests. Shown are three biological replicates, datapoints from the same replicate are indicated with the same symbol. Note that each replicate contains two $Irf1^{+/+}$ samples. Black horizontal lines show the mean of the data.

D *Irf1* expression in murine blastocysts and ESCs (Data ref: Kinoshita *et al*, 2021b). For E5.0, E5.5 and E6.0, entire single blastocysts were sequenced ($n = 3$ biological replicates), for ESCs $n = 4$ biological replicates.

E *Irf1* expression in human blastocysts. Integrated single-cell RNA-seq data from blastocysts collected at or cultured to indicated time points. Data are derived from a total of 16 embryos which passed quality control metrics (Data ref: Molè *et al*, 2021b).

NAGS (Lackner *et al*, 2021)). NAGs are also correlated with pre- to postimplantation expression changes in mice and macaque *in vivo*. As expected, NAGs show strong ESC- and EpiLC-specific expressions, but no drastic change upon *Irf1* deletion (Fig EV3B).

We tested whether *Irf1* KO cells displayed a general exit from naive pluripotency defect using alkaline phosphatase staining (Fig EV3C). We differentiated cells for 72 h to irreversibly exit naive pluripotency and plated them back under naive 2i + LIF conditions.

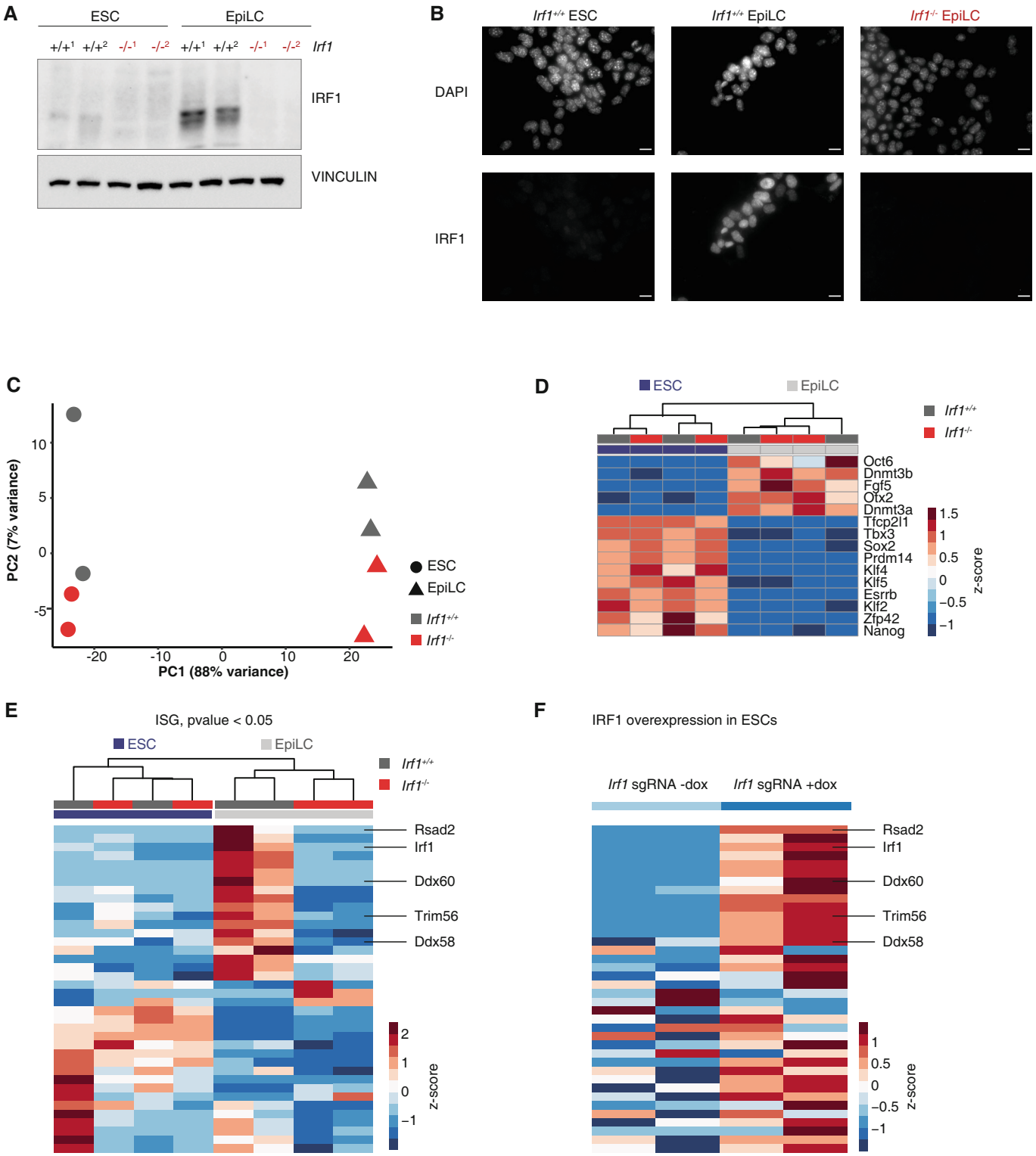


Figure 3.

Figure 3. IRF1 regulates ISG expression in EpiLCs.

A Western blot analysis of *Irf1*$^{-/-}$ in ESC and EpiLC, probed with antibodies against IRF1 and VINCULIN as loading control. Shown are two independent clonal cell lines. See Fig EV3A for long exposure and size marker of the same blot.

B Immunofluorescence with IRF1 antibody in *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ ESC and EpiLC. Nuclei are stained with DAPI. Scale bars are 10 μm.

C Dimension reduction (principal component analysis, PCA) of *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ ESC and EpiLC, based on QuantSeq RNA data. PCA captures global transcriptional states of samples and similarity of these states is reflected in closeness of samples in the plot. $n = 2$ biological replicates.

D Expression changes in selected pluripotency markers in *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ ESC and EpiLC Quantseq RNA data. Data are shown as gene normalized *z*-score. $n = 2$ biological replicates.

E Expression changes of ISGs in *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ ESC and EpiLC Quantseq RNA data. Shown are ISGs differentially expressed (*P*-value < 0.05) between *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ EpiLCs. Data are shown as gene normalized *z*-score. See also Fig EV3E. Selected gene examples are indicated. $n = 2$ biological replicates.

F Expression changes in the same genes as in Fig 3E in dox-inducible SunTag-based IRF1 overexpression. Data are shown gene normalized as *z*-score. Selected gene examples are indicated. $n = 2$ biological replicates.

Genotypes with exit from naive pluripotency defect show an increased fraction of cells still able to proliferate under 2i + LIF conditions (Leeb *et al*, 2014). *Irf1* KO cells did not show increased colony formation and therefore are not impaired in the exit from naive pluripotency. We conclude that IRF1 does not regulate exit from naive or entry into formative pluripotency.

Irf1 is a member of a group of genes previously classified as ISGs. A subset of ISGs are intrinsically expressed in stem cells of several species, even in the absence of viral stimulation (Wu *et al*, 2018). In differentiation to somatic cell types, this intrinsic ISG expression and thereby viral protection is lost (Wu *et al*, 2018).

Given the role of IRF1 as a transcription factor and as a core component of the type I and II IFN responses, we asked whether ISG expression is influenced by IRF1 during the exit from naive pluripotency. We first confirmed high expression levels of ISGs (as defined in Wu *et al*, 2018) in our ESC and EpiLC RNA-seq data (Fig EV3D). As expected, a subset of ISGs was expressed in both ESCs and EpiLCs; however, we also observed changes in the cell state-specific expression of ISGs (Fig EV3D). Notably, *Irf7*, which was shown to be incompatible with maintenance of pluripotency, was absent in either cell state (Eggenberger *et al*, 2019). Next, we analyzed which ISGs are differentially expressed in *Irf1* KO vs. WT in EpiLCs (Fig EV3E). Out of the subset of differentially expressed ISGs, *Irf1* KO most strikingly affects genes which under normal differentiation conditions are upregulated (Fig 3E). Examples for ISGs which are not upregulated in *Irf*1 KO include *Ddx58*, *Ddx60*, *Rsad2*, and *Trim56* (see also below).

We asked whether changes in gene expression are directly controlled by IRF1 and established an IRF1 overexpression system in ESCs using an inducible CRISPR-ON system (Heurtier *et al*, 2019). CRISPR sgRNAs are constitutively expressed. Upon doxycycline (dox) addition, dCAS9 coupled to 10× GCN4 is induced, which in turn recruits several molecules of scFv-linked VP64, a potent transcriptional activator. We used an sgRNA targeting the *Irf1* promoter and validated the expression of full-length IRF1 protein 48 h after dox induction in ESC (Fig EV3F). We performed QuantSeq to detect RNA expression changes after IRF1 overexpression (Dataset EV7). As a control, we included dox induction of dCAS9 in the absence of sgRNA expression. Activity of untargeted dCAS9 had a strong effect on gene expression, potentially because of DNA damage induction and off-target effects (Fig EV3G; Tycko *et al*, 2019). As expected, IRF1 expression was upregulated when sgRNA directed to the promoter of *Irf1* was used under dox induction conditions (Fig EV3H). The expression of NAGs was not influenced by IRF1 overexpression, but rather by non-targeted dCas9 presence (Fig EV3I, left panel). In contrast, ISG and specifically those ISG that were also misregulated in *Irf1* KO, showed upregulation in IRF1 overexpression (Fig EV3I, right panel, Fig 3F).

Pluripotency is associated with lack of epigenetic repression and transcriptional activation of transposable elements (TEs; Peaston *et al*, 2004). Transposable elements include endogenous retroviruses, and IRF1 is involved in repression of endogenous retroviruses (Stoltz *et al*, 2019). We analyzed expression levels of TE families in WT differentiation and in the absence of IRF1 in EpiLCs (Dataset EV6). ESCs and EpiLCs have distinct transcriptional profiles for TE families (Fig EV3J). In contrast, deletion of *Irf1* only has a minor effect on TE families (Fig EV3K). Therefore, we conclude that IRF1 is not a major regulator of TE expression in naive and formative pluripotency. However, *Irf1* is involved in the activation of a subset of ISGs.

The Oct6 enhancer is directly regulated by IRF1

Next, we asked which chromatin sites are bound by endogenous IRF1 in ESCs and EpiLCs. First, we validated IRF1 ChIP efficiency by qPCR. IRF1 ChIP recovers the promoter of *Gbp2* (Fig EV4A), a known IRF1 bound site in macrophages (Ramsauer *et al*, 2007; Langlais *et al*, 2016a). In addition, we also assayed two primer pairs at the *eOct6* enhancer locus, both showed enrichment in IRF1 Chip-qPCR (Fig EV4A).

To test genome-wide binding of IRF1, we sequenced libraries derived from IRF1 ChIP in WT and IRF1 KO for ESCs and EpiLC in replicates (Dataset EV8). Lack of peak calling by MACS2 for IRF1 KO samples confirmed specificity of the used antibody (Fig EV4B). Peaks in ESCs (which have low IRF1 levels) still differed from KO conditions (Fig EV4B). Qualitatively, binding sites of IRF1 did not drastically change in differentiation (71 sites with FDR < 0.05). Quantitatively, IRF1 bound sites are weakly marked already in ESCs, and binding signal increases upon differentiation in EpiLCs (Fig 4A).

We compared IRF1 binding sites to those of OCT4 as a hallmark of pluripotency-related sites (Buecker *et al*, 2014a; Data ref: Buecker *et al*, 2014b). 942/1770 of IRF1 sites overlapped with OCT4 found under either ESC, EpiLC or both conditions (Fig EV4C). The majority of the sites bound by both OCT4 and IRF1 are not changing their OCT4 binding in differentiation. More EpiLC-specific OCT4 sites overlap with IRF1 sites (261) than ESC-specific ones (19; Fig 4B). One example of such a EpiLC-specific site is the *eOct6* enhancer used as reporter in CRISPR-KO screen (Appendix Fig S2B).

Irf1 KO EpiLCs express lower levels of *Oct6* than control, and *Irf1* scored as a screenhit in the set-up which included Oct6 enhancer

activity as a readout. *Irf1* KO EpiLCs are shifted along the *eOct6* enhancer controlled mCherry expression axis and showed lower fluorescent values in FACS analysis (Appendix Fig S2A, C, and E). This effect was already present in ESCs, here also, the basal e*Oct6* enhancer activity was reduced. GFP levels controlled by the e*Tbx3* enhancer were not affected (Appendix Fig S2A, C, and E).

We tested whether *eOct6* enhancer activity can be rescued in *Irf1* KO background by IRF1 expression. We expressed doxycycline-inducible *Irf1* from cDNA in the dual reporter cell lines. Induction with doxycycline under ESCs conditions induced higher IRF1 expression than endogenous expression levels in EpiLCs (Appendix Fig S2D). The *eOct6* reporter was also more active in ESCs compared with WT and *Irf1* KO (Appendix Fig S2C). OCT6 protein levels were not increased in doxycycline-treated ESCs, indicating that even high enhancer activity cannot drive gene expression in a non-permissible chromatin environment (Appendix Fig S2D). Doxycycline treatment of reporter cells during differentiation also increased IRF1 levels and *eOct6* activity to higher levels compared with control differentiation. *eTbx3* always reported the differentiation status of the cells, independent of doxycycline treatment or *Irf1* background.

Doxycycline induced very strong expression of IRF1 and did not recapitulate normal expression levels. Therefore, we additionally used the *eOct6* enhancer to control IRF1 expression to enable differentiation-induced IRF1 expression during the exit from naive pluripotency. In cells transfected with this construct, IRF1 was already increased in ESCs at the protein level (Appendix Fig S2F), most likely due to low enhancer activity in ESCs. This weak expression corresponded to an already increased *eOct6* enhancer controlled mCherry signal in ESCs (Appendix Fig S2E). In EpiLCs, IRF1 expression levels were fully restored or slightly increased in comparison with endogenous levels (Appendix Fig S2F). This correlated with increased *eOct6* enhancer activity driving mCherry. Throughout all rescue conditions, *eTbx3* controlled GFP levels were not influenced and changed according to differentiation status (Appendix Fig S2C and E).

We identified a canonical IRF1 binding motif (gAAAgtGAAA) in the *eOct6* enhancer itself. To test whether this motif is necessary for enhancer activity, we mutated 11 nucleotides within the IRF1 motif (from here referred to as *eOct6-ΔIrf1*). We generated cell lines in WT and *Irf1* KO background which contain *eOct6-ΔIrf1* controlling mCherry expression. In addition, we included the wild-type *eOct6* enhancer controlling GFP expression as an internal control (Appendix Fig S2G). As expected, the unmodified *eOct6* enhancer was activated in WT cells and to a lesser extent than in *Irf1* KO cells. However, with the IRF1 binding motif mutated, the *eOct6* enhancer is activated independent of the presence or absence of IRF1, as *Irf1* KO shows similar activation levels compared with WT cells. The IRF1 binding motif mutated *eOct6-ΔIrf1* enhancer is still increasing activity in differentiation into EpiLCs, stressing that IRF1 is not the sole regulator of *eOct6* activity, but that a combination of factors is involved in the activation of the enhancer. In conclusion, IRF1 directly regulates *eOct6* enhancer activity dependent on the IRF1 binding motif, but IRF1 is not the sole regulator of this enhancer.

IRF1 binding sites are involved in viral defense

As IRF1 binding showed little overlap with cell state-specific OCT4 binding, we asked whether the IRF1 targets in EpiLCs are also activated in the innate immune response. Therefore, we compared ESC and EpiLC IRF1 binding sites to those in BMDMs, before and after IFNγ stimulation (Data ref: Langlais *et al*, 2016b; Fig 4C). 64% of ESC/EpiLC IRF1 binding sites are also bound by IRF1 in BMDMs, either stimulated or unstimulated with IFNγ. We compared genomic annotations between the sites shared between BMDMs and EpiLCs. Shared sites are located more often at promoter regions, while stem cell-specific sites are located in intergenic regions, including enhancer sites (Fig EV4D).

GO term enrichment across all stem cell IRF1 binding sites confirmed integration into biological processes such as defense response to other organisms and innate immunity processes (Fig 4D).

Excitingly, in EpiLCs, IRF1 directly binds to promoters of genes that code for core components of the interferon response: For example, *Ddx58/Rig-I* and its ligand-specific sentinel *Ddx60* (Oshiumi *et al*, 2015) showed increased IRF1 binding at the promoter in differentiation (Fig 4E). Furthermore, both genes are upregulated during the transition from ESC to EpiLCs (Fig 3E), and overexpression of IRF1 in ESCS leads to induction of these genes (Fig 3F), establishing both genes as direct targets of IRF1 in the exit from naive pluripotency. DDX58/RIG-I is known as a sensor of viral RNA, including for vesicular stomatitis virus (Kato *et al*, 2005; Yoneyama *et al*, 2005; Kell & Gale, 2015). Other examples include *Trim56* and *Rsad2/Viperin*, which are also bound and regulated by IRF1 and well-established components of innate immune response (Fig 4E).

Taken together, IRF1 directly regulates a subset of ISGs. Many of these targets are conserved between early embryonic and somatic cell types.

Intrinsic expression of IRF1 defends formative pluripotent cells against viral infections

ISGs play a major role in defense against viral infection, and IRF1 is a direct regulator of a subset of ISGs. Therefore, we investigated whether IRF1-mediated ISG expression in EpiLCs influences susceptibility to viral infection. We first differentiated WT and KO ESCs for 32 h to allow robust induction of IRF1 expression under WT

Figure 4. IRF1 binds to genes involved in the innate immune response.

A IRF1 ChIP-seq in ESC and EpiLC. Binding profiles and heatmaps of the consensus of all binding sites are shown.
B Venn diagram of overlap between chromatin IRF1 binding sites and OCT4 binding sites specific in ESCs (left) or EpiLCs (right). (Data ref: Buecker *et al*, 2014b).
C Venn diagram of overlap between chromatin IRF1 binding sites identified in bone marrow-derived macrophages (BMDM) −IFNγ/+IFNγ (Data ref: Langlais *et al*, 2016b) and ESC/EpiLC.
D GREAT-analysis of GO term enrichment for IRF1 chromatin binding sites in ESC/EpiLC. Binomial *P*-values of top 10 GO biological processes are shown.
E Chromatin loci with IRF1 binding profiles for genes selected in Fig 3F. Called IRF1 peaks at the promoter are highlighted by red boxes.

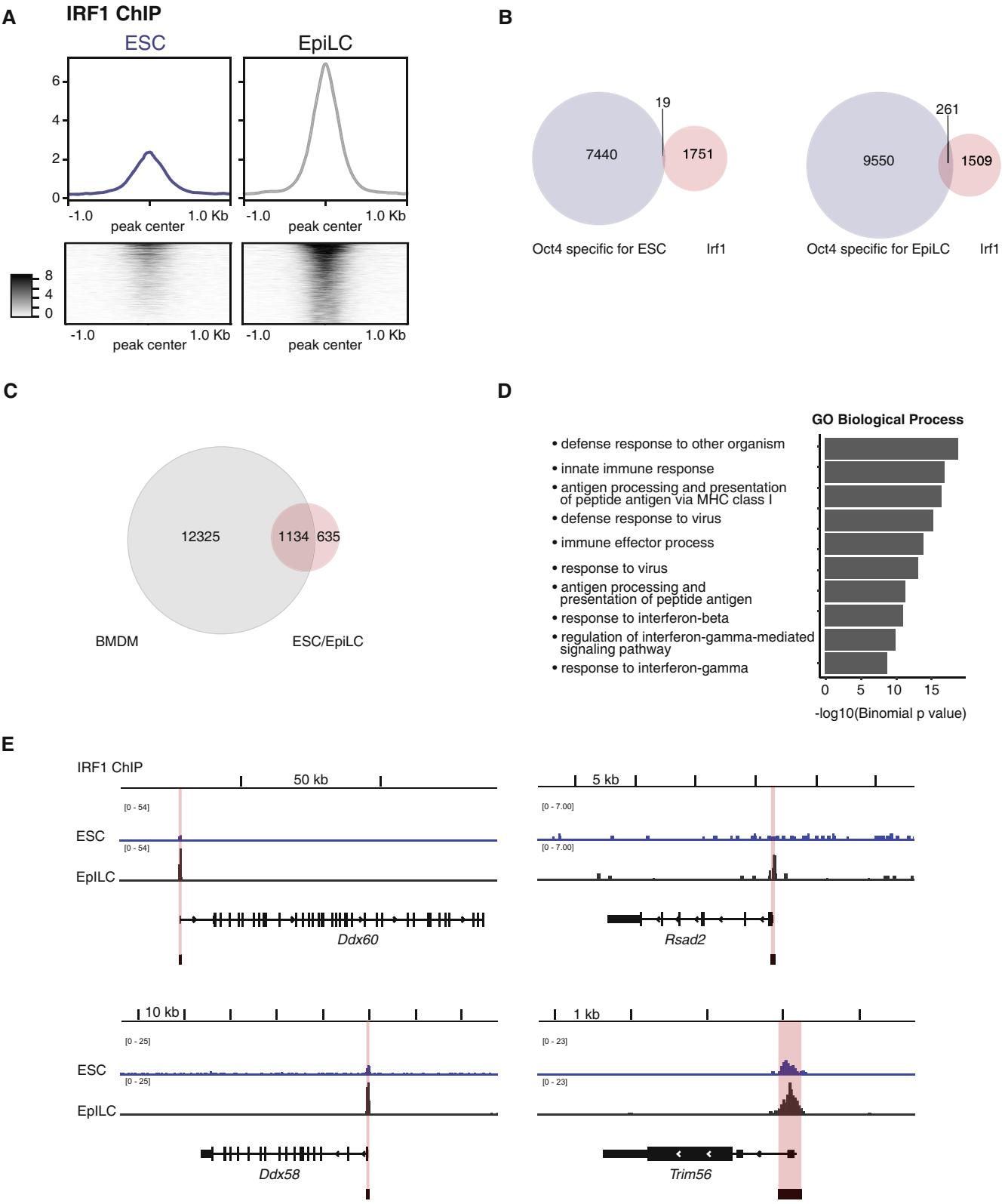


Figure 4.

conditions (Fig 5A). We then treated the differentiated cells for 16 h with vesicular stomatitis virus expressing GFP (VSV-GFP). VSV has broad tissue tropism, suggesting that viral entry is dependent on ubiquitous mechanisms (Finkelshtein *et al*, 2013). GFP expression allows direct readout of infection rates by FACS analysis of GFP-positive cells (Figs 5B, C and E, and EV5A, C and D). Cells were infected with a low multiple of infection (MOI) to establish a multi-cycle infection in *Irf1* WT cells. Strikingly, the percentage of infected cells significantly increased in *Irf1* KO (Fig 5B and C). Furthermore, deletion of the EpiLC-specific enhancer controlling *Irf1* expression also increased viral infection compared with WT cells (Fig 5B and C).

Next, we quantified the viral replication ability by determining the viral titers in the supernatant of infected cells (Fig 5A). Full KO of *Irf*1, but also reduced levels of IRF1 by enhancer KO, increased viral titers (Fig 5D). We tested whether reexpression of IRF1 from a doxycycline-inducible construct could reduce the infection levels (Fig EV5B and C). Indeed, IRF1 overexpression in differentiated *Irf1* KO significantly lowered infection rates (Fig 5E) and viral titers (Fig 5F). In addition, we challenged *Irf1* KO cells with and without IRF1 rescue with higher viral concentrations (Fig EV5D and E). Here, doxycycline treatment alone had an effect, but the reduction in the percentage of infected cells and viral titers was more pronounced in IRF1 overexpression conditions, arguing for a specific effect of exogenous IRF1 expression. The same effect has been described in somatic cells (Pine, 1992). We conclude that intrinsic upregulation of IRF1 is one of the defense mechanisms of EpiLCs against viral infections.

Discussion

In this study, we provide evidence that the pluripotency network ensures constitutive expression of the transcription factor IRF1. This expression in turn is required for intrinsic expression of a subset of ISGs, and the expression of these ISGs defends formative cells against viral infection.

We performed a pooled CRISPR KO screen to identify factors whose deletion impairs the activation of the formative pluripotency gene regulatory network. For this, we used the expression dynamics of a dual reporter system to identify cells which are showing expected behavior—are unimpaired in differentiation—or those who are impaired in differentiation. In contrast to other CRISPR screens that for example depend on proliferation of cells over prolonged time frames, our screening set-up is more sensitive to fluctuation due to noise. We counteracted these issues by focusing on a smaller library selected for nuclear factors, which could miss important factors such as metabolic regulators involved in differentiation (Moussaieff *et al*, 2015). Nevertheless, we identified several factors such as *Fgfr1* and *Zic3* which have already been studied in the context of the exit from naive pluripotency, stressing the validity of this approach (Molotkov *et al*, 2017; Yang *et al*, 2019).

We infected cells with a low MOI to ensure single knockout of one gene per cell. However, recent studies have shown the markable redundancy ensuring differentiation from naive to formative pluripotency: knockout of single factors delays, but not completely abolishes differentiation (Lackner *et al*, 2021). Differentiation is only abolished by simultaneous genetic deletion of three complementary drivers (Kalkan *et al*, 2019). Furthermore, the naive pluripotency network is supported by functionally overlapping factors of the same orphan nuclear receptor family (Festuccia *et al*, 2021). We therefore hypothesize that cooperativity and redundancy between two or even more factors are ensuring proper execution of this cell state transition, and that screening approaches should be adapted to this in the future.

We identified *Irf1* in our CRISPR KO screen as influencing the reporter activity of an enhancer region of *Oct6*, which we used as a proxy for the establishment of formative pluripotency. However, IRF1 deletion does not drastically change gene expression of other formative markers such as *Otx2*, *Fgf5*, and *Dnmt3a/b*, indicating that observed change in reporter activity is not strictly a readout of cell state. IRF1 is directly interacting with the chosen enhancer region through a canonical IRF1 chromatin binding motif. *Oct6* gene expression is reduced in the absence of IRF1. *Oct6* belongs to the family of POU TFs that share a highly similar DNA recognition motif. In fact, it was shown that OCT6 can replace OCT4 during reprogramming into human iPS cells due to the similarity in DNA motif binding (Kim *et al*, 2020). OCT6 and OCT4 occupy many of the same sites in epiblast stem cells (EpiSCs), and it is therefore not surprising that lower levels of OCT6 have only minor effects in the exit from naive pluripotency (Matsuda *et al*, 2017). Furthermore, deletion of *Oct6* in the mouse does not severely impact embryogenesis, again probably due to functional redundancy among different members of the POU transcription factor family (Bermingham *et al*, 1996; Kim *et al*, 2021).

Redundancy seems to be present not just in naive pluripotency, but also among many formative pluripotency regulators: The transcription factor OTX2 is sufficient to drive gene changes associated with formative pluripotency, but its deletion only has limited effect on the expression of formative genes (Yang *et al*, 2014; Buecker

Figure 5. IRF1 protects EpiLCs from viral infection.

A Experimental strategy to analyze viral infection in ESC to EpiLC transition. Cells were differentiated for 32 h and then infected with GFP-VSV. GFP^+ cells were scored by FACS and titers were determined by plaque formation assays.

B Representative FACS profiles of GFP-VSV infected $Irf1^{+/+}$, *Irf1* enhancer KO and $Irf1^{-/-}$ cells. See Fig 5C for quantification.

C Quantification of GFP^+ cells after GFP-VSV infection. *P*-values were calculated per Wilcoxon test. $n = 6$ biological replicates. The central band shows the median, the box 25^{th} and 75^{th} percentiles, whiskers show 1.5*IGR (interquartile range).

D Viral titer as determined by plaque formation assays. *P*-values were calculated per Wilcoxon test. $n = 6$ biological replicates. The central band shows the median, the box 25^{th} and 75^{th} percentiles, whiskers show 1.5*IGR (interquartile range).

E Quantification of GFP^+ cells after GFP-VSV infection, IRF1 doxycycline-inducible overexpression in $Irf1^{+/+}$ and $Irf1^{-/-}$ cells, without and with doxycycline induction. *P*-values were calculated per Wilcoxon test. $n = 3$ biological replicates, black horizontal lines show the mean of the data.

F Viral titer as determined by plaque formation assays, IRF1 doxycycline-inducible overexpression in $Irf1^{-/-}$ cells, without and with doxycycline induction. *P*-values were calculated per Wilcoxon test. $n = 3$ biological replicates, black horizontal lines show the mean of the data.

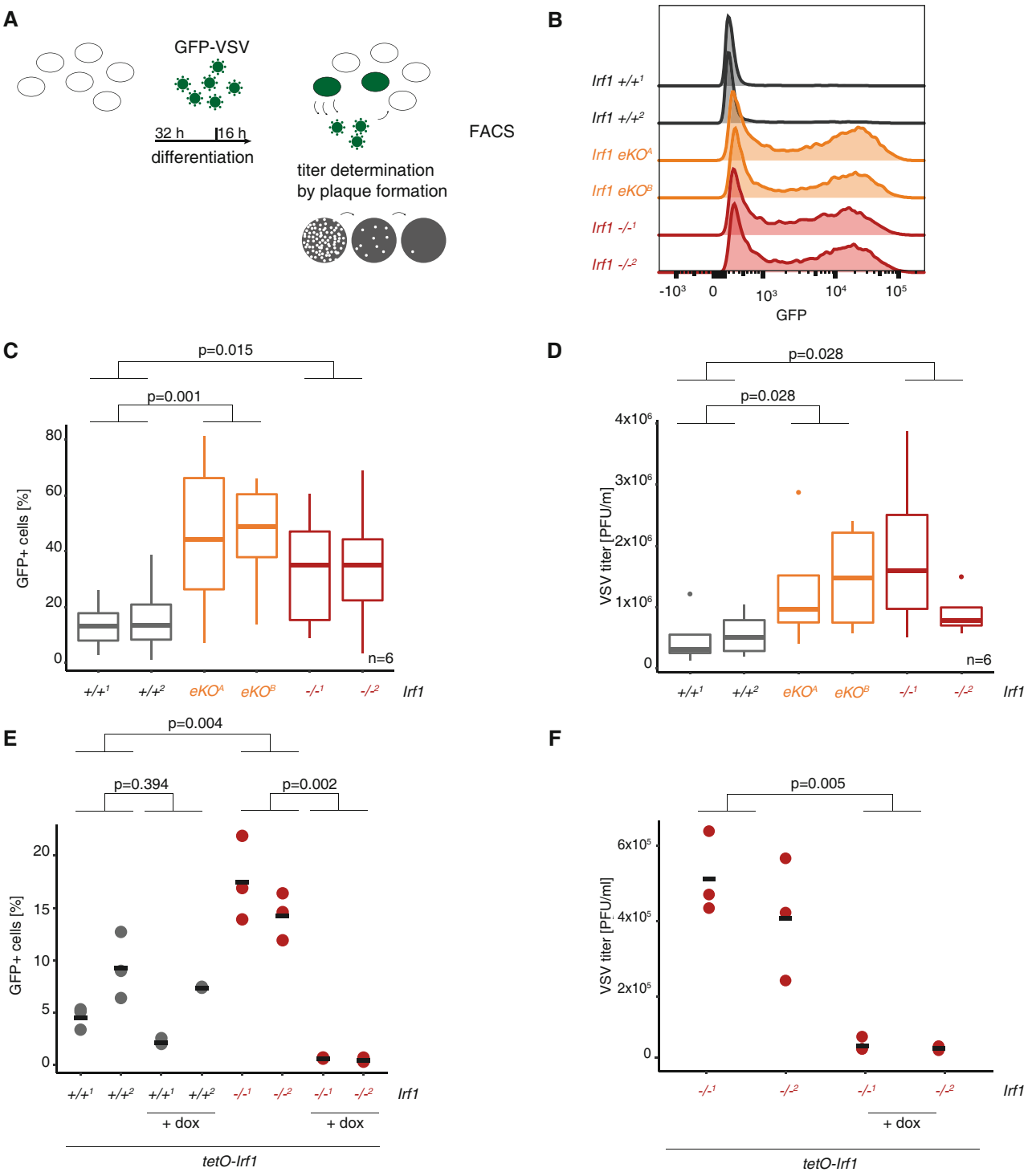


Figure 5.

et al, 2014a). Misregulation of OCT6 could therefore also be compensated by other members of the formative gene regulatory network. While OCT6 has been studied mostly in the context of development and neurogenesis, it might have additional roles upon interferon stimulation. OCT6 itself is upregulated by interferons in macrophages and other cell types; however, its role in interferon signaling is unclear (Hofmann *et al*, 2010).

Importantly, *Irf1* itself is directly regulated by the formative pluripotency network through an enhancer, which is marked by activating chromatin marks as well as OCT4 and OTX2 in formative EpiLCs. This enhancer is stem cell specific as it does not get activated in stimulated BMDMs (Lara-Astiaso *et al*, 2014). A subset of ISGs is differentially expressed in the exit from naïve pluripotency. IRF1 activates EpiLC-specific ISGs, but so far, it is unclear what drives ESC-specific ISG

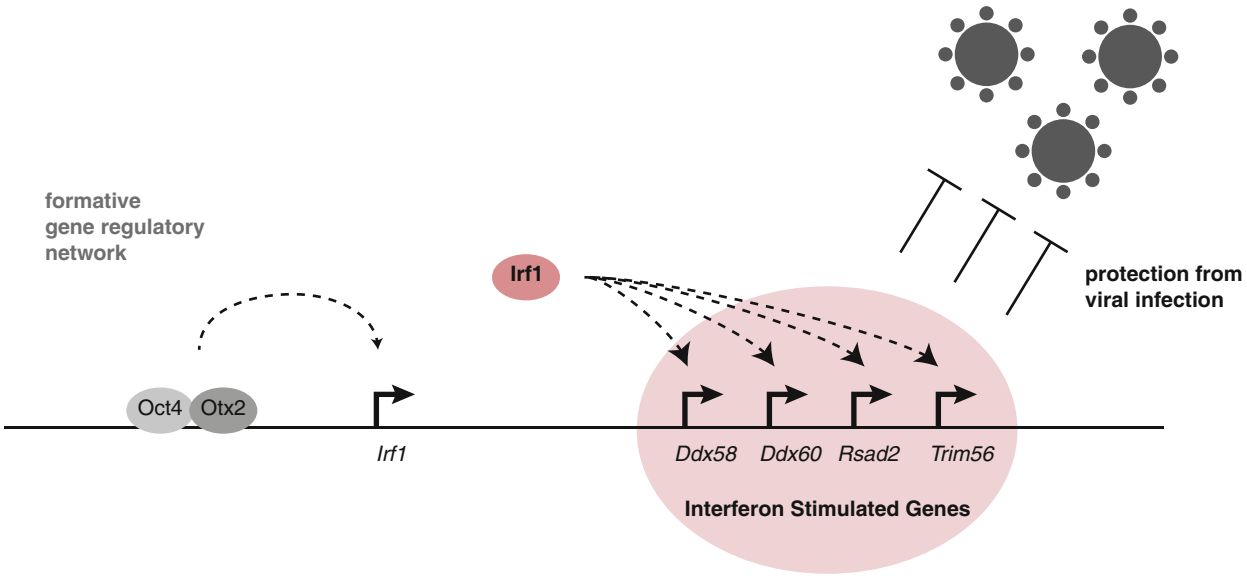


Figure 6. The formative pluripotency network regulates IRF1, which is required for defense against viral infections.

The formative pluripotency network directly activates Irf1 expression during the exit from naïve pluripotency through an EpiLC-specific enhancers. IRF1 in turn activates a subset of ISGs. This proactively sets the machinery for defense against viral infections in place. In the presence of IRF1, cells therefore are stronger protected against viral infections.

expression. ISG expression in early embryonic tissues, iPS or also somatic stem cells is conserved in mouse, pig, chimpanzee, and human (Wu *et al*, 2018; Shi *et al*, 2020). This expression is independent from viral stimulation and was previously only described as intrinsic (Wu *et al*, 2018) or as a consequence of endogenous retroviral expression observed in hypomethylated stem cells (Grow *et al*, 2015).

Wu *et al* (2018) showed that ISG expression in human stem cells can protect these cells from viral infection. Here, we show that similar mechanisms are at play in mouse stem cells, as IRF1 is a key factor required for the activation of several ISGs. Of note, the expression of ISGs is a physiological and constitutive property of differentiating cells, as we observed these relationships in the absence of viral infection. It has also been shown in hepatocytes that constitutive IRF1 expression maintains ISG expression and confers immediate defense against virus (Yamane *et al*, 2019). The strong protection against viral infection suggests that combined action of the ISGs regulated by IRF1 is required for viral defense. The naive/formative culture system recapitulates changes occuring at the time of implantation of the embryo into the uterus. We therefore speculate that prophylactic priming of the defense system could be crucial at a vulnerable moment in development, as implantation could be a risk for viral exposure. Interestingly, the upregulation of IRF1 in the embryo is transient and *Irf1* expression is already reduced right after implantation. We therefore speculate that *Irf1* expression might only be needed at this very short time period when the epiblast first encounters the mother's uterus.

We show here that the formative pluripotency network itself controls the expression of a subset of ISGs through upregulation of IRF1, and thereby protects this stage from viral infection (Fig 6). Robustness against viral infection appears to be an important property of proliferating stem cells and is controlled by the specific gene regulatory network that safeguards the stem cell state itself.

Materials and Methods

Reagents and Tools table

Reagent/Resource	Reference or Source	Identifier or Catalog Number
Experimental Models		
mESC R1	Buecker *et al* (2014a)	N/A
mESC E14Tg2a	Acampora *et al* (2013	N/A
Recombinant DNA		
pB-eOct6-mCherry-Puro	This study	N/A
pB-e/Oct6--GFP--Neo	Thomas *et al* (2021)	N/A
pB-eOct6ΔIrf1motif-mCherry-Puro	This study	N/A

Reagents and Tools table (continued)

Reagent/Resource	Reference or Source	Identifier or Catalog Number
pB-eOct6ΔIrf1motif_extended-mCherry-Puro	This study	N/A
pB-eTbx3-dGDP-Hygro	This study	N/A
pX330-U6-Chimeric_BB_CBh_hSpCas9	Addgene Plasmid	#42230
pX330-U6-Chimeric_noCas9	This study	N/A
pB-transposase	System Biosciences Cat	#PB210PA-1
PB-gRNA-Puro	Addgene	#121121
PB-TetON-dual-SunTag-Hygro	Addgene	#121119
PB-CA-rtTA Adv	Addgene	#20910
pB-Oct6-Irf1-Neo	This study	N/A
PB-tetO-Irf1-Neo	This study	N/A
PB-tetO-Irf1-Blasti	This study	N/A
Antibodies		
GAPDH (WB 1:10,000)	Santa Cruz	sc-365062
IRF1 (IF 1:200, WB 1:1,000)	CST	#8478
OTX2 (WB 1:1,000)	R&D Systems	AF1979
VINCULIN (WB 1:5,000)	Santa Cruz	sc-73614
OCT6/Pou3F1 (WB 1:1,000)	Sigma/Aldrich	MABN738
Phospho-STAT3 (Tyr705) (3E2) (WB 1:1,000)	CST	#9138
Oligonucleotides and sequence-based reagents		
CRISPR KO library	This study	Dataset EV1
PCR primers	This study	Dataset EV2
pPCR primers	This study	Dataset EV2
CRISPR guides (validations)	This study	Dataset EV2
Chemicals, enzymes and other reagents		
poly-L-ornithine hydrobromide	Sigma-Aldrich	P4638
laminin	Sigma-Aldrich	L2020
human Plasma Fibronectin Purified Protein	Sigma-Aldrich	FC010-10MG
AlbuMAX™ II Lipid-Rich Bovine Serum Albumin	Fisher Scientific	11021-029
MACS NeuroBrew-21 with Vitamin A (Miltenyi Biotec)	Miltenyi Biotec	130-093-566
MEM NEAA	Thermo Fisher Scientific	12084947
Penicillin-Streptomycin	Thermo Fisher Scientific	15070063
Sodium Pyruvate	Thermo Fisher Scientific	15070063
2-Mercaptoethanol	Fisher Scientific	11508916
CHIR-99021	Selleck Chemicals	S1036
PD0325901	Selleck Chemicals	S1263
hLIF	Provided by the VBCF Protein Technologies Facility	https://www.viennabiocenter.org/facilities/
Recombinant Human FGF-basic	peprotech-eubio	100-18B-100µG
KnockOutTM Serum Replacement	Fisher Scientific	Gibco™ 10828028
Trypsin-EDTA solution	Sigma-Aldrich	T3924-100ML
FSC	Sigma-Aldrich	F7524-500ML
ruxolitinib	Invivogen	INCB018424
Lipofectamine® 2000	Invitrogen	10696153
G415 Disulfate solution	Sigma-Aldrich	G8168-10ML

Reagents and Tools table (continued)

Reagent/Resource	Reference or Source	Identifier or Catalog Number
doxycyclin	Sigma-Aldrich	D9891-1G
hygromycin B	Sigma-Aldrich	10843555001
puromycin	Invivogen	ant-pr-1
HEPES-buffered saline, pH 7.0 (2× for transfection)	VWR	J62623.AK
SpeedBeads™	Sigma-Aldrich	GE45152105050250
pepGOLD TriFastTM reagent	VWR	130-2010
cOmpleteTM Protease Inhibitor Cocktail	Sigma-Aldrich	11836145001
DAPI	Sigma-Aldrich	D9542
RIPA lysis buffer	Merck	20-188
DMEM high glucose	Sigma-Aldrich	D6429
Avicel-RC591	FMC BioPolymer	N/A
Software		
bbmap	https://jgi.doe.gov/data-and-tools/software-tools/bbtools/	N/A
ChIPpeakAnno (3.24.2)	http://www.bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html	N/A
deeptools 3.5.1	https://deeptools.readthedocs.io/en/develop/	N/A
htseq/0.11.2-foss-2018b-python-3.6.6	https://htseq.readthedocs.io/en/master/	N/A
MAGeCK 0.5.9.	https://sourceforge.net/p/mageck/wiki/Home/	N/A
Nextflow/21.04.1 nf-core	https://nf-co.re	N/A
nf-core/chipseq v1.2.2	https://nf-co.re/chipseq/1.2.2/usage	N/A
star/2.7.1a-foss-2018b	https://github.com/alexdobin/STAR	N/A
tophat/2.1.2-foss-2018b	http://ccb.jhu.edu/software/tophat/index.shtml	N/A
R4.0.4	https://cran.r-project.org	N/A
ggplot2 (3.3.5)	https://cran.r-project.org/web/packages/ggplot2/index.html	N/A
DESeq2 (1.30.1)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html	N/A
pheatmap (1.0.12)	https://cran.r-project.org/web/packages/pheatmap/index.html	N/A
GREAT (4.0.4)	http://great.stanford.edu/public/html/	N/A
ImageJ	https://imagej.net	N/A
Other		
BD FACSMelody Cell Sorter	BD Biosciences	N/A
BD FACSAria	BD Biosciences	N/A
BD FACSFortessa	BD Biosciences	N/A
Zeiss Axio Oberserver Z1	Zeiss	N/A
Ilumina HiSeqV4 SR50	Illumina	N/A
Illumina NextSeq550 PE75	Illumina	N/A
Chemidoc Touch	Bio-Rad	N/A
Bioanalyser	Agilent	N/A
µ-Slide 8 Well Chambered Coverslip	ibidi	#80827
AMPure XP beads	Beckman Coulter	A63880
PVDF Transfer Membranes	Thermo Scientific	#88518
NucleoSpin PCR and Gel Purification Kit	Macherey-Nagel	#15559212
NEBNext Library Quant Kit for Illumina	NEB	E7630S
SensiFast™ cDNA Synthesis kit	Bioline	67.BIO-65054
SensiFASTTM SYBR® No-ROX kit	Bioline	67.BIO-98020
3' mRNA Seq Library Prep Kit	Lexogen	No. 081.96

Reagents and Tools table (continued)

Reagent/Resource	Reference or Source	Identifier or Catalog Number
sparQ DNA Library Prep Kit	Quanta Bio	#95191-096
Alkaline Phosphatase Detection Kit	Sigma-Aldrich	SCR004
Bio-Rad Protein Assay	BioRad	#5000006

Methods and Protocols

ESC maintenance and differentiation conditions

Murine embryonic stem cells (E14Tg2a for screening, R1 for all other cell lines) were cultured and differentiated as described in Thomas *et al* (2021). For maintenance, cells were grown on CELLSTAR® 6/12 wells coated with first poly-L-ornithine hydrobromide (6 μg/ml in PBS, 1 h at 37°C, sigma P4638) and then laminin (1.2 μg/ml in PBS, 1 h at 37°C, Sigma L2020). Cultures were tested regularly for mycoplasma contamination. For differentiation and viral infection, plates were coated with fibronectin (Human Plasma Fibronectin Purified Protein, Sima-Aldrich, 5 μg/ml in PBS, 1 h at RT).

Cells were cultured in base medium HyClone DMEM/F12 without Hepes (Cytiva) with 4 mg/ml AlbuMAX™ II Lipid-Rich Bovine Serum Albumin (GIBCO™), 1× MACS NeuroBrew-21 with Vitamin A (Miltenyi Biotec), 1× MEM NEAA (GIBCO™), 50 U/ml Penicillin–Streptomycin (GIBCO™), 1 mM Sodium Pyruvate (GIBCO™), and 1× 2-Mercaptoethanol (GIBCO™). For 2i + LIF culturing conditions, base medium was supplemented with with 3.3 mM CHIR-99021 (Selleckchem), 0.8 mM PD0325901 (Selleckchem), and 10 ng/ml hLIF (provided by the VBCF Protein Technologies Facility, https://www.viennabiocenter.org/facilities/). For differentiation, base medium was supplemented with 12 mg/ml Recombinant Human FGF-basic (PEPROTECH) and KnockOutTM Serum Replacement (1:100, GIBCO™).

For splitting, cells were treated with 1× Trypsin–EDTA solution (sigma T3924) at 37°C until cells detached. Trypsinization was stopped with 2i + LIF medium with 10% FSC (Sigma F7524), cells were harvested by centrifugation at 300 *g* for 3 min, resuspended in 2i + LIF and seeded in appropriate ratios.

For differentiation, cells were seeded a day prior on fibronectin-coated plates in 2i + LIF medium (100 k per 12 well, 200 k per 6 well). The next day, differentiation was started by removing the medium, two washes of the attached cells with PBS, and the addition of differentiation medium. Cells were collected as EpiLCs after 48 h.

For JAK inhibition experiments, cells were treated as indicated on the onset of differentiation with 1.5 μM ruxolitinib (Invivogen).

Generation of reporter and screening cell lines

Cells used for the dual fluorescent reporter system and CRISPR screening were based on an $Otx2^{flox/-}$; $R26^{CreER/+}$ ESC cell line (Acampora *et al*, 2013). Otx2 heterozygous expression does not impair ESC to EpiLC differentiation (Acampora *et al*, 2013). pB-transposase, pB-eOct6-mCherry-Puro and pB-eTbx3-dGFP-Hygro were lipofectamine transfected. After selection, two clonal cell lines with clear distinction of ESC/EpiLC states in FACS analysis were selected. Lentivirus containing Cas9-Blasticidin was used for constitutive CAS9 expression. Neomycin resistance derived from the original present LacZ allele was removed by CRISPR KO.

For pB-eOct6ΔIrf1-mCherry-Puro, the mutated Irf1 binding sites were placed on overlapping primers, and the Oct6 enhancer was PCR amplified and the plasmid constructed with three fragment Gibson Assembly. pB-transposase, pB-eOct6ΔIrf1-mCherry-Puro and pB-eOct6-mCherry-Neo were co-transfected into ESCs with lipofectamine.

Generation of KO and rescue cell lines

Knockouts of coding sequences or enhancer regions were performed with dual CRISPR sgRNA KO. For this, CRISPR sgRNAs were inserted into pX330-U6-Chimeric_noCas9 (for cell lines already expressing Cas9) or pX330-U6-Chimeric_BB_CBh_hSpCas9 using BbsI (NEB) directed cloning. sgRNA plasmid combinations and substoichiometric amounts of dsRed plasmid were cotransfected by Lipofectamine® 2000 (Invitrogen) transfection (375 ng each sgRNA plasmid, 50 ng dsRed, 5 μl lipofectamine in total 100 μl DMEM/F12) into cells seeded on 12-well plates 1 day prior. Two to three days after transfection, $dsRed^{+}$ cells as proxy for sgRNA transfection were FACS-sorted (BD FACSMelody Cell Sorter) as single cells on 96-well plates coated with fibronectin. Successful genome editing was confirmed by genotyping PCRs and Western blotting where applicable.

For doxycycline-inducible reexpression of Irf1 in Irf1 KO cell lines, we cloned Irf1 cDNA under tetO control with either Neomycin or Blasticidin selection cassettes. Transfected and selected cells were analyzed as pools without clonal selection. Dox treatment (1 μg/ml) was performed for 48 h, in ESC or EpiLC medium as indicated.

For differentiation-induced rescue of Irf1 expression, we cloned pB expression constructs containing Irf1 cDNA under control of the Oct6 enhancer. This resulted in low Irf1 expression levels in ESCs and upregulation of Irf1 upon differentiation. Transfection of pB-Oct6enh-Irf1-Neo was performed with lipofectamine transfection as described for reporter constructs, albeit with lower DNA amounts (250 ng pB plasmid). After neomycin selection (400 μg/ml, G415 Disulfate solution, sigma G8168-10ML), pools of transfected cells were differentiated and analyzed by FACS and Western blotting.

IRF1 overexpression with SunTag system

Overexpression of Irf1 from the endogenous locus was based on a SunTag overexpression system published in Heurtier *et al*, 2019. Following plasmids were ordered from Addgene: 121121 PB-gRNA-Puro, 121119 PB-TetON-dual-SunTag-Hygro, 20910 PB-CA-rtTA Adv. Ten microgram PB Transposase, 500 ng PB-TetON-dual-SunTag-Hygro and 500 ng PB-CA-rtTA were transfected into R1 ESCs by electroporation (500 μF, 240 V, 4 mm). Cells were treated with doxycycline (1 μg/ml) and hygromycin (400 μg/ml). After 5 days, BFP^{+}/GFP^{+} double positive cells were selected by FACS. Cells were grown in the absence of doxycycline and hygromycin and BFP^{-}/GFP^{-} double negative cells were selected by FACS. Single

clones were selected and tested by FACS and Western blotting against Cas9 to select for response to doxycycline treatment.

sgRNAs targeting the Irf1 promoter were designed in benchling, which is based on (Hsu *et al*, 2013) and cloned into PB-gRNA-Puro by BbsI directed cloning. The sgRNA plasmid and PB transposase were lipofectamin transfected into parental SunTag cell lines and selected by puromycin (2 μg/ml, Invivogen) treatment. Irf1 upregulation after dox treatment was confirmed by qPCR and Western blotting.

CRISPR screening

CRISPR screening was based on (Michlits *et al*, 2017), but without using clonal dilution steps. DNA pools of indicated subpools were combined according to sgRNA number and CaCl Hepes transfected into PlatE cells, including helper plasmid. Twenty-four hours after transfection, medium was exchanged to 2i + LIF. Virus containing 2i + LIF was harvested in two batches over the next 36 h. Both batches were pooled and frozen. MOI was determined by determining infection efficiency without selection.

Execution of CRISPR screen

The screen was performed with two independent replicates with different screening cell lines. To reach sufficient coverage of the library, we infected 100 million cells at a MOI of 0.3 by seeding 7.5 million cells per 15 cm dish coated with fibronectin (Dataset EV1). Starting library representation was accessed by infecting cells without Cas9 (R1) in parallel. In addition, no selection, kill and no sgRNA library control plates were prepared. Usage of fibronectin coated plates allows to omit polybrene addition during infection. As soon as cells were attached (2 h), virus was added to the cells without additives or medium exchanges (day1). The next day, selection with neomycin was started and continued for 2 days. This short but stringent selection (400 μg/ml, G415 Disulfate solution, sigma G8168-10ML) resulted in complete elimination of nonresistant cells. From now on, each day cells were either splitted or medium was exchanged. On day 4, the kill control was dead and 30 mio library infected cells were seeded on 15 cm PLOL plates with 7.5 mio per plate without selection. Empty library controls were carried throughout the screen. On day 6 and 8, again 30 mio cells were seeded without selection, allowing time for depletion of essential genes. On day 10, 60 mio cells were harvested and analyzed later as samples representing genes essential for ESC maintenance. For differentiation, 30 mio cells were seeded on fibronectin-coated dishes (7.5 mio per 15 cm dish) In addition, extra dishes as 2i + LIF control and no sgRNA library controls were seeded. On day 11, differentiation was induced by 2 + iLIF medium removal, 2 washed in PBS and addition of differentiation medium. Starting time points for different plates were staggered to allow FACS analysis after 48 of differentiation for each plate.

On day 13, cells were harvested and sorted by BD FACSAria (BD Biosciences). For this, gates were set up using no sgRNA library control cells, which were processed in parallel. At least 60 mio cells were sorted in the unimpaired gate and all possible cells in the impaired gate (N1 = 64.7–10.3 mio, N2 = 71.5–5.6 mio cells). In addition, 60 mio cells for essential ESC and 60 mio for unsorted EpiLCs were harvested for each condition. The screen was performed as replicate using two independently generated cell lines.

gDNA isolation, PCR amplification, and NGS

CRISPR KO libraries were generated from collected cells as described by (Michlits *et al*, 2017), scaled according to the cell numbers used.

In brief, cell pellets were resuspended in SDS-lysis buffer and incubated at 55°C overnight to lyse cells, followed by RNaseA treatment. Protein was precipitated by NaCl precipitation, the supernatant was then purified with phenol/chloroform. DNA was precipitated with isopropanol and dissolved in TE buffer. DNA was PacI digested for a total of 48 h and size selected with SpeedBeads™ (Millipore Sigma) in two rounds to enrich fragments < 2 kb.

CRISPR-UMI cassettes were PCR amplified and multiplexed with 25 cycles of 1:1 KlenTag/Phusion PCR (Jena Bioscience and in house generated) using primers specified in Michlits *et al* (2017). Reactions were purified using the NucleoSpin PCR and Gel Purification Kit (Macherey-Nagel). Samples were pooled after qPCR quantification, purified from 1.5% agarose, and quantified with NEBNext Library Quant Kit for Illumina (NEB). Sequencing was performed on Illumina HiSeqV4 SR50 (Dataset EV3, Dataset EV4).

RT-qPCR analysis

ESC and EpiLC were directly lysed on the plate using pepGOLD TriFastTM reagent (Peqlab) according to the manufacturer's instructions. Five hundred nanogram of RNA were reversed transcribed to cDNA with the SensiFast™ cDNA Synthesis kit (Bioline) according to the manufacturer's instructions.

For qPCR, cDNA was 1:5 diluted and 0.5 μl were used per 10 μl qPCR reaction with SensiFASTTM SYBR® No-ROX kit (Bioline) and 125 nM forward and reverse primer. Irf1-specific primers were based on (Platanitis *et al*, 2019).

qPCRs were analyzed by $\Delta\Delta C_T$ and normalized to ESCs samples of the corresponding cell lines. Statistical tests were performed as homoscedastic one-sided *t*-tests.

FACS analysis

ESC and EpiLC were harvested by trypsin treatment, which was stopped by 1:1 addition of 10% FSC. Samples were strained through 5-ml polystyrene round-bottom tubes with cell-strainer caps (Falcon). For analytical purposes, the BD Fortessa was used, for cell sorting in high-throughput BD FACSAria or in low-throughput BD FACSMelody. Single cells were gated according to FFW and SSW scattering.

Quantseq

RNA QuantSeq was prepared according to the manufacturer's instructions (Lexogen 3' mRNA Seq Library Prep Kit). Five hundred nanogram RNA was used as starting material. qPCRs were used to multiplex samples and final libraries were quantified by NEBNext Library Quant Kit for Ilumina (#E7630S).

Chromatin immunoprecipitation

ChIP with IRF1 was performed as described in Thomas *et al* (2021). In brief, 3 mio cells were seeded on 15 cm fibronectin-coated dishes, medium exchanged to 2i + LIF or differentiation the next day and cells harvested after 48 h.

For harvesting, cells were cross-linked in 1% formaldehyde in PBS for 10 min. Cross-linking was quenched in 0.125 M glycine. From now on, samples were kept at 4°C. Fixed cells were washed

on the plate twice with PBS and scraped off in 0.01% Triton in PBS. Cells were harvested by 500 *g* 5 min centrifugation and flash-frozen in liquid nitrogen. Cell pellets were resuspended in 5 ml LB1 (50 mM Hepes pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% TX-100, 1 mM PMSF, 1× cOmpleteTM Protease Inhibitor Cocktail (Roche) by rotating vertically for 10 min at 4°C and harvested (5 min, 1,350 *g*, 4°C)). The pellet was resuspended in 5 ml LB2 (10 mM Tris pH 8. 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA1, mM PMSF, 1× cOmpleteTM Protease Inhibitor Cocktail (Roche)), 10 min vertical rotation, room temperature. Centrifugation was repeated and pellets were suspended in 1.5 ml LB3 (10 mM Tris–HCl pH 8, 100 mM NaCl, 1 mM EDTA 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine, 1 mM PMSF, 1× cOmpleteTM Protease Inhibitor Cocktail (Roche)) and 200 µl sonification beads (diagenode) in Bioruptor® Pico Tubes (diagenode). Sonification was performed for 13 cycles 30 s on/45 s off. Supernatant, but not the beads, were transferred to fresh tubes and cellular debris was pelleted at 16,000 *g* at 4°C. 1.1 ml were transferred to fresh tubes and 110 µl 10% triton to final 1% were added. Fifty microliter was saved as input and 1 ml was used for ChIP.

Antibody precipitation was performed with 5 µl IRF1 antibody overnight at 4°C with vertical rotation. Next day, Dynabeads protein G for Immunoprecipitation (Thermo Fisher Scientific) were used. One hundred microliter beads were washed in block solution (0.5% BSA in PBS) and incubated with chromatin/antibody solutions for at least 2 h. Bound beads were washed five times in cold RIPA wash buffer (50 mM Hepes pH 7.5, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate), followed by three washes in TE + 50 mM NaCl. Bound fractions were eluted in 210 µl elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) for 15 min at 65°C. Supernatant was removed from the beads. Input samples were diluted with three volumes of elution buffer. Input and ChIP samples were decrosslinked at 65°C overnight.

One volume of TE was added to all samples as well as RNase A (0.2 mg/ml final) and incubated at 37°C for 2 h. Salt concentration was adjusted to final 5.25 $CaCl_2$ with 300 mM $CaCl_2$ in 10 mM Tris pH 8.0 and Proteinase K added to final 0.2 mg/ml. Digestion was performed at 55°C for 30 min. DNA was phenol-chloroform extracted in Phase Lock GelTM tubes (Quantabio), ethanol precipitated and dissolved in H20.

ChIP was checked by performing qPCRs and calculating recovery of input (%).

Sequencing libraries were generated with sparQ DNA Library Prep Kit (Quanta Bio #95191–096) according to the manufacturer's instruction and using AMPure XP beads. Adapter contamination was cleaned up with an additional round of AMPure XP bead purification. Quality of the libraries were checked by Bioanalyser (Agilent) and final concentrations were determined with the NEB library Quant Kit.

Alkaline phosphatase staining

For AP staining, cells were seeded in different densities on gelatine-coated plates. The next day, differentiation was performed for 72 h, controls were kept in 2i + LIF medium. All cells were placed back in 2i + LIF medium. After 2–4 day, colonies were stained with the Alkaline Phosphatase Detection Kit (Sigma-Aldrich) according to the manufacturer's instructions.

Immunofluorescence

Ten k cells were seeded per chamber of a µ-Slide 8 Well Chambered Coverslip (ibidi). Cells were differentiated for 48 h, washed in PBS and fixed with 4% PFA for 15 min at RT. Cells were washed 3× in PBST (0.1% Tween in PBS). Permeabilization was performed in 0.1% Triton-X in PBS for 10 min at RT. Cells were washed 3× in PBST and blocked in 5% BSA in PBST for 30 min at 4°C. Primary antibody was diluted in blocking buffer and incubated overnight at 4°C. Cells were washed 5× in PBST and incubated with secondary antibody in blocking buffer for 1 h at RT. Cells were washed 3× in PBST, followed by 2× PBS washes. Nuclei were stained with 20 ng/ml DAPI (Sigma, D9542) for 10 min. Cells were washed 3× in PBS and stored in PBS at 4°C until imaging. Imaging was performed on a Zeiss Axio Oberserver Z1 microscope with 63× oil objective. Images were processed using ImageJ.

Western blot analysis

For Western blot analysis, cells were harvested by trypsin treatment. The cell pellet was washed once in PBS, after supernatant removal dry pellets were frozen. Cells were lysed in 1× RIPA lysis buffer (Merck, 20-188), including 1× cOmpleteTM Protease Inhibitor Cocktail (Roche) and phosphatase inhibitors (1 mM NaF, 20 mM β-glycerophosphate, 1 mM Na_3VO_4) for 1 h on ice. Cellular debris was removed by centrifugation (16,000 *g*, 10 min, 4°C). Protein was quantified using Bio-Rad Protein Assay (5000006). Twenty-five microgram protein per sample were separated on 10% Tris-gylcine SDS polyacrylamid electrophoresis. Wet-blotting was performed onto PVDF Transfer Membranes (Thermo Scientific #88518). Membranes were blocked in 5% milk in PBST. Primary antibodies were incubated overnight at 4°C, secondary for 2 h at RT. HRP coupled secondary antibodies were used for detecting with GE Healthcare LS ECL Select WB detection reagent. Images were acquired on a Chemidoc Touch (Bio-Rad).

Note that αIRF1 primary detected two bands, both which were specific to IRF1. The ratio of these bands is dependent on the presence or absence of phosphatase inhibitor (Roche PhosSTOP) in lysis conditions.

Viral infection

2.5×10^4 cells were seeded a day before the start of differentiation on 24-well plates in 2i + LIF medium. Differentiation was induced as described before, as indicated cells were doxycycline treated at the same time. After 32 h, virus was added to the cells. For this, a VSV-GFP stock with a titer of 1×10^8 PFU/ml (determined on A549 cells) was diluted 1:20,000 or 1:5,000 in differentiation medium. Two hundred microliter of virus dilutions were added to the existing 500 µl of medium on the cells, resulting in estimated MOIs of 0.02 and 0.08 PFU/ml, respectively. Cells were incubated for an additional 16 h.Subsequently, supernatants were collected, cleaned from any cells by centrifugation and stored at −80°C until titration by plaque assays. EpiLCs were collected by trypsinization and the cell pellets were fixed in 4% PFA (Electron Microscope Sciences #1570) in PBS for 15 min. One milliliter of PBS was added to dilute the PFA, cells were collected by centrifugation and resuspended in 10% FCS in DMEM/F12 for FACS analysis.

Plaque assay

Vero cells were cultured in DMEM high glucose (Sigma-Aldrich D6429) with Penicillin–Streptomycin (Thermo-Fischer 15070063)

and 10% FCS (Sigma-Aldrich, F7524). They were seeded at 50% confluency on day 1 in 6-well plates, so that they were > 90% confluent on day 2. On day 2, sequential dilutions of viral supernatants were made in DMEM without additives, ranging from 10^{-3} to 10^{-6}. Vero cells were incubated with 400 µl of viral dilutions for 1 h at room temperature in technical replicates. Afterward, viral dilutions were removed and cells were overlaid with 2 ml of 0.4%, in DMEM high glucose with 4.2% FCS to stop cell growth (Matrosovich *et al*, 2006). Cells were incubated at 37°C, 5% CO_2 for 24–30 h. Then, cells were fixed in 2% formaldehyde (Sigma-Aldrich F8775) in PBS for > 30 min. Plates were washed in tap water and stained with Crystal Violet Staining Solution (0.005% Crystal Violet, 1% Formaldehyde, 1X PBS, 1% methanol). Plaques in suitable dilutions were manually counted, blinded toward sample identity. Titers were averaged from the technical replicates.

Data analysis

For CRISPR KO screening, reads were aligned to the library using custom scripts as published in Michlits *et al* (2017). Starting from count tables, analysis was performed with MAGeCK 0.5.9.2 and R 4.0.4. Plots were generated with ggplot2 (3.3.5).

QuantSeq RNA-seq data was processed following Lexogen's standard pipeline on bluebee. Adapter contamination was removed with bbmap, mapped using STAR and counted with HTScount. Downstream analysis was performed in R using DESeq2 (1.30.1) and pheatmap (1.0.12).

QuantSeq RNA data were used to also analyze repeat elements as in Percharde *et al* (2017). For this, Tophat2 with the setting g-1 was used to retain multimappers mapped to one random location, repeat elements were counted using a custom gtf file. Downstream analysis was performed in R using DESeq2 (1.30.1) and pheatmap (1.0.12).

For ChIP-Seq analysis, the Nextflow/21.04.1 nf-core/chipseq v1.2.2 pipeline was used with mm10 as reference genome. Profiles and heatmaps were generated with deeptools 3.5.1. ChIP data were analyzed with ChIPpeakAnno (3.24.2) and GO term were determined with GREAT (4.0.4) analysis (McLean *et al*, 2010).

ISGs were defined based on (Wu *et al*, 2018).

Data availability

The NGS datasets produced in this study are available in the European Nucleotide Archive, https://www.ebi.ac.uk/ena/browser/home and assigned the identifier PRJEB53216 (https://www.ebi.ac.uk/ena/browser/view/PRJEB53216?show=reads).

Processed data have been included as extended view Datasets in this manuscript.

Expanded View for this article is available online.

Acknowledgements

We would like to thank all members of the Buecker laboratory for discussions and feedback throughout the project, Martin Leeb and Thomas Decker for critical feedback and discussion on the manuscript, Michelle Perchard for help with analysis of retrovirus and transposable elements and the BioOptics-FACS and BioOptics-Light Microscopy facility at Max Perutz Labs. ChIP-seq and QuantSeq was performed by the Next-Generation Sequencing Facility at Vienna BioCenter Core Facilities (VBCF), member of the Vienna BioCenter (VBC), Austria. This work was supported by the Austrian Science Fund FWF (P30599 and P34123 to CB and W1261 DK SMICH) and Uni:Docs fellowships from the University of Vienna to MR.

Author contributions

Merrit Romeike: Conceptualization; formal analysis; investigation; writing—original draft; writing—review and editing. **Stephanie Spach:** Data curation; investigation. **Marie Huber:** Data curation; investigation. **Songjie Feng:** Data curation; investigation. **Gintautas Vainorius:** Resources; methodology. **Ulrich Elling:** Resources; methodology. **Gjis A Versteeg:** Investigation; methodology. **Christa Buecker:** Conceptualization; supervision; funding acquisition; writing—original draft; project administration; writing—review and editing.

In addition to the CRediT author contributions listed above, the contributions in detail are:

CB and MR conceived and designed the study. MR, SS, and MH generated and validated cell lines and carried out experiments. SF generated and validated enhancer KO cell lines. MR carried out the screen. MR and GV analyzed screening data, with input from UE. GAV provided VSV-GFP and expertise in viral infections and titer determination. MR analyzed all sequencing data. MR and CB wrote the manuscript, with input from all co-authors. CB supervised the project.

Disclosure and competing interests statement

The authors declare that they have no conflict of interest.

References

Acampora D, Di Giovannantonio LG, Simeone A (2013) Otx2 is an intrinsic determinant of the embryonic stem cell state and is required for transition to a stable epiblast stem cell condition. *Development* 140: 43 – 55

Bermingham JR, Scherer SS, O'Connell S, Arroyo E, Kalla KA, Powell FL, Rosenfeld MG (1996) Tst-1/oct-6/SCIP regulates a unique step in peripheral myelination and is required for normal respiration. *Genes Dev* 10: 1751 – 1762

Betschinger J, Nichols J, Dietmann S, Corrin PD, Paddison PJ, Smith A (2013) Exit from pluripotency is gated by intracellular redistribution of the BHLH transcription factor Tfe3. *Cell* 153: 335 – 347

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, Bertone P (2015a) Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. *Dev Cell* 35: 366 – 382

Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, Nichols J, Smith A, Bertone P (2015b) Arrax Express E-MTAB-2958 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2958/). [DATASET]

Buecker C, Srinivasan R, Zhixiang W, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J (2014a) Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell* 14: 838 – 853

Buecker C, Srinivasan R, Wu Z, Calo E, Acampora D, Faial T, Simeone A, Tan M, Swigut T, Wysocka J (2014b) Gene Expression Omnibus GSE56138 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56138). [DATASET]

Burke DC, Graham CF, Lehman JM (1978) Appearance of interferon inducibility and sensitivity during differentiation of murine teratocarcinoma cells *in vitro*. *Cell* 13: 243 – 248

Chen B, Gurung C, Guo J, Kwon C, Guo Y-L (2020) Pluripotent stem cells are insensitive to the cytotoxicity of TNFα and IFNγ. *Reproduction* 160: 547 – 560

Dunn S-J, Martello G, Yordanov B, Emmott S, Smith AG (2014) Defining an essential transcription factor program for naive pluripotency. *Science* 344: 1156 – 1160

Eggenberger J, Blanco-Melo D, Panis M, Brennand KJ, ten Oever BR (2019) Type I interferon response impairs differentiation potential of pluripotent stem cells. *Proc Natl Acad Sci USA* 116: 1384 – 1393

Festuccia N, Owens N, Chervova A, Dubois A, Navarro P (2021) The combined action of Esrrb and Nr5a2 is essential for murine Naïve pluripotency. *Development* 148: dev199604

Finkelshtein D, Werman A, Novick D, Barak S, Rubinstein M (2013) LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc Natl Acad Sci USA* 110: 7306 – 7311

Gao X, Tate P, Hu P, Tjian R, Skarnes WC, Wang Z (2008) ES cell pluripotency and germ-layer formation require the SWI/SNF chromatin remodeling component BAF250a. *Proc Natl Acad Sci USA* 105: 6656 – 6661

Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY *et al* (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522: 221 – 225

Hackett JA, Huang Y, Günesdogan U, Gretarsson KA, Kobayashi T, Azim Surani M (2018) Tracing the transitions from pluripotency to germ cell fate with CRISPR screening. *Nat Commun* 9: 4292

Harada H, Willison K, Sakakibara J, Miyamoto M, Fujita T, Taniguchi T (1990) Absence of the type I IFN system in EC cells: Transcriptional activator (IRF-1) and repressor (IRF-2) genes are developmentally regulated. *Cell* 63: 303 – 312

Hart T, Tong AHY, Chan K, Van Leeuwen J, Seetharaman A, Aregger M, Chandrashekhar M, Hustedt N, Seth S, Noonan A *et al* (2017) Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* 7: 2719 – 2727

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M (2011) Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146: 519 – 532

Heurtier V, Owens N, Gonzalez I, Mueller F, Proux C, Mornico D, Clerc P, Dubois A, Navarro P (2019) The molecular logic of Nanog-induced self-renewal in mouse embryonic stem cells. *Nat Commun* 10: 1109

Hofmann E, Reichart U, Gausterer C, Guelly C, Meijer D, Müller M, Strobl B (2010) Octamer-binding factor 6 (oct-6/Pou3f1) is induced by interferon and contributes to DsRNA-mediated transcriptional responses. *BMC Cell Biol* 11: 61

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O *et al* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31: 827 – 832

Kalkan T, Bornelöv S, Mulas C, Diamanti E, Lohoff T, Ralser M, Middelkamp S, Lombard P, Nichols J, Smith A (2019) Complementary activity of ETV5, RBPJ, and TCF3 drives formative transition from naive pluripotency. *Cell Stem Cell* 24: 785 – 801

Kato H, Sato S, Yoneyama M, Yamamoto M, Uematsu S, Matsui K, Tsujimura T, Takeda K, Fujita T, Takeuchi O *et al* (2005) Cell type-specific involvement of RIG-I in antiviral response. *Immunity* 23: 19 – 28

Kell AM, Gale M (2015) RIG-I in RNA virus recognition. *Virology* 479–480: 110 – 121

Kim K-P, Han DW, Kim J, Schöler HR (2021) Biological importance of OCT transcription factors in reprogramming and development. *Exp Mol Med* 53: 1018 – 1028

Kim K-P, Wu Y, Yoon J, Adachi K, Wu G, Velychko S, MacCarthy CM, Shin B, Röpke A, Arauzo-Bravo MJ *et al* (2020) Reprogramming competence of OCT factors is determined by transactivation domains. *Sci Adv* 6: eaaz7364

Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, Dietmann S, Nichols J, Smith A (2021a) Capture of mouse and human stem cells with features of formative pluripotency. *Cell Stem Cell* 28: 453 – 471

Kinoshita M, Barber M, Mansfield W, Cui Y, Spindlow D, Stirparo GG, Dietmann S, Nichols J, Smith A (2021b) Gene Expression Omnibus GSE131556 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131556). [DATASET]

Lackner A, Sehlke R, Garmhausen M, Stirparo GG, Huth M, Titz-Teixeira F, van der Lelij P, Ramesmayer J, Thomas HF, Ralser M *et al* (2021) Cooperative genetic networks drive embryonic stem cell transition from Naïve to formative pluripotency. *EMBO J* 40: e105776

Langlais D, Barreiro LB, Gros P (2016a) The macrophage IRF8/IRF1 regulome is required for protection against infections and is associated with chronic inflammation. *J Exp Med* 213: 585 – 603

Langlais D, Barreiro LB, Gros P (2016b) Gene Expression Omnibus GSE77886 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77886). [DATASET]

Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, Keren-Shaul H, Mildner A, Winter D, Jung S *et al* (2014) Chromatin state dynamics during blood formation. *Science* 345: 943 – 949

Leeb M, Dietmann S, Paramor M, Niwa H, Smith A (2014) Genetic exploration of the exit from self-renewal using haploid embryonic stem cells. *Cell Stem Cell* 14: 385 – 393

Li M, Yu JSL, Tilgner K, Ong SH, Koike-Yusa H, Yusa K (2018) Genome-wide CRISPR-KO screen uncovers MTORC1-mediated Gsk3 regulation in naive pluripotency maintenance and dissolution. *Cell Rep* 24: 489 – 502

Matrosovich M, Matrosovich T, Garten W, Klenk H-D (2006) New low-viscosity overlay medium for viral plaque assays. *Virol J* 3: 63

Matsuda K, Mikami T, Oki S, Iida H, Andrabi M, Boss JM, Yamaguchi K, Shigenobu S, Kondoh H (2017) ChIP-seq analysis of genomic binding regions of five major transcription factors highlights a central role for ZIC2 in the mouse epiblast stem cell gene regulatory network. *Development* 144: 1948 – 1958

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495 – 501

Michlits G, Hubmann M, Wu S-H, Vainorius G, Budusan E, Zhuk S, Burkard TR, Novatchkova M, Aichinger M, Lu Y *et al* (2017) CRISPR-UMI: single-cell lineage tracing of pooled CRISPR–Cas9 screens. *Nat Methods* 14: 1191 – 1197

Mittnenzweig M, Mayshar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, Chomsky E, Reines N, Uzonyi A, Lumerman L *et al* (2021) A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* 184: 2825 – 2842

Molè MA, Coorens THH, Shahbazi MN, Weberling A, Weatherbee BAT, Gantner CW, Sancho-Serra C, Richardson L, Drinkwater A, Syed N *et al* (2021a) A single cell characterisation of human embryogenesis identifies pluripotency transitions and putative anterior hypoblast centre. *Nat Commun* 12: 3679

Molè MA, Coorens THH, Shahbazi MN, Weberling A, Weatherbee BAT, Gantner CW, Sancho-Serra C, Richardson L, Drinkwater A, Syed N *et al* (2021b) Array Express E-MTAB-8060 (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8060/). [DATASET]

Molotkov A, Pierre Mazot J, Brewer R, Cinalli RM, Soriano P (2017) Distinct requirements for FGFR1 and FGFR2 in primitive endoderm development and exit from pluripotency. *Dev Cell* 41: 511 – 526

Moussaieff A, Rouleau M, Kitsberg D, Cohen M, Levy G, Barasch D, Nemirovski A, Shen-Orr S, Laevsky I, Amit M *et al* (2015) Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab* 21: 392 – 402

Oshiumi H, Miyashita M, Okamoto M, Morioka Y, Okabe M, Matsumoto M, Seya T (2015) DDX60 is involved in RIG-I-dependent and independent antiviral responses, and its function is attenuated by virus-induced EGFR activation. *Cell Rep* 11: 1193 – 1207

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7: 597 – 606

Percharde M, Wong P, Ramalho-Santos M (2017) Global Hypertranscription in the mouse embryonic germline. *Cell Rep* 19: 1987 – 1996

Pine R (1992) Constitutive expression of an ISGF2/IRF1 transgene leads to interferon-independent activation of interferon-inducible genes and resistance to virus infection. *J Virol* 66: 4470 – 4478

Platanitis E, Demiroz D, Schneller A, Fischer K, Capelle C, Hartl M, Gossenreiter T, Müller M, Novatchkova M, Decker T (2019) A molecular switch from STAT2-IRF9 to ISGF3 underlies interferon-induced gene transcription. *Nat Commun* 10: 2921

Ramsauer K, Farlik M, Zupkovitz G, Seiser C, Kroger A, Hauser H, Decker T (2007) Distinct modes of action applied by transcription factors STAT1 and IRF1 to initiate transcription of the IFN-γ-inducible Gbp2 gene. *Proc Natl Acad Sci USA* 104: 2849 – 2854

Sabapathy K (1997) Regulation of ES cell differentiation by functional and conformational modulation of P53. *EMBO J* 16: 6217 – 6229

Sachs P, Ding D, Bergmaier P, Lamp B, Schlagheck C, Finkernagel F, Nist A, Stiewe T, Mermoud JE (2019) SMARCAD1 ATPase activity is required to silence endogenous retroviruses in embryonic stem cells. *Nat Commun* 10: 1335

Seruggia D, Oti M, Tripathi P, Canver MC, LeBlanc L, Di Giammartino DC, Bullen MJ, Nefzger CM, Sun YBY, Farouni R *et al* (2019) TAF5L and TAF6L maintain self-renewal of embryonic stem cells via the MYC regulatory network. *Mol Cell* 74: 1148 – 1163

Shi B, Gao D, Zhong L, Zhi M, Weng X, Xu J, Li J, Du X, Xin Y, Gao J *et al* (2020) IRF-1 expressed in the inner cell mass of the porcine early blastocyst enhances the pluripotency of induced pluripotent stem cells. *Stem Cell Res Ther* 11: 505

Smith A (2017) Formative pluripotency: the executive phase in a developmental continuum. *Development* 144: 365 – 373

Stoltz KP, Jondle CN, Pulakanti K, Sylvester PA, Urrutia R, Rao S, Tarakanova VL (2019) Tumor suppressor interferon regulatory factor 1 selectively blocks expression of endogenous retrovirus. *Virology* 526: 52 – 60

Swartzendruber DE, Lehman JM (1975) Neoplastic differentiation: interaction of simian virus 40 and polyoma virus with murine teratocarcinoma cells *in vitro*. *J Cell Physiol* 85: 179 – 187

Thomas HF, Kotova E, Jayaram S, Pilz A, Romeike M, Lackner A, Penz T, Bock C, Leeb M, Halbritter F *et al* (2021) Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol Cell* 81: 969 – 982

Tycko J, Wainberg M, Marinov GK, Ursu O, Hess GT, Ego BK, Aradhana AL, Truong A, Trevino AE, Spees K *et al* (2019) Mitigation of off-target toxicity in CRISPR-Cas9 screens for essential non-coding elements. *Nat Commun* 10: 4063

Wang R, Wang J, Paul AM, Acharya D, Bai F, Huang F, Guo Y-L (2013) Mouse embryonic stem cells are deficient in type I interferon expression in response to viral infections and double-stranded RNA. *J Biol Chem* 288: 15926 – 15936

Wolf D, Goff SP (2009) Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 458: 1201 – 1204

Wray J, Kalkan T, Gomez-Lopez S, Eckardt D, Cook A, Kemler R, Smith A (2011) Inhibition of glycogen synthase Kinase-3 alleviates Tcf3 repression of the pluripotency network and increases embryonic stem cell resistance to differentiation. *Nat Cell Biol* 13: 838 – 845

Wu X, Thi VLD, Huang Y, Billerbeck E, Saha D, Hoffmann H-H, Wang Y, Vale Silva LA, Sarbanes S, Sun T *et al* (2018) Intrinsic immunity shapes viral resistance of stem cells. *Cell* 172: 423 – 438

Yamane D, Feng H, Rivera-Serrano EE, Selitsky SR, Hirai-Yuki A, Das A, McKnight KL, Misumi I, Hensley L, Lovell W *et al* (2019) Basal expression of interferon regulatory factor 1 drives intrinsic hepatocyte resistance to multiple RNA viruses. *Nat Microbiol* 4: 1096 – 1104

Yang S-H, Andrabi M, Biss R, Baker SM, Iqbal M, Sharrocks AD (2019) ZIC3 controls the transition from naive to primed pluripotency. *Cell Rep* 27: 3215 – 3227

Yang S-H, Kalkan T, Morissroe C, Marks H, Stunnenberg H, Smith A, Sharrocks AD (2014) Otx2 and Oct4 drive early enhancer activation during embryonic stem cell transition from naive pluripotency. *Cell Rep* 7: 1968 – 1981

Ye Y, Chen X, Zhang W (2021) Mammalian SWI/SNF chromatin remodeling complexes in embryonic stem cells: regulating the balance between pluripotency and differentiation. *Front Cell Dev Biol* 8. https://doi.org/10.3389/fcell.2020.626383

Yoneyama M, Kikuchi M, Matsumoto K, Imaizumi T, Miyagishi M, Taira K, Foy E, Loo Y-M, Michael Gale SA Jr, Yonehara S *et al* (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol* 175: 2851 – 2858

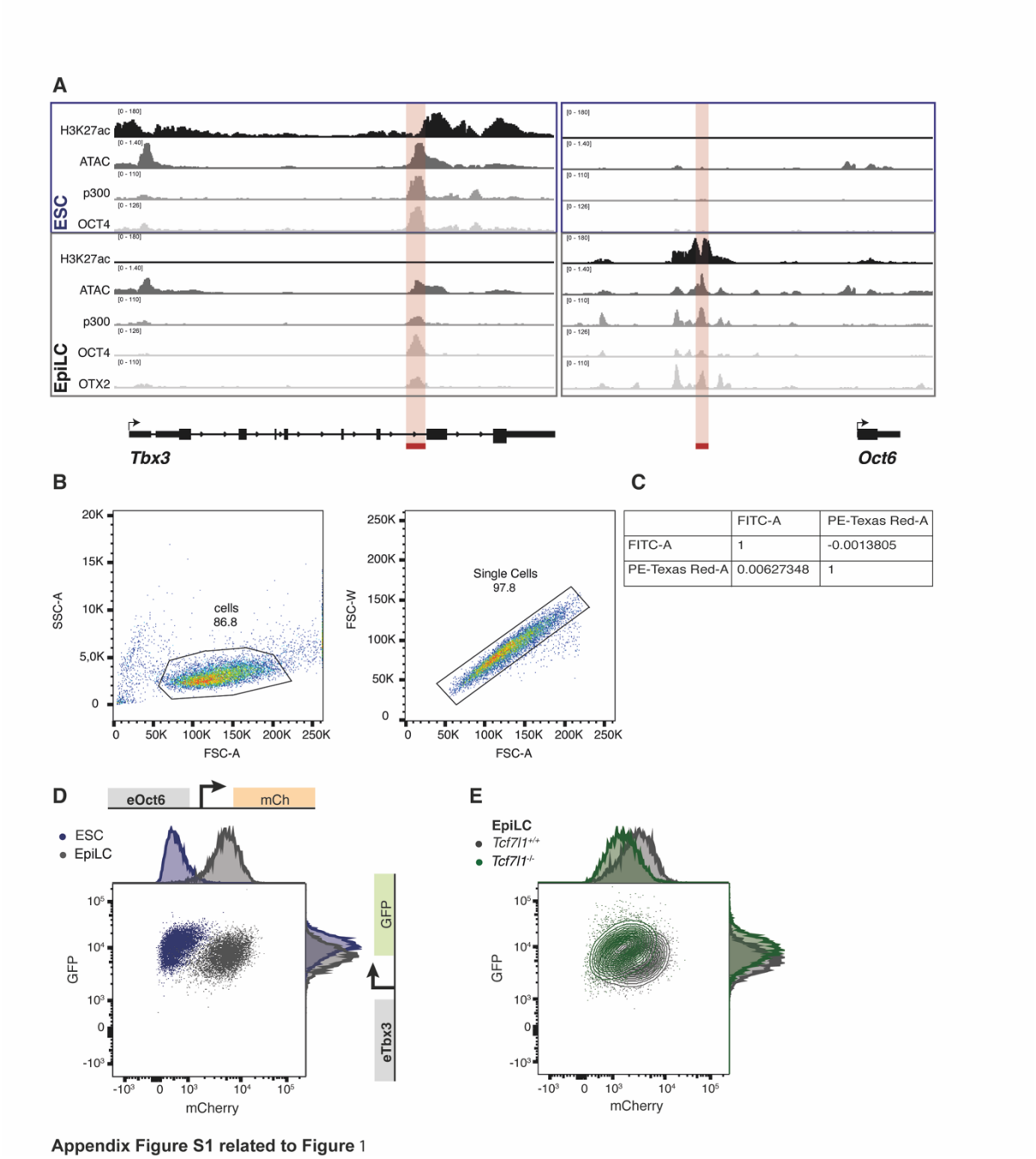
Transient upregulation of IRF1 during the exit from naive pluripotency confers viral protection

Merrit Romeike[1,3], Stephanie Spach[1], Marie Huber[1], Songjie Feng[1,3], Gintautas Vainorius[2,3], Ulrich Elling[2], Gjis A. Versteeg[1], Christa Buecker[1]*

[1]Max Perutz Labs Vienna, Vienna Biocenter (VBC), University of Vienna, Dr.Bohr-Gasse 9, 1030 Vienna, Austria
[2]Institute of Molecular Biotechnology of the Austrian Academy of Science (IMBA), Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[3]Vienna Biocenter PhD Program, a Doctoral School of the University of Vienna and Medical University of Vienna, A-1030 Vienna, Austria
*corresponding author

Appendix

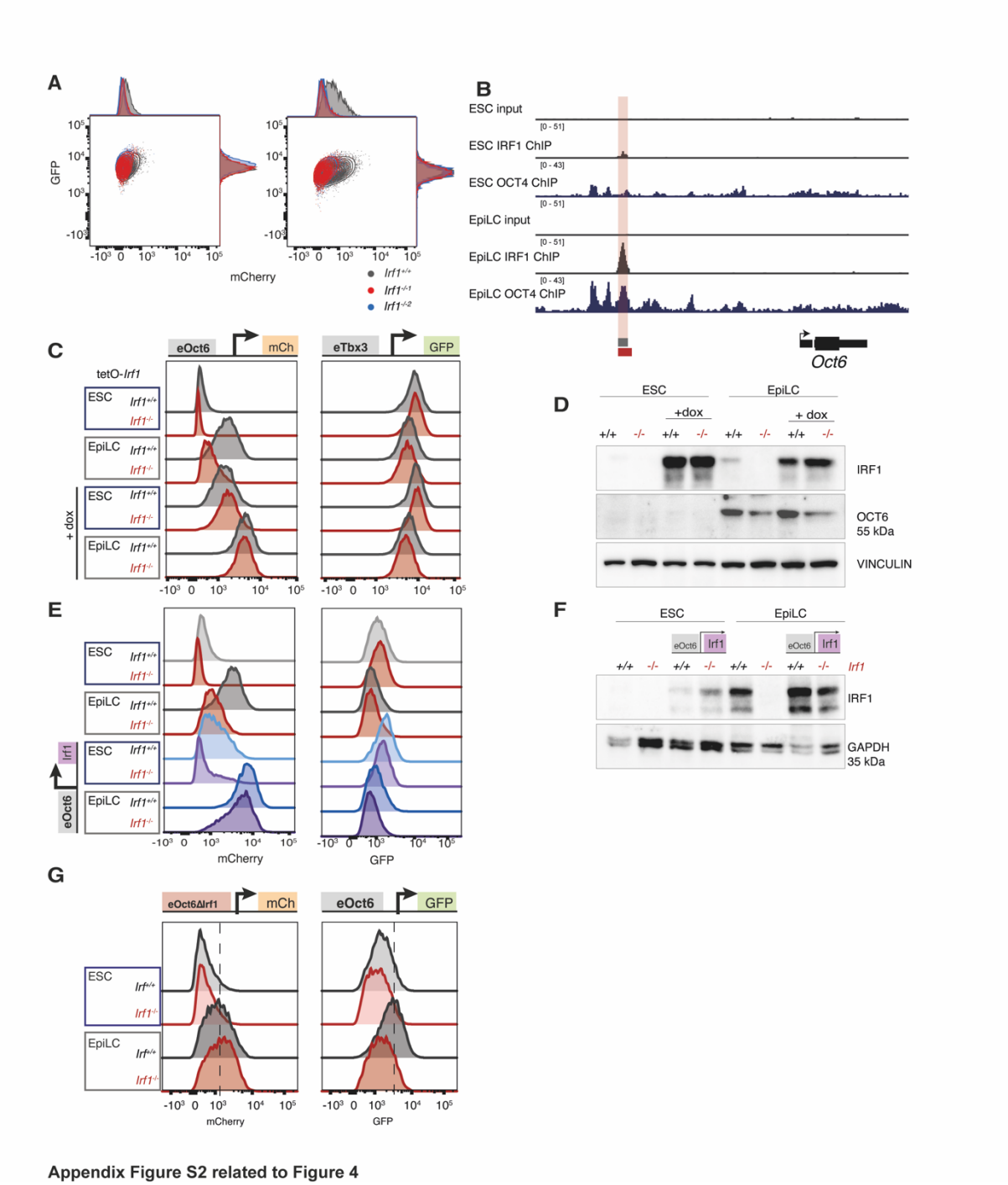
Table of Contents



Appendix Figure S1 related to Figure 1

Appendix Figure S1 related to Figure 1

A Chromatin context of *Tbx3* and *Oct6* loci in ESC and EpiLC. Enhancer regions used as reporters are indicated by red boxes. ChIP data was generated by Buecker et al. 2014.
B Example gating strategy for FACS analysis of screening cell lines.
C Compensation matrix for the used fluorescent reporter constructs, eTbx3-GFP measured in FITC-A, eOtc6-mCherry measured in PE-Texas-Red channels.
D Representative flow cytometry profiles of reporter cell lines in ESC and differentiated to EpiLC conditions.
E Representative flow cytometry profiles of reporter cell lines in EpiLC condition, *Tcf7l1*$^{+/+}$ and *Tcf7l1*$^{-/-}$.



Appendix Figure S2 related to Figure 4

Appendix Figure S2 related to Figure 4

A Representative flow cytometry profiles of reporter cell lines in ESC and EpiLC conditions, *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$. Shown are two independent cell lines.

B Chromatin context of *eOct6* locus, with a called IRF1 peak highlighted. IRF1 ChIP-seq tracks and OCT4 ChIP-seq tracks for ESC and EpiLCs are shown. The *eOct6* region used as reporter is marked with the lower red box. OCT4 ChIP-Seq data from Buecker et al., 2014.

C Representative flow cytometry profiles of reporter cell lines in ESC and EpiLC condition, *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ and doxycycline-induced IRF1 expression.

D Western Blot analysis of doxycycline-induced IRF1 expression in *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ background, in ESC and EpiLC cells, probed with antibodies against IRF1, OCT6 and GAPDH as loading control.

E Representative flow cytometry profiles of reporter cell lines in ESC and EpiLC condition, *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ and ectopic IRF1 expression driven by the *eOct6* enhancer.
F Western Blot analysis of ectopic IRF1 expression in Irf1$^{+/+}$ and Irf1$^{-/-}$ background, in ESC and EpiLC cells, probed with antibodies against IRF1 and GAPDH as loading control.
G Representative flow cytometry profiles of ESC and EpiLC cells, *Irf1*$^{+/+}$ and *Irf1*$^{-/-}$ with *eOct6ΔIrf* and *eOct6* controlling mCherry and GFP, respectively.

--- end of appendix --

Expanded View Figures

A

set	genes	guides
Transcription Factors	845	3290
Epigenetic Regulators	548	2134
Nuclear Genes Set 1	1994	7765
Nuclear Genes Set 2	2072	8069
DNA repair	265	1437
non-targeting control		86
total	2524	22781

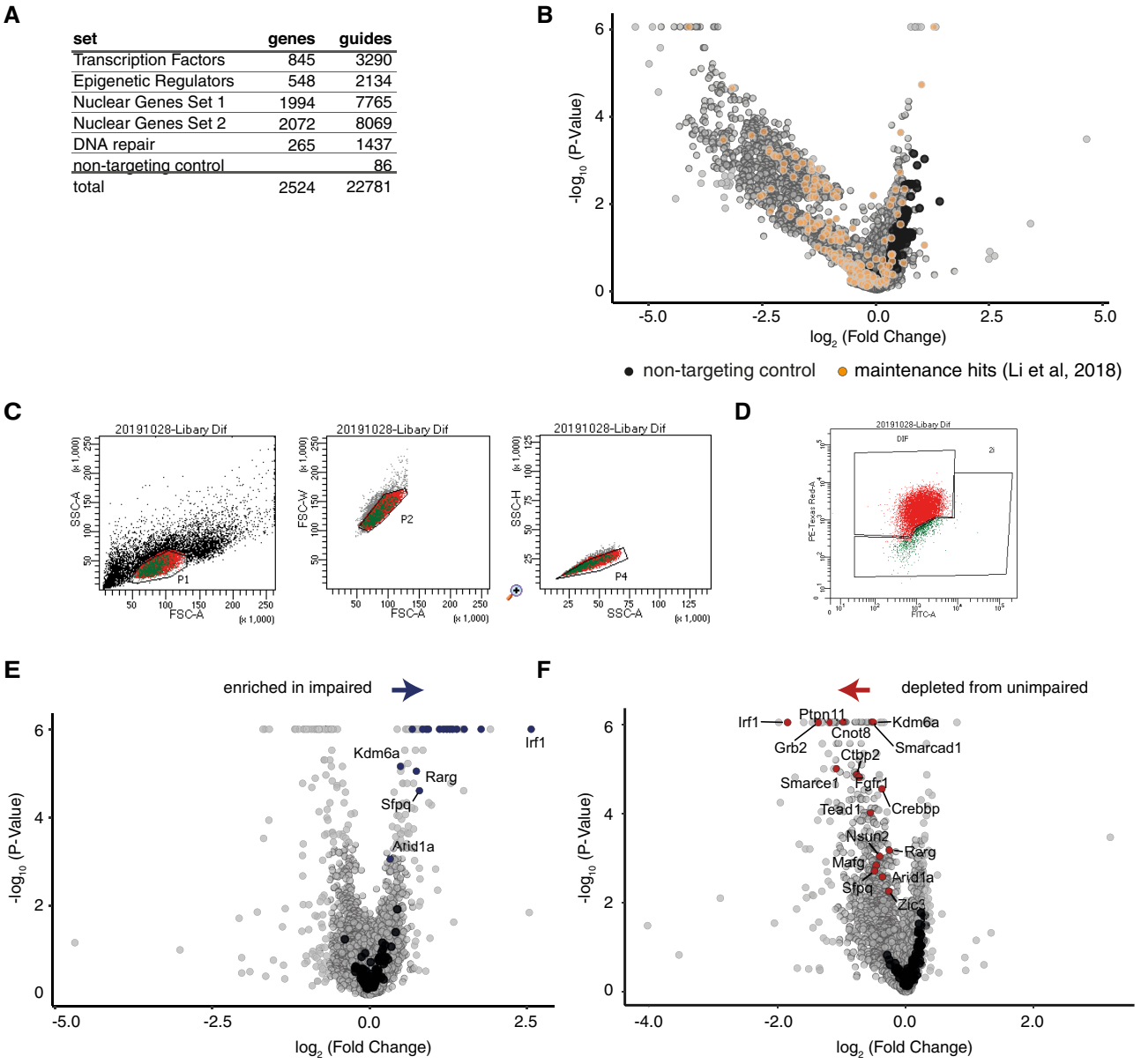


Figure EV1. Screen strategy to identify factors required for activation of formative pluripotency.

A Information about sgRNA library used in CRISPR-KO screen. Shown are number of targeted genes and unique guides used for each class of genes.

B As in Fig 1B, sgRNA representation in ESCs with CAS9 vs. ESCs without CAS9. Indicated are known ESC maintenance genes (Li *et al*, 2018).

C Example screening FACS gating strategy for CRISPR-KO screen for selection of single cells. Live cells were sorted according to FSC-A/SSC-A gates, followed by gating for FSC-A/FSC-W and SSC-A/SSC-H for individual cells.

D Example for gating strategy to sort impaired and unimpaired populations. Differentiation was scored by FITC-A/PE-TexasRed-A gating.

E As Fig 1C, sgRNA representation in EpiLC sorted as impaired vs. non-sorted EpiLC. Indicated in blue are factors identified as shared candidates (see Fig 1E).

F As Fig 1D, sgRNA representation in EpiLC sorted as unimpaired vs. non-sorted ESCs. Indicated in red are factors identified as shared candidates (see Fig 1E).

Figure EV2. IRF1 expression in formative pluripotency is controlled by an EpiLC-specific enhancer.

A Genotyping PCR of enhancer KO. Primers bind outside the edited site, producing smaller PCR products than on the control locus. Enhancer KO A and C were generated with one set of sgRNA, enhancer KO B and D with a second, different set.

B Examples for Sanger sequencing of PCR products, shown are enhancer KO B and D. Gapped alignment is indicated by dashed lines.

C Western blot analysis of *Irf1*$^{+/+}$, I*rf1*$^{-/-}$ and *Irf1* enhancer KO ESC and EpiLC, probed with antibodies against IRF1, formative marker OTX2 and VINCULIN as loading control.

D Western blot analysis of *Irf1*$^{+/+}$ and *Irf1* enhancer KO ESC and EpiLC with ruxolitinib treatment, probed with antibodies against phosphorylated STAT3, IRF1 and VINCULIN as loading control.

E, F RT-qPCR analysis of *Tbx3* (E, naive marker) and *Otx2* (F, formative marker) mRNA in *Irf1*$^{+/+}$ and *Irf1* enhancer KO in differentiation, treated with ruxolitinib. Fold change is normalized against *Rpl13a* housekeeping mRNA expression and is calculated against ESC for each indicated cell line. This is shown as baseline with the dashed line at fold change = 1. Shown are three biological replicates, datapoints from the same replicate are indicated with the same symbol. Note that each replicate contains two *Irf1*$^{+/+}$ samples. Black horizontal lines show the mean of the data.

G *Irf1* expression [FPKM] in murine embryos (Data ref: Boroviak *et al*, 2015b); $n = 3$ biological replicates.

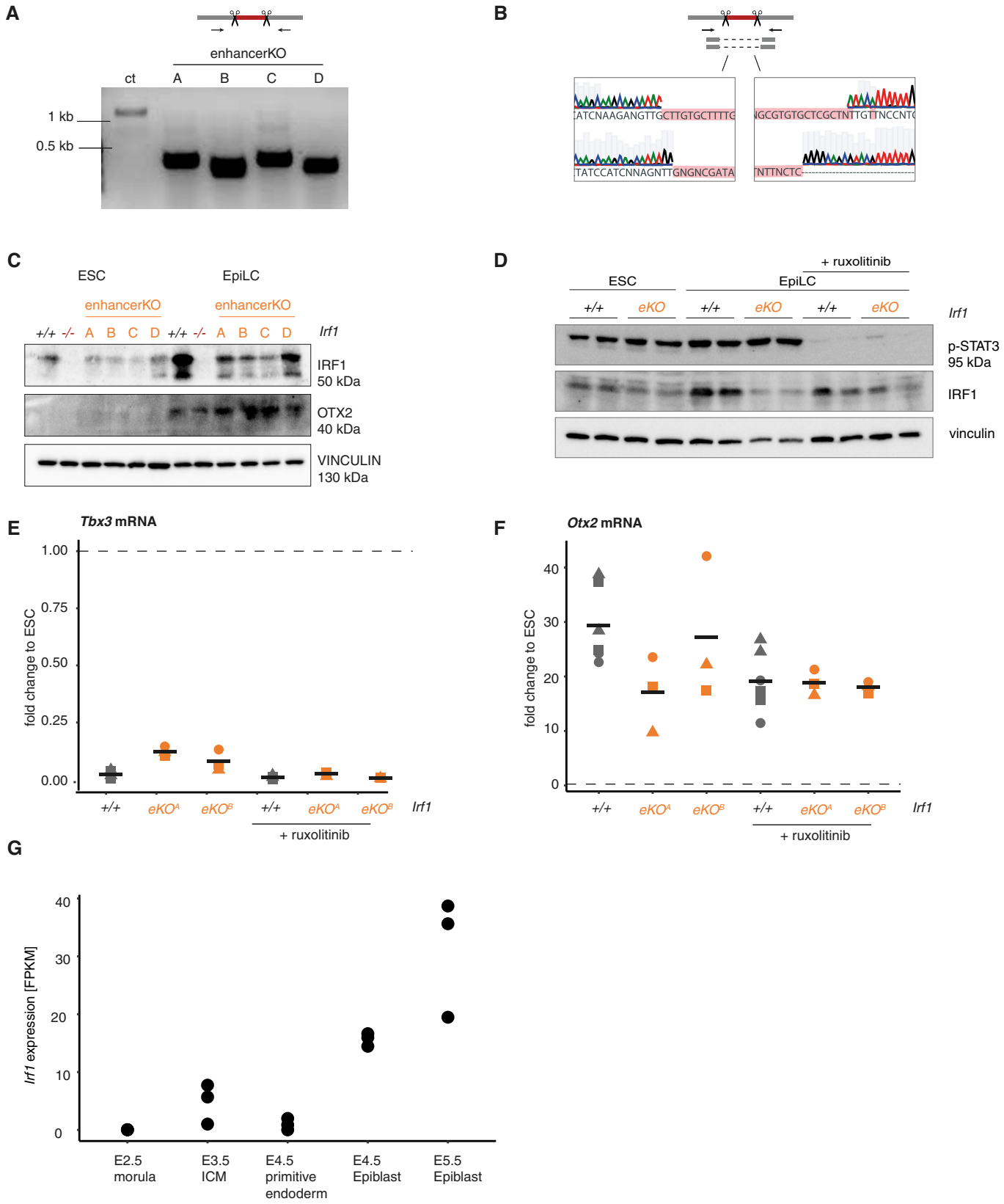


Figure EV2.

Figure EV3. IRF1 is not a major regulator of pluripotency gene or repeat element expression.

A Same Western blot as Fig 3A, long exposure and size marker of IRF1, VINCULIN is used as loading control. Overexposure is indicated with red signal.
B Expression changes in naive-associated genes (NAGs, Lackner *et al*, 2021) in $Irf1^{+/+}$ and $Irf1^{-/-}$ ESC and EpiLC based on QuantSeq RNA data. Data are shown as *z*-score. $n = 2$ biological replicates.
C Alkaline phosphatase staining of $Irf1^{+/+}$ and $Irf1^{-/-}$ in different cell seeding densities. Left panel: Cells were kept in 2i + LIF medium to control for seeding density, right panel: Cells were differentiated for 72 h before being placed back in 2i + LIF medium.
D MAplot of gene expression changes in EpiLC vs. ESC, ISGs are indicated. $n = 2$ biological replicates.
E MAplot of gene expression changes in EpiLC $Irf1^{-/-}$ vs. $Irf1^{+/+}$, ISGs are indicated. Differentially expressed (*P*-value < 0.05) Differentially expressed ISGs are plotted in Fig 3E. $n = 2$ biological replicates.
F Western blot analysis of dox-inducible SunTag-based IRF1 overexpression in ESCs, probed with antibodies against IRF1 and VINCULIN as loading control. Two different sgRNAs were tested, the sgRNA marked with an asterix (Irf1*) was used for RNA-sequencing.
G Dimension reduction (principal component analysis, PCA) plot of dox-inducible SunTag-based IRF1 overexpression in ESCs, based on QuantSeq RNA data. $n = 2$ biological replicates.
H Normalized counts for *Irf1* mRNA in dox-inducible SunTag-based IRF1 overexpression in ESCs, based on QuantSeq RNA data. *P*-value calculated by DESeq2. $n = 2$ biological replicates.
I Expression changes in dox-inducible SunTag-based IRF1 overexpression, left panel NAGS, right panel full set of ISGs. Data are shown as gene normalized *z*-score. Samples are color-coded as in Fig EV3G. $n = 2$ biological replicates.
J MAplot of TE family expression changes in EpiLC vs. ESC, adjusted *P*-values < 0.1 are indicated. $n = 2$ biological replicates.
K MAplot of TE family expression changes in EpiLC $Irf1^{-/-}$ vs. $Irf1^{+/+}$, adjusted *P*-values < 0.1 are indicated. $n = 2$ biological replicates.

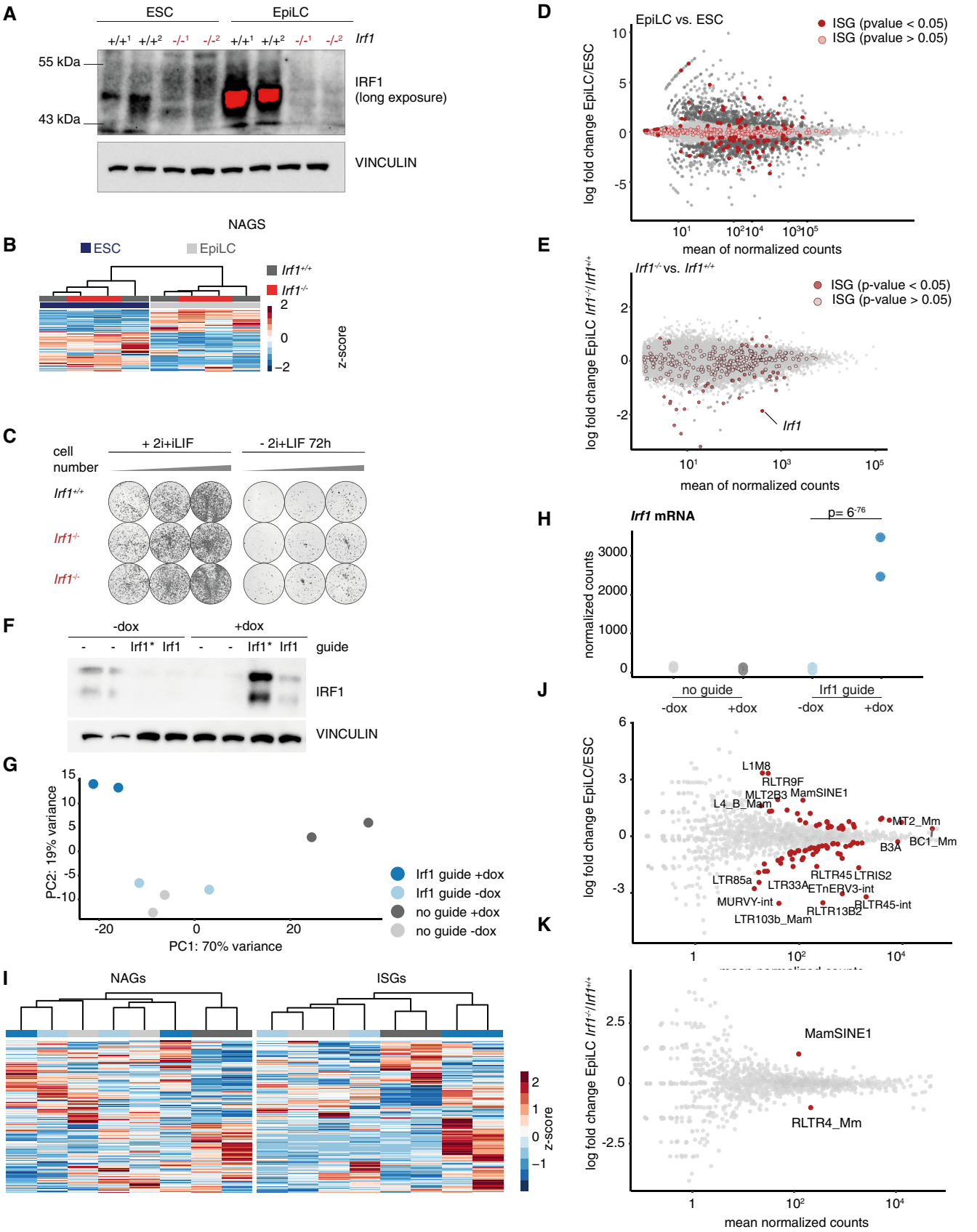


Figure EV3.

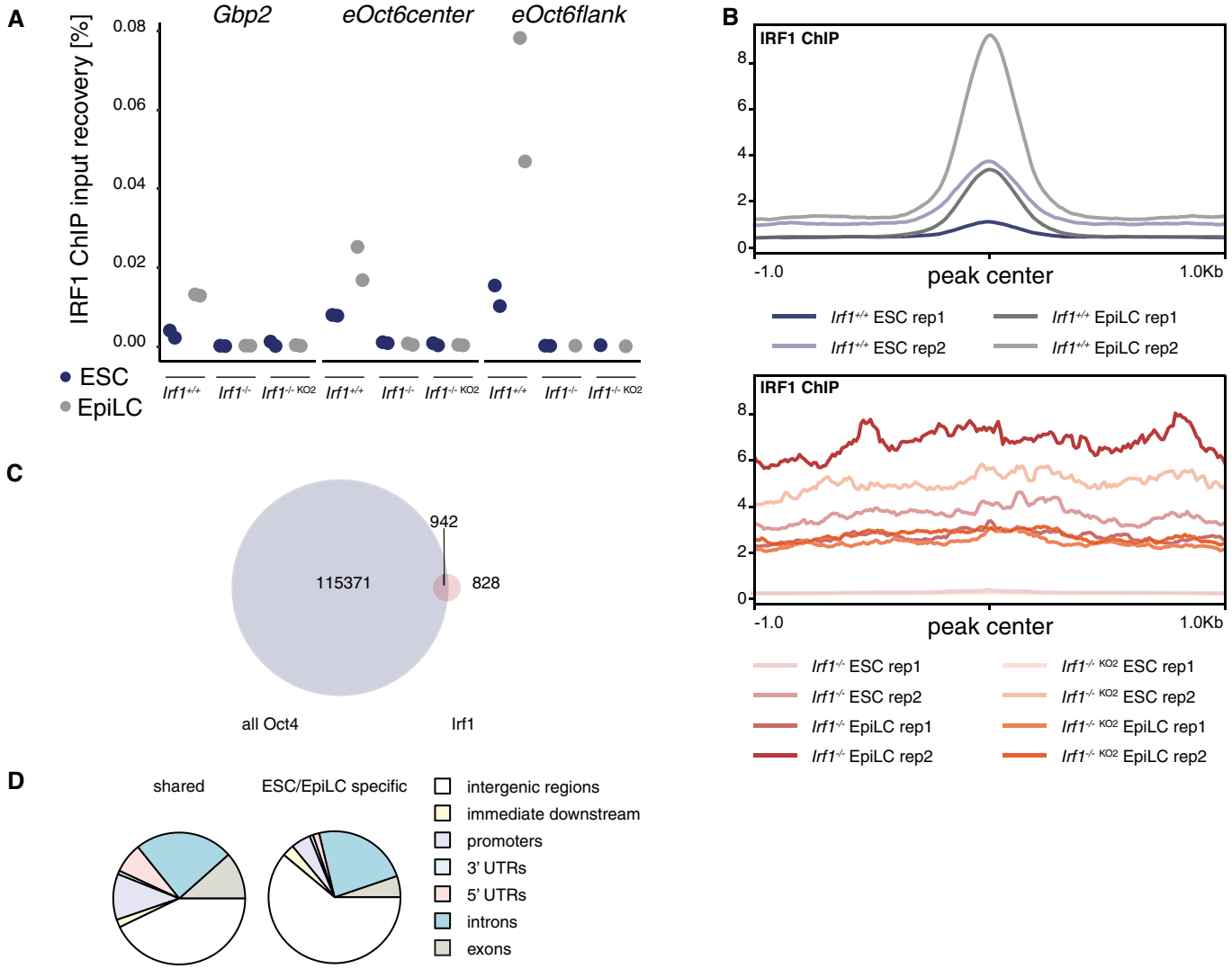


Figure EV4. Chromatin binding of IRF1.

A ChIP-qPCR analysis for the *Gbp2* promoter as known IRF1 binding site and two primer sets for the *eOct6* enhancer. Values are calculated as input recovery [%]. $n = 2$ biological replicates.

B Signal strength around all identified IRF1 binding sites for indicated conditions. Top panel: $Irf1^{+/+}$ ESC and EpiLCs, bottom panel: two independent $Irf1^{-/-}$ cell lines in ESC and EpiLC conditions. $n = 2$ biological replicates.

C Overlap of IRF1 binding sites and all OCT4 binding sites in ESCs and EpiLCs. (Data ref: Buecker *et al*, 2014b).

D Assigned chromatin regions of IRF1 binding sites shared between BMDMs and ESC/EpiLC (left) and IRF1 binding sites only detected in ESC/EpiLC (right).

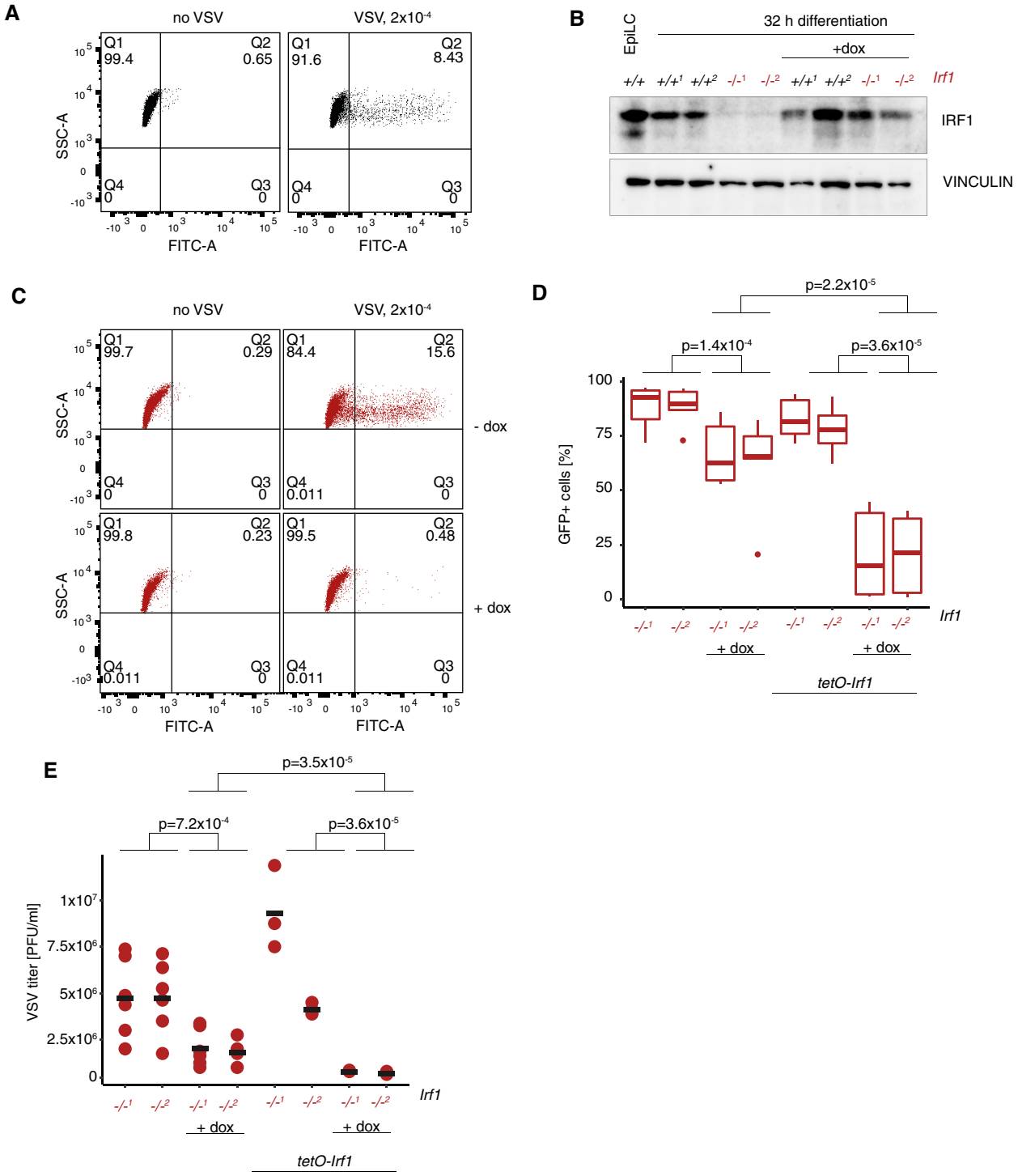


Figure EV5.

Figure EV5. IRF1 in EpiLCs is required for protection against viral infection.

A Representative FACS profiles of cells infected with GFP-VSV. Cells were scored according to FITC-A signal.

B Western blot analysis of IRF1 rescue with doxycycline-inducible *Irf1* construct. VINCULIN is used as loading control. All rescue samples were analyzed in two biological replicates.

C Representative FACS profiles of IRF1 doxycycline-inducible overexpression in cells infected with GFP-VSV, without and with doxycycline treatment. Cells were scored according to FITC-A signal.

D Quantification of GFP^+ cells after GFP-VSV infection (higher virus concentration, see Methods), $Irf1^{-/-}$ and doxycycline-inducible overexpression in $Irf1^{-/-}$ cells, without and with doxycycline induction. *P*-values were calculated per Wilcoxon test. $n = 6$ biological replicates. The central band shows the median, the box 25^{th} and 75^{th} percentiles, whiskers show 1.5*IGR (interquartile range).

E Viral titer as determined by plaque formation assays, $Irf1^{-/-}$ and IRF1 doxycycline-inducible overexpression in $Irf1^{-/-}$ cells, without and with doxycycline induction, as in panel (D). *P*-values were calculated per Wilcoxon test. $n = 3$ biological replicates for rescue, $n = 6$ for non-rescue conditions. Black horizontal lines show the mean of the data.

Discussion

In this thesis, I established a bottom-up, synthetic genomics platform to dissect how multiple enhancers cooperate across distance to regulate a shared promoter. This platform allows for the precise integration of single enhancers or enhancer pairs at defined distances and in specified orders upstream of a reporter gene, providing a controlled framework to study regulatory element function. Three core principles were discovered. First, as the distance between an enhancer and its target promoter increases, transcriptional output declines in a non-linear manner. Notably, strong and weak enhancers respond differently to increasing distance: strong enhancers tend to maintain a higher level of activity over longer distances, whereas weak enhancers exhibit a more pronounced loss of function as the distance increases. Second, super-additive activation is observed only when a strong enhancer is placed at a distal position, and a weak enhancer resides in the intervening region. The intrinsic strength of the enhancers plays a decisive role, since swapping the order of the strong and weak enhancers eliminates the super-additive activity. Moreover, the relative position of the weak enhancer further influences enhancer cooperation, as relocating it from 25 kb to 1.5 kb upstream of the promoter abolishes the synergistic effect. Third, CTCF sites positioned between the 75 kb and 25 kb loading panels attenuate the activity of a single strong enhancer at 75 kb as well as that of synergistic enhancer pairs, while leaving the synergistic interaction itself unaffected. Taken together, these results argue that long-range gene regulation is not a simple linear sum of independent inputs; instead, it is shaped by the joint effects of distance, order, and enhancer strength.

I also quantified the function of an EpiLC-specific enhancer regulating *Irf1*. My data indicate that this enhancer is regulated by the formative pluripotency gene regulatory network. As observed in EpiLCs, it is bound by OCT4, P300, and the formative transcription factor OTX2. In addition, it carries the active histone modifications H3K27ac and H3K4me1.

My synthetic reconstructions confirmed that enhancer activity declines with increasing distance, extending prior observations (Rinzema et al., 2022; Zuin et al., 2022). This observation is consistent with contact-facilitated models in which the effective rate of

productive E–P encounters decreases with genomic separation. Indeed, recently quantitative analyses from live-cell imaging have shown that the genomic distance between an enhancer and its target promoter primarily influences burst frequency, while having minimal impact on burst duration or burst size (Tünnermann et al., 2025). Consistently, single-molecule RNA FISH at the β-globin locus demonstrated that forced chromatin looping between LCR and the β-globin promoter selectively enhances burst frequency without altering burst size (Bartman et al., 2016).

However, alternative mechanisms may also contribute to the distance dependence of enhancer activity that I observed. The TAG model offers a complementary perspective on distance-dependent enhancer function and underscores the capacity of enhancer strength to buffer against the loss of regulatory activity imposed by genomic distance (Karr et al., 2022). In this framework, enhancer–promoter communication is mediated not by physical contacts but by the diffusion of active transcription factor complexes emanating from enhancers. As genomic distance increases, the effective concentration of ac-TFs at the promoter is predicted to decline, leading to reduced regulatory influence. Importantly, this attenuation can be partially compensated by enhancer strength: highly active enhancers generate larger pools of ac-TFs, which can sustain promoter activation even over extended genomic distances. This compensatory effect aligns with experimental observations that strong enhancers are less constrained by distance.

Perhaps the most intriguing finding is that enhancer synergy emerged only when a strong enhancer was distal and a weaker one was inserted proximally. Reversing the order or moving the weak element immediately proximal to the promoter largely eliminates synergy. Notably, *Map4k3 E1* can function as a relatively strong enhancer at 75 kb when paired with *Map4k3 E2* positioned at 25 kb, or as a weak enhancer at 25 kb in combination with stronger enhancers such as *Nanog* or *Rybp E2* at 75 kb. This suggests that the relative strength and positioning of cooperating enhancers, rather than their absolute activity or distance alone, may be key determinants of regulatory output.

Enhancers and promoters can assemble into hubs enriched with Mediator, BRD4, sequence-specific TFs, and RNA Polymerase II (Uyehara & Apostolou, 2023). While the precise mechanism of enhancer hubs remains debated, which may arise either through phase separation or multivalent scaffolding, a growing number of studies have described facilitator or tethering elements that lack autonomous enhancer activity yet boost the action of distal regulatory elements (Blayney et al., 2023; Bower et al., 2025; Brosh et al., 2023; X. Li & Levine, 2024). In such hubs, the strong enhancer may act as a nucleation site, recruiting initiator TFs and providing high-valency interactions that stabilize the overall structure. A weak enhancer may contribute in more subtle ways, such as promoting docking interactions or prolonging the residence time of transcriptional machinery at the promoter. In such configurations, a proximal weak element may not activate transcription independently but can nonetheless facilitate or stabilize the enhancer–promoter contacts. This offers a plausible explanation for how weak elements can amplify the output of more potent, distal enhancers.

Consistent with this model, one study proposed that enhancer compatibility and coordination rely on the formation of multiple cooperative environments. Importantly, the capacity for such coordination appears to reside primarily with the weaker, dependent enhancer element (Kribelbauer-Swietek et al., 2024). Another study at the *Sox2* locus showed that local genomic elements such as weak enhancers are required for condensates to enhance transcriptional bursting (Du et al., 2024). Removal of these elements eliminates the condensate-associated increases in burst size or frequency, indicating that nearby regulatory sequences are indispensable for translating condensate proximity into transcriptional output. This requirement is conceptually consistent with the roles of facilitator or tethering elements, which, despite lacking strong autonomous activity, enable distal enhancers to achieve their full regulatory potential. My data further identify enhancer order as a determinant of cooperation. Specifically, I show that enhancer pairs exhibiting additive behavior can be functionally shifted toward cooperation when the proximal, weaker element enhances the effective engagement of a distal, stronger partner. These observations support the idea that such elements represent a functionally analogous class of regulatory components that act by enabling or stabilizing long-range enhancer–promoter communication.

I also examined the role of CTCF-binding sites in modulating enhancer activity within the system. The presence of two CTCF sites reduced transcriptional output but did not fully disrupt enhancer–promoter communication, and the extent of attenuation appeared to depend on both enhancer strength and local chromatin architecture. In the wild-type background lacking integrated reporters or enhancers, these CTCF sites did not exhibit any detectable structural configuration. These findings suggest that CTCF does not act as an absolute insulator capable of blocking enhancer–promoter communication under these conditions. Further studies are needed to determine how deletion of CTCF-binding sites alters local three-dimensional chromatin architecture.

Limitations

While our synthetic platform illuminates rules of enhancer cooperation, it has limitations. First, our primary readout is reporter protein, and I did not directly image nascent transcription. Although analysis by Flow-FISH confirmed that mCherry transcript abundance was consistently aligned with reporter protein output, I did not characterize how synergistic effects influence burst frequency, transcriptional output per burst, and transcriptional activity windows. Our reporter platform contains 24×PP7 in the intron of mCherry and PCP-GFP, enabling live-cell imaging of nascent RNA. Applying PP7/PCP-based live imaging will allow us to quantify how enhancer–promoter distance and order influence burst kinetics. Such analyses will help determine whether synergistic effects arise through increased encounter frequency, enhanced burst size, prolonged burst duration, or a combination of these mechanisms. Second, our synthetic locus uses a minimal TK promoter. At close distances, many strong enhancers reach similar expression levels, indicating potential promoter saturation. Future work could explore how different classes of core promoters, such as housekeeping and developmental promoters, interact with diverse enhancer pairs. Examining these combinations may clarify how enhancer–enhancer–promoter configurations shape cooperative behavior, influence distance-dependent effects, and define enhancer–promoter compatibility. Third, our analysis was limited to a small number of enhancer pairs tested across defined distances. Higher-throughput approaches, such as improved integration strategies or PiggyBac-based looping systems, could allow the systematic interrogation of a broader spectrum of cis-regulatory elements. In addition, our synthetic constructs were integrated only at the β-globin locus. Although this site is considered

relatively neutral, it does not fully recapitulate locus-specific chromatin features such as subnuclear positioning or the presence of pre-existing regulatory hubs. Lastly, I did not perturb cohesin, CTCF, Mediator, or BRD4. Our inferences about the 3D structure of enhancer hubs and condensates remain indirect. Future experiments that acutely perturb architectural proteins and coactivators, for example by degron-based depletion or by genetic disruption of these factors, will be required to clarify whether enhancer synergy is driven mainly by extrusion-mediated looping or by condensates.

In conclusion, my thesis identifies principles underlying long-range enhancer cooperation. These rules explain how many enhancer pairs act additively while a subset displays synergy, and how elements with low intrinsic activity can become essential in favorable spatial arrangements. The findings highlight that information on enhancer positioning is as critical as intrinsic enhancer strength for the interpretation of cis-regulatory systems.

Reference

Abdella, R., Talyzina, A., Chen, S., Inouye, C. J., Tjian, R., & He, Y. (2021). Structure of the human Mediator-bound transcription preinitiation complex. *Science (New York, N.Y.)*, *372*(6537), 52–56. https://doi.org/10.1126/science.abg3074

Aboreden, N. G., Zhao, H., Shan, F., Liu, F., Zhang, H., & Blobel, G. A. (2025). *Cis-regulatory chromatin contacts form de novo in the absence of loop extrusion* (p. 2025.01.12.632634). bioRxiv. https://doi.org/10.1101/2025.01.12.632634

Acuña, L. I. G., Flyamer, I., Boyle, S., Friman, E. T., & Bickmore, W. A. (2024). Transcription decouples estrogen-dependent changes in enhancer-promoter contact frequencies and spatial proximity. *PLOS Genetics*, *20*(5), e1011277. https://doi.org/10.1371/journal.pgen.1011277

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., Huang, B., Lomvardas, S., & Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife*, *8*, e41769. https://doi.org/10.7554/eLife.41769

Allahyar, A., Vermeulen, C., Bouwman, B. A. M., Krijger, P. H. L., Verstegen, M. J. A. M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J. H. I., Jalink, K., Teunissen, H., Renkens, I. J., Kloosterman, W. P., Rowland, B. D., de Wit, E., de Ridder, J., & de Laat, W. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. *Nature Genetics*, *50*(8), 1151–1160. https://doi.org/10.1038/s41588-018-0161-5

Andrey, G., Montavon, T., Mascrez, B., Gonzalez, F., Noordermeer, D., Leleu, M., Trono, D., Spitz, F., & Duboule, D. (2013). A switch between topological domains underlies HoxD

genes collinearity in mouse limbs. *Science (New York, N.Y.)*, *340*(6137), 1234167. https://doi.org/10.1126/science.1234167

Andrey, G., Schöpflin, R., Jerković, I., Heinrich, V., Ibrahim, D. M., Paliou, C., Hochradel, M., Timmermann, B., Haas, S., Vingron, M., & Mundlos, S. (2017). Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Research*, *27*(2), 223–233. https://doi.org/10.1101/gr.213066.116

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)*, *339*(6123), 1074–1077. https://doi.org/10.1126/science.1232542

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2), 299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Barbadilla-Martínez, L., Klaassen, N., Franceschini-Santos, V. H., Breda, J., Hernandez-Quiles, M., Lieshout, T. van, Traslaviña, C. G. U., Yücel, H., Boi, M. C. L., Hermana-Garcia-Agullo, C., Gregoricchio, S., Zwart, W., Voest, E., Franke, L., Vermeulen, M., Ridder, J. de, & Steensel, B. van. (2024). *The regulatory grammar of human promoters uncovered by MPRA-trained deep learning* (p. 2024.07.09.602649). bioRxiv. https://doi.org/10.1101/2024.07.09.602649

Bartman, C. R., Hsu, S. C., Hsiung, C. C.-S., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, *62*(2), 237–247. https://doi.org/10.1016/j.molcel.2016.03.007

Baudic, M., Murata, H., Bosada, F. M., Melo, U. S., Aizawa, T., Lindenbaum, P., van der Maarel, L. E., Guedon, A., Baron, E., Fremy, E., Foucal, A., Ishikawa, T., Ushinohama,

H., Jurgens, S. J., Choi, S. H., Kyndt, F., Le Scouarnec, S., Wakker, V., Thollet, A., … Barc, J. (2024). TAD boundary deletion causes PITX2-related cardiac electrical and structural defects. *Nature Communications*, *15*(1), 3380. https://doi.org/10.1038/s41467-024-47739-x

Beliveau, B. J., Joyce, E. F., Apostolopoulos, N., Yilmaz, F., Fonseka, C. Y., McCole, R. B., Chang, Y., Li, J. B., Senaratne, T. N., Williams, B. R., Rouillard, J.-M., & Wu, C. (2012). Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(52), 21301–21306. https://doi.org/10.1073/pnas.1213818110

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., Grimes, G. R., Therizols, P., & Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell*, *76*(3), 473-484.e7. https://doi.org/10.1016/j.molcel.2019.07.038

Bestor, T., Laudano, A., Mattaliano, R., & Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *Journal of Molecular Biology*, *203*(4), 971–983. https://doi.org/10.1016/0022-2836(88)90122-2

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., & Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science (New York, N.Y.)*, *313*(5793), 1642–1645. https://doi.org/10.1126/science.1127344

Bianco, S., Lupiáñez, D. G., Chiariello, A. M., Annunziatella, C., Kraft, K., Schöpflin, R., Wittler, L., Andrey, G., Vingron, M., Pombo, A., Mundlos, S., & Nicodemi, M. (2018).

Polymer physics predicts the effects of structural variants on chromatin architecture. *Nature Genetics*, *50*(5), 662–667. https://doi.org/10.1038/s41588-018-0098-8

Blayney, J. W., Francis, H., Rampasekova, A., Camellato, B., Mitchell, L., Stolper, R., Cornell, L., Babbs, C., Boeke, J. D., Higgs, D. R., & Kassouf, M. (2023). Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell*, *186*(26), 5826-5839.e18. https://doi.org/10.1016/j.cell.2023.11.030

Bossard, P., & Zaret, K. S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. *Development*, *125*(24), 4909–4917. https://doi.org/10.1242/dev.125.24.4909

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T., & Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *eLife*, *4*, e07956. https://doi.org/10.7554/eLife.07956

Bower, G., Hollingsworth, E. W., Jacinto, S. H., Alcantara, J. A., Clock, B., Cao, K., Liu, M., Dziulko, A., Alcaina-Caro, A., Xu, Q., Skowronska-Krawczyk, D., Lopez-Rios, J., Dickel, D. E., Bardet, A. F., Pennacchio, L. A., Visel, A., & Kvon, E. Z. (2025). Range extender mediates long-distance enhancer activity. *Nature*, *643*(8072), 830–838. https://doi.org/10.1038/s41586-025-09221-6

Bozhilov, Y. K., Downes, D. J., Telenius, J., Marieke Oudelaar, A., Olivier, E. N., Mountford, J. C., Hughes, J. R., Gibbons, R. J., & Higgs, D. R. (2021). A gain-of-function single nucleotide variant creates a new promoter which acts as an orientation-dependent enhancer-blocker. *Nature Communications*, *12*(1), 3806. https://doi.org/10.1038/s41467-021-23980-6

Brosh, R., Coelho, C., Ribeiro-Dos-Santos, A. M., Ellis, G., Hogan, M. S., Ashe, H. J., Somogyi, N., Ordoñez, R., Luther, R. D., Huang, E., Boeke, J. D., & Maurano, M. T. (2023).

Synthetic regulatory genomics uncovers enhancer context dependence at the Sox2 locus. *Molecular Cell*, *83*(7), 1140-1152.e7. https://doi.org/10.1016/j.molcel.2023.02.027

Bu, H., Gan, Y., Wang, Y., Zhou, S., & Guan, J. (2017). A new method for enhancer prediction based on deep belief network. *BMC Bioinformatics*, *18*(12), 418. https://doi.org/10.1186/s12859-017-1828-0

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & 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*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Calhoun, V. C., Stathopoulos, A., & Levine, M. (2002). Promoter-proximal tethering elements regulate enhancer-promoter specificity in the *Drosophila Antennapedia* complex. *Proceedings of the National Academy of Sciences*, *99*(14), 9243–9247. https://doi.org/10.1073/pnas.142291299

Caravaca, J. M., Donahue, G., Becker, J. S., He, X., Vinson, C., & Zaret, K. S. (2013). Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes & Development*, *27*(3), 251–260. https://doi.org/10.1101/gad.206458.112

Carey, M. (1998). The Enhanceosome and Transcriptional Synergy. *Cell*, *92*(1), 5–8. https://doi.org/10.1016/S0092-8674(00)80893-4

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. *Nature Genetics*, *50*(9), 1296–1303. https://doi.org/10.1038/s41588-018-0175-z

Chen, J., Zhang, Z., Li, L., Chen, B.-C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R., & Liu, Z. (2014). Single-molecule dynamics of enhanceosome

assembly in embryonic stem cells. *Cell*, *156*(6), 1274–1285. https://doi.org/10.1016/j.cell.2014.01.062

Chen, L., & Capra, J. A. (2020). Learning and interpreting the gene regulatory grammar in a deep learning framework. *PLOS Computational Biology*, *16*(11), e1008334. https://doi.org/10.1371/journal.pcbi.1008334

Chereji, R. V., Bryson, T. D., & Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. *Genome Biology*, *20*(1), 198. https://doi.org/10.1186/s13059-019-1815-z

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T. V., Graf, T., & Cramer, P. (2021). Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. *eLife*, *10*, e65381. https://doi.org/10.7554/eLife.65381

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., & Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*. https://doi.org/10.1126/science.aar2555

Cirillo, L. A., Lin, F. R., Cuesta, I., Friedman, D., Jarnik, M., & Zaret, K. S. (2002). Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, *9*(2), 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S., & Stern, D. L. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell*, *160*(1–2), 191–203. https://doi.org/10.1016/j.cell.2014.11.041

Davies, J. O. J., Telenius, J. M., McGowan, S. J., Roberts, N. A., Taylor, S., Higgs, D. R., & Hughes, J. R. (2016). Multiplexed analysis of chromosome conformation at vastly

improved sensitivity. *Nature Methods*, *13*(1), 74–80. https://doi.org/10.1038/nmeth.3664

de Almeida, B. P., Reiter, F., Pagani, M., & Stark, A. (2022). DeepSTARR predicts enhancer activity from DNA sequence and enables the de novo design of synthetic enhancers. *Nature Genetics*, *54*(5), 613–624. https://doi.org/10.1038/s41588-022-01048-5

de Almeida, B. P., Schaub, C., Pagani, M., Secchia, S., Furlong, E. E. M., & Stark, A. (2024). Targeted design of synthetic enhancers for selected tissues in the Drosophila embryo. *Nature*, *626*(7997), 207–211. https://doi.org/10.1038/s41586-023-06905-9

De Gobbi, M., Viprakasit, V., Hughes, J. R., Fisher, C., Buckle, V. J., Ayyub, H., Gibbons, R. J., Vernimmen, D., Yoshinaga, Y., de Jong, P., Cheng, J.-F., Rubin, E. M., Wood, W. G., Bowden, D., & Higgs, D. R. (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science (New York, N.Y.)*, *312*(5777), 1215–1217. https://doi.org/10.1126/science.1126431

de la Serna, I. L., Ohkawa, Y., Berkes, C. A., Bergstrom, D. A., Dacwag, C. S., Tapscott, S. J., & Imbalzano, A. N. (2005). MyoD Targets Chromatin Remodeling Complexes to the Myogenin Locus Prior to Forming a Stable DNA-Bound Complex. *Molecular and Cellular Biology*, *25*(10), 3997–4009. https://doi.org/10.1128/MCB.25.10.3997-4009.2005

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–1311. https://doi.org/10.1126/science.1067799

Delaneau, O., Zazhytska, M., Borel, C., Giannuzzi, G., Rey, G., Howald, C., Kumar, S., Ongen, H., Popadin, K., Marbach, D., Ambrosini, G., Bielser, D., Hacker, D., Romano, L., Ribaux, P., Wiederkehr, M., Falconnet, E., Bucher, P., Bergmann, S., … Dermitzakis,

E. T. (2019). Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*. https://doi.org/10.1126/science.aat8266

Deluz, C., Friman, E. T., Strebinger, D., Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S., & Suter, D. M. (2016). A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. *Genes & Development*, *30*(22), 2538–2550. https://doi.org/10.1101/gad.289256.116

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., Dean, A., & Blobel, G. A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell*, *149*(6), 1233–1244. https://doi.org/10.1016/j.cell.2012.03.051

Deng, W., Rupon, J. W., Krivega, I., Breda, L., Motta, I., Jahn, K. S., Reik, A., Gregory, P. D., Rivella, S., Dean, A., & Blobel, G. A. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell*, *158*(4), 849–860. https://doi.org/10.1016/j.cell.2014.05.050

Deplancke, B., Alpern, D., & Gardeux, V. (2016). The Genetics of Transcription Factor DNA Binding Variation. *Cell*, *166*(3), 538–554. https://doi.org/10.1016/j.cell.2016.07.012

Di Giammartino, D. C., Kloetgen, A., Polyzos, A., Liu, Y., Kim, D., Murphy, D., Abuhashem, A., Cavaliere, P., Aronson, B., Shah, V., Dephoure, N., Stadtfeld, M., Tsirigos, A., & Apostolou, E. (2019). KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nature Cell Biology*, *21*(10), 1179–1190. https://doi.org/10.1038/s41556-019-0390-6

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

Du, M., Stitzinger, S. H., Spille, J.-H., Cho, W.-K., Lee, C., Hijaz, M., Quintana, A., & Cissé, I. I. (2024). Direct observation of a condensate effect on super-enhancer controlled gene bursting. *Cell*, *187*(2), 331-344.e17. https://doi.org/10.1016/j.cell.2023.12.005

Erwin, G. D., Oksenberg, N., Truty, R. M., Kostka, D., Murphy, K. K., Ahituv, N., Pollard, K. S., & Capra, J. A. (2014). Integrating Diverse Datasets Improves Developmental Enhancer Prediction. *PLOS Computational Biology*, *10*(6), e1003677. https://doi.org/10.1371/journal.pcbi.1003677

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A. D., & Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Research*, *26*(12), 1345–1348. https://doi.org/10.1038/cr.2016.137

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S., & Levine, M. S. (2015). Suboptimization of developmental enhancers. *Science (New York, N.Y.)*, *350*(6258), 325–328. https://doi.org/10.1126/science.aac6948

Festuccia, N., Dubois, A., Vandormael-Pournin, S., Gallego Tejeda, E., Mouren, A., Bessonnard, S., Mueller, F., Proux, C., Cohen-Tannoudji, M., & Navarro, P. (2016). Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. *Nature Cell Biology*, *18*(11), 1139–1148. https://doi.org/10.1038/ncb3418

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., Spielmann, M., Timmermann, B., Wittler, L., Kurth, I., Cambiaso, P., Zuffardi, O., Houge, G., Lambie, L., Brancati, F., … Mundlos, S. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, *538*(7624), 265–269. https://doi.org/10.1038/nature19800

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368. https://doi.org/10.1016/j.cell.2016.05.025

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G. Y., Huang, P. Y. H., Welboren, W.-J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., … Ruan, Y. (2009). An oestrogen-receptor-α-bound human chromatin interactome. *Nature*, *462*(7269), 58–64. https://doi.org/10.1038/nature08497

Furlong, E. E. M., & Levine, M. (2018). Developmental enhancers and chromosome topology. *Science (New York, N.Y.)*, *361*(6409), 1341–1345. https://doi.org/10.1126/science.aau0320

Gassler, J., Kobayashi, W., Gáspár, I., Ruangroengkulrith, S., Mohanan, A., Hernández, L. G., Kravchenko, P., Kümmecke, M., Lalic, A., Rifel, N., Ashburn, R. J., Zaczek, M., Vallot, A., Rico, L. C., Ladstätter, S., & Tachibana, K. (2022). Zygotic genome activation by the totipotency pioneer factor Nr5a2. *Science*. https://doi.org/10.1126/science.abn7478

Georgakopoulos-Soares, I., Deng, C., Agarwal, V., Chan, C. S. Y., Zhao, J., Inoue, F., & Ahituv, N. (2023). Transcription factor binding site orientation and order are major drivers of gene regulatory activity. *Nature Communications*, *14*(1), 2333. https://doi.org/10.1038/s41467-023-37960-5

Ghavi-Helm, Y., Jankowski, A., Meiers, S., Viales, R. R., Korbel, J. O., & Furlong, E. E. M. (2019). Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics*, *51*(8), 1272–1282. https://doi.org/10.1038/s41588-019-0462-3

Goel, V. Y., Huseyin, M. K., & Hansen, A. S. (2023). Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nature Genetics*, *55*(6), 1048–1056. https://doi.org/10.1038/s41588-023-01391-1

Gosai, S. J., Castro, R. I., Fuentes, N., Butts, J. C., Mouri, K., Alasoadura, M., Kales, S., Nguyen, T. T. L., Noche, R. R., Rao, A. S., Joy, M. T., Sabeti, P. C., Reilly, S. K., &

Tewhey, R. (2024). Machine-guided design of cell-type-targeting cis-regulatory elements. *Nature*, *634*(8036), 1211–1220. https://doi.org/10.1038/s41586-024-08070-z

Gottesfeld, J. M., & Forbes, D. J. (1997). Mitotic repression of the transcriptional machinery. *Trends in Biochemical Sciences*, *22*(6), 197–202. https://doi.org/10.1016/s0968-0004(97)01045-1

Guckelberger, P., Doughty, B. R., Munson, G., Rao, S. S. P., Tan, Y., Cai, X. S., Fulco, C. P., Nasser, J., Mualim, K. S., Bergman, D. T., Ray, J., Jagoda, E., Munger, C. J., Gschwind, A. R., Sheth, M. U., Tan, A. S., Steinmetz, L. M., Lander, E. S., Meissner, A., … Engreitz, J. M. (2024). *Cohesin-mediated 3D contacts tune enhancer-promoter regulation* (p. 2024.07.12.603288). bioRxiv. https://doi.org/10.1101/2024.07.12.603288

Hafner, A., Park, M., Berger, S. E., Murphy, S. E., Nora, E. P., & Boettiger, A. N. (2023). Loop stacking organizes genome folding from TADs to chromosomes. *Molecular Cell*, *83*(9), 1377-1392.e6. https://doi.org/10.1016/j.molcel.2023.04.008

Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J., & Fazzio, T. G. (2019). Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell*, *177*(5), 1319-1329.e11. https://doi.org/10.1016/j.cell.2019.03.014

Hansen, J. L., & Cohen, B. A. (2022). A quantitative metric of pioneer activity reveals that HNF4A has stronger in vivo pioneer activity than FOXA1. *Genome Biology*, *23*(1), 221. https://doi.org/10.1186/s13059-022-02792-x

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., Gosden, M., Sharpe, J. A., Sloane-Stanley, J. A., Hughes, J. R., Davies, B., & Higgs, D. R. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, *19*(8), 952–961. https://doi.org/10.1038/ncb3573

Harris, H. L., Gu, H., Olshansky, M., Wang, A., Farabella, I., Eliaz, Y., Kalluchi, A., Krishna, A., Jacobs, M., Cauer, G., Pham, M., Rao, S. S. P., Dudchenko, O., Omer, A., Mohajeri, K., Kim, S., Nichols, M. H., Davis, E. S., Gkountaroulis, D., … Rowley, M. J. (2023). Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nature Communications*, *14*(1), 3303. https://doi.org/10.1038/s41467-023-38429-1

Hecker, N., Kempynck, N., Mauduit, D., Abaffyová, D., Vandepoel, R., Dieltiens, S., Borm, L., Sarropoulos, I., González-Blas, C. B., Man, J. D., Davie, K., Leysen, E., Vandensteen, J., Moors, R., Hulselmans, G., Lim, L., Wit, J. D., Christiaens, V., Poovathingal, S., & Aerts, S. (2025). Enhancer-driven cell type comparison reveals similarities between the mammalian and bird pallium. *Science*. https://doi.org/10.1126/science.adp3957

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, *39*(3), 311–318. https://doi.org/10.1038/ng1966

Heintzmann, R., & Cremer, C. G. (1999). Laterally modulated excitation microscopy: Improvement of resolution by using a diffraction grating. *Optical Biopsies and Microscopic Techniques III*, *3568*, 185–196. https://doi.org/10.1117/12.336833

Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. *Cell*, *169*(1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007

Hsieh, T.-H. S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., & Rando, O. J. (2015). Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell*, *162*(1), 108–119. https://doi.org/10.1016/j.cell.2015.05.048

Huang, J., Li, K., Cai, W., Liu, X., Zhang, Y., Orkin, S. H., Xu, J., & Yuan, G.-C. (2018). Dissecting super-enhancer hierarchy based on chromatin interactions. *Nature Communications*, *9*(1), 943. https://doi.org/10.1038/s41467-018-03279-9

Jäger, R., Migliorini, G., Henrion, M., Kandaswamy, R., Speedy, H. E., Heindl, A., Whiffin, N., Carnicer, M. J., Broome, L., Dryden, N., Nagano, T., Schoenfelder, S., Enge, M., Yuan, Y., Taipale, J., Fraser, P., Fletcher, O., & Houlston, R. S. (2015). Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nature Communications*, *6*(1), 6178. https://doi.org/10.1038/ncomms7178

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Várnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnström, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., … Fraser, P. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, *167*(5), 1369-1384.e19. https://doi.org/10.1016/j.cell.2016.09.037

Jindal, G. A., Bantle, A. T., Solvason, J. J., Grudzien, J. L., D'Antonio-Chronowska, A., Lim, F., Le, S. H., Song, B. P., Ragsac, M. F., Klie, A., Larsen, R. O., Frazer, K. A., & Farley, E. K. (2023). Single-nucleotide variants within heart enhancers increase binding affinity and disrupt heart development. *Developmental Cell*, *58*(21), 2206-2216.e5. https://doi.org/10.1016/j.devcel.2023.09.005

Jindal, G. A., & Farley, E. K. (2021). Enhancer grammar in development, evolution, and disease: Dependencies and interplay. *Developmental Cell*, *56*(5), 575–587. https://doi.org/10.1016/j.devcel.2021.02.016

Jolma, A., Yin, Y., Nitta, K. R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., & Taipale, J. (2015). DNA-dependent formation of transcription factor

pairs alters their binding specificity. *Nature*, *527*(7578), 384–388. https://doi.org/10.1038/nature15518

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E. H., Birney, E., & Furlong, E. E. M. (2012). A Transcription Factor Collective Defines Cardiac Cell Fate and Reflects Lineage History. *Cell*, *148*(3), 473–486. https://doi.org/10.1016/j.cell.2012.01.030

Kadauke, S., Udugama, M. I., Pawlicki, J. M., Achtman, J. C., Jain, D. P., Cheng, Y., Hardison, R. C., & Blobel, G. A. (2012). Tissue-specific mitotic bookmarking by hematopoietic transcription factor GATA1. *Cell*, *150*(4), 725–737. https://doi.org/10.1016/j.cell.2012.06.038

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*, *458*(7236), 362–366. https://doi.org/10.1038/nature07667

Karr, J. P., Ferrie, J. J., Tjian, R., & Darzacq, X. (2022). The transcription factor activity gradient (TAG) model: Contemplating a contact-independent mechanism for enhancer-promoter communication. *Genes & Development*, *36*(1–2), 7–16. https://doi.org/10.1101/gad.349160.121

Kawasaki, K., & Fukaya, T. (2023). Functional coordination between transcription factor clustering and gene activity. *Molecular Cell*, *83*(10), 1605-1622.e9. https://doi.org/10.1016/j.molcel.2023.04.018

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, *10*(1), 1930. https://doi.org/10.1038/s41467-019-09982-5

Kempfer, R., & Pombo, A. (2020). Methods for mapping 3D chromosome architecture. *Nature Reviews. Genetics*, *21*(4), 207–226. https://doi.org/10.1038/s41576-019-0195-2

Kim, S., Morgunova, E., Naqvi, S., Goovaerts, S., Bader, M., Koska, M., Popov, A., Luong, C., Pogson, A., Swigut, T., Claes, P., Taipale, J., & Wysocka, J. (2024). DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell*, *187*(3), 692-711.e26. https://doi.org/10.1016/j.cell.2023.12.032

King, D. M., Hong, C. K. Y., Shepherdson, J. L., Granas, D. M., Maricque, B. B., & Cohen, B. A. (2020). Synthetic and genomic regulatory elements reveal aspects of cis-regulatory grammar in mouse embryonic stem cells. *eLife*, *9*, e41279. https://doi.org/10.7554/eLife.41279

Klar, T. A., Jakobs, S., Dyba, M., Egner, A., & Hell, S. W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(15), 8206–8210. https://doi.org/10.1073/pnas.97.15.8206

Ko, J. Y., Oh, S., & Yoo, K. H. (2017). Functional Enhancers As Master Regulators of Tissue-Specific Gene Regulation and Cancer Development. *Molecules and Cells*, *40*(3), 169–177. https://doi.org/10.14348/molcells.2017.0033

Kribelbauer-Swietek, J. F., Pushkarev, O., Gardeux, V., Faltejskova, K., Russeil, J., van Mierlo, G., & Deplancke, B. (2024). Context transcription factors establish cooperative environments and mediate enhancer communication. *Nature Genetics*, *56*(10), 2199–2212. https://doi.org/10.1038/s41588-024-01892-7

Krivega, I., & Dean, A. (2017). LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Research*, *45*(14), 8255–8268. https://doi.org/10.1093/nar/gkx433

Kvon, E. Z., Waymack, R., Gad, M., & Wunderlich, Z. (2021). Enhancer redundancy in development and disease. *Nature Reviews. Genetics*, *22*(5), 324–336. https://doi.org/10.1038/s41576-020-00311-x

Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., Garvin, T. H., Pham, Q., Harrington, A. N., Hunter, R. D., Godoy, J., Meky, E. M., Akiyama, J. A., Afzal, V., Tran, S., Escande, F., Gilbert-Dussardier, B., Jean-Marçais, N., Hudaiberdiev, S., … Pennacchio, L. A. (2020). Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell*, *180*(6), 1262-1271.e15. https://doi.org/10.1016/j.cell.2020.02.031

Lagha, M., Bothma, J. P., & Levine, M. (2012). Mechanisms of transcriptional precision in animal development. *Trends in Genetics*, *28*(8), 409–416. https://doi.org/10.1016/j.tig.2012.03.006

Lando, D., Ma, X., Cao, Y., Jartseva, A., Stevens, T. J., Boucher, W., Reynolds, N., Montibus, B., Hall, D., Lackner, A., Ragheb, R., Leeb, M., Hendrich, B. D., & Laue, E. D. (2024). Enhancer-promoter interactions are reconfigured through the formation of long-range multiway hubs as mouse ES cells exit pluripotency. *Molecular Cell*, *84*(8), 1406-1421.e8. https://doi.org/10.1016/j.molcel.2024.02.015

Lettice, L. A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P. S., Gautier, P., Morrison, H., Fantes, J., Hill, R. E., & FitzPatrick, D. R. (2011). Enhancer-adoption as a mechanism of human developmental disease. *Human Mutation*, *32*(12), 1492–1499. https://doi.org/10.1002/humu.21615

Lettice, L. A., Hill, A. E., Devenney, P. S., & Hill, R. E. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Human Molecular Genetics*, *17*(7), 978–985. https://doi.org/10.1093/hmg/ddm370

Levo, M., Raimundo, J., Bing, X. Y., Sisco, Z., Batut, P. J., Ryabichko, S., Gregor, T., & Levine, M. S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. *Nature*, *605*(7911), 754–760. https://doi.org/10.1038/s41586-022-04680-7

Li, J., Zhang, P., Xi, X., Liu, L., Wei, L., & Wang, X. (2025). Modeling and designing enhancers by introducing and harnessing transcription factor binding units. *Nature Communications*, *16*(1), 1469. https://doi.org/10.1038/s41467-025-56749-2

Li, Q., Zhang, M., Han, H., Rohde, A., & Stamatoyannopoulos, G. (2002). Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Research*, *30*(11), 2484–2491. https://doi.org/10.1093/nar/30.11.2484

Li, X., & Levine, M. (2024). What are tethering elements? *Current Opinion in Genetics & Development*, *84*, 102151. https://doi.org/10.1016/j.gde.2023.102151

Lieberman-Aiden, E., Berkum, N. L. van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*. https://doi.org/10.1126/science.1181369

Lim, F., Solvason, J. J., Ryan, G. E., Le, S. H., Jindal, G. A., Steffen, P., Jandu, S. K., & Farley, E. K. (2024). Affinity-optimizing enhancer variants disrupt development. *Nature*, *626*(7997), 151–159. https://doi.org/10.1038/s41586-023-06922-8

Liu, Y., Pelham-Webb, B., Di Giammartino, D. C., Li, J., Kim, D., Kita, K., Saiz, N., Garg, V., Doane, A., Giannakakou, P., Hadjantonakis, A.-K., Elemento, O., & Apostolou, E. (2017). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors

in Pluripotent Stem Cells. *Cell Reports*, *19*(7), 1283–1293. https://doi.org/10.1016/j.celrep.2017.04.067

Loubiere, V., de Almeida, B. P., Pagani, M., & Stark, A. (2024). Developmental and housekeeping transcriptional programs display distinct modes of enhancer-enhancer cooperativity in Drosophila. *Nature Communications*, *15*(1), 8584. https://doi.org/10.1038/s41467-024-52921-2

Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., … Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, *161*(5), 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Lupo, O., Kumar, D. K., Livne, R., Chappleboim, M., Levy, I., & Barkai, N. (2023). The architecture of binding cooperativity between densely bound transcription factors. *Cell Systems*, *14*(9), 732-745.e5. https://doi.org/10.1016/j.cels.2023.06.010

Mann, I. K., Chatterjee, R., Zhao, J., He, X., Weirauch, M. T., Hughes, T. R., & Vinson, C. (2013). CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. *Genome Research*, *23*(6), 988–997. https://doi.org/10.1101/gr.146654.112

Martinez-Ara, M., Comoglio, F., & van Steensel, B. (2024). Large-scale analysis of the integration of enhancer-enhancer signals by promoters. *eLife*, *12*, RP91994. https://doi.org/10.7554/eLife.91994

McDaniel, S. L., Gibson, T. J., Schulz, K. N., Fernandez Garcia, M., Nevil, M., Jain, S. U., Lewis, P. W., Zaret, K. S., & Harrison, M. M. (2019). Continued Activity of the Pioneer

Factor Zelda Is Required to Drive Zygotic Genome Activation. *Molecular Cell*, *74*(1), 185-195.e4. https://doi.org/10.1016/j.molcel.2019.01.014

McKay, D. J., Stutzman, A. V., & Dowen, J. M. (2020). Advancements in mapping 3D genome architecture. *Methods*, *170*, 75–81. https://doi.org/10.1016/j.ymeth.2019.06.002

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C. G., Kinney, J. B., Kellis, M., Lander, E. S., & Mikkelsen, T. S. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature Biotechnology*, *30*(3), 271–277. https://doi.org/10.1038/nbt.2137

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S. W., Andrews, S., Grey, W., Ewels, P. A., Herman, B., Happe, S., Higgs, A., LeProust, E., Follows, G. A., Fraser, P., Luscombe, N. M., & Osborne, C. S. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature Genetics*, *47*(6), 598–606. https://doi.org/10.1038/ng.3286

Monahan, K., Horta, A., & Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature*, *565*(7740), 448–453. https://doi.org/10.1038/s41586-018-0845-0

Mumbach, M. R., Rubin, A. J., Flynn, R. A., Dai, C., Khavari, P. A., Greenleaf, W. J., & Chang, H. Y. (2016). HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods*, *13*(11), 919–922. https://doi.org/10.1038/nmeth.3999

Nagano, M., Hu, B., Ogata, K., Umemura, F., Ishikura, Y., Suzuki, S., Katsifis, C. C., Yoshinaga, M., Litos, G., Nagasaka, K., Tang, W., Nosaka, Y., Sasada, H., Wang, H., Kondo, D., Katou, Y., Mizuta, K., Yabuta, Y., Ohta, H., … Saitou, M. (2025). The mitotic STAG3–cohesin complex shapes male germline nucleome. *Nature Structural & Molecular Biology*, 1–16. https://doi.org/10.1038/s41594-025-01647-w

Neumayr, C., Pagani, M., Stark, A., & Arnold, C. D. (2019). STARR-seq and UMI-STARR-seq: Assessing Enhancer Activities for Genome-Wide-, High-, and Low-Complexity Candidate Libraries. *Current Protocols in Molecular Biology*, *128*(1), e105. https://doi.org/10.1002/cpmb.105

Nguyen, H. Q., Chattoraj, S., Castillo, D., Nguyen, S. C., Nir, G., Lioutas, A., Hershberg, E. A., Martins, N. M. C., Reginato, P. L., Hannan, M., Beliveau, B. J., Church, G. M., Daugharthy, E. R., Marti-Renom, M. A., & Wu, C. -ting. (2020). 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nature Methods*, *17*(8), 822–832. https://doi.org/10.1038/s41592-020-0890-0

Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3), 247–257. https://doi.org/10.1016/s0092-8674(00)81656-6

Oudelaar, A. M., Davies, J. O. J., Hanssen, L. L. P., Telenius, J. M., Schwessinger, R., Liu, Y., Brown, J. M., Downes, D. J., Chiariello, A. M., Bianco, S., Nicodemi, M., Buckle, V. J., Dekker, J., Higgs, D. R., & Hughes, J. R. (2018). Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nature Genetics*, *50*(12), 1744–1751. https://doi.org/10.1038/s41588-018-0253-2

Oudelaar, A. M., Harrold, C. L., Hanssen, L. L. P., Telenius, J. M., Higgs, D. R., & Hughes, J. R. (2019). A revised model for promoter competition based on multi-way chromatin interactions at the α-globin locus. *Nature Communications*, *10*(1), 5412. https://doi.org/10.1038/s41467-019-13404-x

Pachano, T., Sánchez-Gaya, V., Ealo, T., Mariner-Faulí, M., Bleckwehl, T., Asenjo, H. G., Respuela, P., Cruz-Molina, S., Muñoz-San Martín, M., Haro, E., van IJcken, W. F. J., Landeira, D., & Rada-Iglesias, A. (2021). Orphan CpG islands amplify poised enhancer

regulatory activity and determine target gene responsiveness. *Nature Genetics*, *53*(7), 1036–1049. https://doi.org/10.1038/s41588-021-00888-x

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., Lee, C., Andrie, J. M., Lee, S.-I., Cooper, G. M., Ahituv, N., Pennacchio, L. A., & Shendure, J. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. *Nature Biotechnology*, *30*(3), 265–270. https://doi.org/10.1038/nbt.2136

Perry, M. W., Boettiger, A. N., Bothma, J. P., & Levine, M. (2010). Shadow enhancers foster robustness of Drosophila gastrulation. *Current Biology: CB*, *20*(17), 1562–1567. https://doi.org/10.1016/j.cub.2010.07.043

Prescott, S. L., Srinivasan, R., Marchetto, M. C., Grishina, I., Narvaiza, I., Selleri, L., Gage, F. H., Swigut, T., & Wysocka, J. (2015). Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell*, *163*(1), 68–83. https://doi.org/10.1016/j.cell.2015.08.036

Quang, D., & Xie, X. (2016). DanQ: A hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Research*, *44*(11), e107. https://doi.org/10.1093/nar/gkw226

Rao, S., Ahmad, K., & Ramachandran, S. (2021). Cooperative binding between distant transcription factors is a hallmark of active enhancers. *Molecular Cell*, *81*(8), 1651-1665.e4. https://doi.org/10.1016/j.molcel.2021.02.014

Reiter, F., de Almeida, B. P., & Stark, A. (2023). Enhancers display constrained sequence flexibility and context-specific modulation of motif function. *Genome Research*, *33*(3), 346–358. https://doi.org/10.1101/gr.277246.122

Reiter, F., Wienerroither, S., & Stark, A. (2017). Combinatorial function of transcription factors and cofactors. *Current Opinion in Genetics & Development*, *43*, 73–81. https://doi.org/10.1016/j.gde.2016.12.007

Richter, W. F., Nayak, S., Iwasa, J., & Taatjes, D. J. (2022). The Mediator complex as a master regulator of transcription by RNA polymerase II. *Nature Reviews. Molecular Cell Biology*, *23*(11), 732–749. https://doi.org/10.1038/s41580-022-00498-3

Rinzema, N. J., Sofiadis, K., Tjalsma, S. J. D., Verstegen, M. J. A. M., Oz, Y., Valdes-Quezada, C., Felder, A.-K., Filipovska, T., van der Elst, S., de Andrade dos Ramos, Z., Han, R., Krijger, P. H. L., & de Laat, W. (2022). Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nature Structural & Molecular Biology*, *29*(6), 563–574. https://doi.org/10.1038/s41594-022-00787-7

Rodríguez-Carballo, E., Lopez-Delisle, L., Zhan, Y., Fabre, P. J., Beccari, L., El-Idrissi, I., Huynh, T. H. N., Ozadam, H., Dekker, J., & Duboule, D. (2017). The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes & Development*, *31*(22), 2264–2281. https://doi.org/10.1101/gad.307769.117

Ryan, R. J. H., Drier, Y., Whitton, H., Cotton, M. J., Kaur, J., Issner, R., Gillespie, S., Epstein, C. B., Nardi, V., Sohani, A. R., Hochberg, E. P., & Bernstein, B. E. (2015). Detection of Enhancer-Associated Rearrangements Reveals Mechanisms of Oncogene Dysregulation in B-cell Lymphoma. *Cancer Discovery*, *5*(10), 1058–1071. https://doi.org/10.1158/2159-8290.CD-15-0370

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)*, *361*(6400), eaar3958. https://doi.org/10.1126/science.aar3958

Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C. O., Cramer, P., Kivioja, T., & Taipale, J. (2022). Sequence determinants of human gene regulatory elements. *Nature Genetics*, *54*(3), 283–294. https://doi.org/10.1038/s41588-021-01009-4

Sanyal, A., Lajoie, B. R., Jain, G., & Dekker, J. (2012). The long-range interaction landscape of gene promoters. *Nature*, *489*(7414), 109–113. https://doi.org/10.1038/nature11279

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews. Genetics*, *20*(8), 437–455. https://doi.org/10.1038/s41576-019-0128-0

Scholes, C., Biette, K. M., Harden, T. T., & DePace, A. H. (2019). Signal Integration by Shadow Enhancers and Enhancer Duplications Varies across the Drosophila Embryo. *Cell Reports*, *26*(9), 2407-2418.e5. https://doi.org/10.1016/j.celrep.2019.01.115

Schones, D. E., Cui, K., Cuddapah, S., Roh, T.-Y., Barski, A., Wang, Z., Wei, G., & Zhao, K. (2008). Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell*, *132*(5), 887–898. https://doi.org/10.1016/j.cell.2008.02.022

Schreiber, J., Lorbeer, F. K., Heinzl, M., Lu, Y. Y., Stark, A., & Noble, W. S. (2025). *Programmatic design and editing of cis-regulatory elements* (p. 2025.04.22.650035). bioRxiv. https://doi.org/10.1101/2025.04.22.650035

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N. A., Huber, W., Haering, C. H., Mirny, L., & Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, *551*(7678), 51–56. https://doi.org/10.1038/nature24281

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., & Cavalli, G. (2012). Three-dimensional folding and functional organization

principles of the Drosophila genome. *Cell*, *148*(3), 458–472. https://doi.org/10.1016/j.cell.2012.01.010

Shlyueva, D., Stampfel, G., & Stark, A. (2014). Transcriptional enhancers: From properties to genome-wide predictions. *Nature Reviews. Genetics*, *15*(4), 272–286. https://doi.org/10.1038/nrg3682

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., & de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genetics*, *38*(11), 1348–1354. https://doi.org/10.1038/ng1896

Singh, G., Mullany, S., Moorthy, S. D., Zhang, R., Mehdi, T., Tian, R., Duncan, A. G., Moses, A. M., & Mitchell, J. A. (2021). A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. *Genome Research*, *31*(4), 564–575. https://doi.org/10.1101/gr.272468.120

Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols*, *13*(5), 1006–1019. https://doi.org/10.1038/nprot.2018.015

Slattery, M., Zhou, T., Yang, L., Machado, A. C. D., Gordân, R., & Rohs, R. (2014). Absence of a simple code: How transcription factors read the genome. *Trends in Biochemical Sciences*, *39*(9), 381–399. https://doi.org/10.1016/j.tibs.2014.07.002

Snetkova, V., Ypsilanti, A. R., Akiyama, J. A., Mannion, B. J., Plajzer-Frick, I., Novak, C. S., Harrington, A. N., Pham, Q. T., Kato, M., Zhu, Y., Godoy, J., Meky, E., Hunter, R. D., Shi, M., Kvon, E. Z., Afzal, V., Tran, S., Rubenstein, J. L. R., Visel, A., … Dickel, D. E. (2021). Ultraconserved enhancer function does not require perfect sequence conservation. *Nature Genetics*, *53*(4), 521–528. https://doi.org/10.1038/s41588-021-00812-3

Socha, M., Sowińska-Seidler, A., Melo, U. S., Kragesteen, B. K., Franke, M., Heinrich, V., Schöpflin, R., Nagel, I., Gruchy, N., Mundlos, S., Sreenivasan, V. K. A., López, C., Vingron, M., Bukowska-Olech, E., Spielmann, M., & Jamsheer, A. (2021). Position effects at the FGF8 locus are associated with femoral hypoplasia. *The American Journal of Human Genetics*, *108*(9), 1725–1734. https://doi.org/10.1016/j.ajhg.2021.08.001

Song, B. P., Ragsac, M. F., Tellez, K., Jindal, G. A., Grudzien, J. L., Le, S. H., & Farley, E. K. (2023). Diverse logics and grammar encode notochord enhancers. *Cell Reports*, *42*(2), 112052. https://doi.org/10.1016/j.celrep.2023.112052

Soutourina, J. (2018). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology*, *19*(4), 262–274. https://doi.org/10.1038/nrm.2017.115

Speicher, M. R., Gwyn Ballard, S., & Ward, D. C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics*, *12*(4), 368–375. https://doi.org/10.1038/ng0496-368

Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: From enhancer binding to developmental control. *Nature Reviews Genetics*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

Stoeber, S., Godin, H., Xu, C., & Bai, L. (2024). Pioneer factors: Nature or nurture? *Critical Reviews in Biochemistry and Molecular Biology*, *59*(3–4), 139–153. https://doi.org/10.1080/10409238.2024.2355885

Szabo, Q., Donjon, A., Jerković, I., Papadopoulos, G. L., Cheutin, T., Bonev, B., Nora, E. P., Bruneau, B. G., Bantignies, F., & Cavalli, G. (2020). Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics*, *52*(11), 1151–1157. https://doi.org/10.1038/s41588-020-00716-8

Taskiran, I. I., Spanier, K. I., Dickmänken, H., Kempynck, N., Pančíková, A., Ekşi, E. C., Hulselmans, G., Ismail, J. N., Theunis, K., Vandepoel, R., Christiaens, V., Mauduit, D.,

& Aerts, S. (2024). Cell-type-directed design of synthetic enhancers. *Nature*, *626*(7997), 212–220. https://doi.org/10.1038/s41586-023-06936-2

Thomas, H. F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., Wysocka, J., & Buecker, C. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Molecular Cell*, *81*(5), 969-982.e13. https://doi.org/10.1016/j.molcel.2020.12.047

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T. K., Diegel, M., Dunn, D., Ebersol, A. K., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232

Tournamille, C., Colin, Y., Cartron, J. P., & Le Van Kim, C. (1995). Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy–negative individuals. *Nature Genetics*, *10*(2), 224–228. https://doi.org/10.1038/ng0695-224

Tsai, A., Muthusamy, A. K., Alves, M. R., Lavis, L. D., Singer, R. H., Stern, D. L., & Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. *eLife*, *6*, e28975. https://doi.org/10.7554/eLife.28975

Tünnermann, J., Roth, G., Cramard, J., & Giorgetti, L. (2025). *Enhancer control of promoter activity and variability via frequency modulation of clustered transcriptional bursts* (p. 2025.03.26.645410). bioRxiv. https://doi.org/10.1101/2025.03.26.645410

Uyehara, C. M., & Apostolou, E. (2023). 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Reports*, *42*(4), 112068. https://doi.org/10.1016/j.celrep.2023.112068

Vaishnav, E. D., de Boer, C. G., Molinet, J., Yassour, M., Fan, L., Adiconis, X., Thompson, D. A., Levin, J. Z., Cubillos, F. A., & Regev, A. (2022). The evolution, evolvability and engineering of gene regulatory DNA. *Nature*, *603*(7901), 455–463. https://doi.org/10.1038/s41586-022-04506-6

van Arensbergen, J., Pagie, L., FitzPatrick, V. D., de Haas, M., Baltissen, M. P., Comoglio, F., van der Weide, R. H., Teunissen, H., Võsa, U., Franke, L., de Wit, E., Vermeulen, M., Bussemaker, H. J., & van Steensel, B. (2019). High-throughput identification of human SNPs affecting regulatory element activity. *Nature Genetics*, *51*(7), 1160–1169. https://doi.org/10.1038/s41588-019-0455-2

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., & Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–858. https://doi.org/10.1038/nature07730

Wang, J., Yu, H., Ma, Q., Zeng, P., Wu, D., Hou, Y., Liu, X., Jia, L., Sun, J., Chen, Y., Guallar, D., Fidalgo, M., Chen, J., Yu, Y., Jiang, S., Li, F., Zhao, C., Huang, X., Wang, J., … Ding, J. (2021). Phase separation of OCT4 controls TAD reorganization to promote cell fate transitions. *Cell Stem Cell*, *28*(10), 1868-1883.e11. https://doi.org/10.1016/j.stem.2021.04.023

Waszak, S. M., Delaneau, O., Gschwind, A. R., Kilpinen, H., Raghav, S. K., Witwicki, R. M., Orioli, A., Wiederkehr, M., Panousis, N. I., Yurovsky, A., Romano-Palumbo, L., Planchon, A., Bielser, D., Padioleau, I., Udin, G., Thurnheer, S., Hacker, D., Hernandez, N., Reymond, A., … Dermitzakis, E. T. (2015). Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell*, *162*(5), 1039–1050. https://doi.org/10.1016/j.cell.2015.08.001

Weintraub, A. S., Li, C. H., Zamudio, A. V., Sigova, A. A., Hannett, N. M., Day, D. S., Abraham, B. J., Cohen, M. A., Nabet, B., Buckley, D. L., Guo, Y. E., Hnisz, D., Jaenisch, R., Bradner, J. E., Gray, N. S., & Young, R. A. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*, *171*(7), 1573-1588.e28. https://doi.org/10.1016/j.cell.2017.11.008

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2), 307–319. https://doi.org/10.1016/j.cell.2013.03.035

Xie, Z., Sokolov, I., Osmala, M., Yue, X., Bower, G., Pett, J. P., Chen, Y., Wang, K., Cavga, A. D., Popov, A., Teichmann, S. A., Morgunova, E., Kvon, E. Z., Yin, Y., & Taipale, J. (2025). DNA-guided transcription factor interactions extend human gene regulatory code. *Nature*, *641*(8065), 1329–1338. https://doi.org/10.1038/s41586-025-08844-z

Yang, J. H., & Hansen, A. S. (2024). Enhancer selectivity in space and time: From enhancer–promoter interactions to promoter activation. *Nature Reviews Molecular Cell Biology*, *25*(7), 574–591. https://doi.org/10.1038/s41580-024-00710-6

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*. https://doi.org/10.1126/science.aaj2239

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2015). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature*, *518*(7540), 556–559. https://doi.org/10.1038/nature13994

Zhang, Q., Zhong, W., Zhu, G., Cheng, L., Yin, C., Deng, L., Yang, Y., Zhang, Z., Shen, J., Fu, T., Zhu, J.-K., & Zhao, L. (2024). aChIP is an efficient and sensitive ChIP-seq technique for economically important plant organs. *Nature Plants*, *10*(9), 1317–1329. https://doi.org/10.1038/s41477-024-01743-7

Zhao, Y., Vartak, S. V., Conte, A., Wang, X., Garcia, D. A., Stevens, E., Kyoung Jung, S., Kieffer-Kwon, K.-R., Vian, L., Stodola, T., Moris, F., Chopp, L., Preite, S., Schwartzberg, P. L., Kulinski, J. M., Olivera, A., Harly, C., Bhandoola, A., Heuston, E. F., … Casellas, R. (2022). "Stripe" transcription factors provide accessibility to co-binding partners in mammalian genomes. *Molecular Cell*, *82*(18), 3398-3411.e11. https://doi.org/10.1016/j.molcel.2022.06.029

Zheng, L., & Wang, W. (2022). Regulation associated modules reflect 3D genome modularity associated with chromatin activity. *Nature Communications*, *13*(1), 5281. https://doi.org/10.1038/s41467-022-32911-y

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., Collura, F., & Mitchell, J. A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes & Development*, *28*(24), 2699–2711. https://doi.org/10.1101/gad.248526.114

Zhou, P., Gu, F., Zhang, L., Akerberg, B. N., Ma, Q., Li, K., He, A., Lin, Z., Stevens, S. M., Zhou, B., & Pu, W. T. (2017). Mapping cell type-specific transcriptional enhancers using high affinity, lineage-specific Ep300 bioChIP-seq. *eLife*, *6*, e22039. https://doi.org/10.7554/eLife.22039

Zhou, Z., Li, A., Zhang, J., Yu, H., Ozer, A., & Lis, J. T. (2025). *Robust regulatory interplay of enhancers, facilitators, and promoters in a native chromatin context* (p. 2025.07.07.663560). bioRxiv. https://doi.org/10.1101/2025.07.07.663560

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through enhancer–promoter interactions. *Nature*, *604*(7906), 571–577. https://doi.org/10.1038/s41586-022-04570-y