UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC)

CNRS UMR 7104 – Inserm U 1258

THÈSE présentée par :

Angeliki PLATANIA

soutenue le : **23 juillet 2021**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Aspects moléculaires et cellulaire de la biologie

Assessing the links between enhancer-promoter proximity, local chromatin dynamics and gene activity

THÈSE dirigée par :
Dr. SEXTON Thomas CR, IGBMC, Strasbourg, France

RAPPORTEURS :
Pr. BICKMORE Wendy Professor, Institute of Genetics and Cancer. MRC Human Genetics Unit, University of Edinburgh, Scotland
Dr. GIORGETTI Luca DR, Friedrich Miescher Institute for Biomedical Research (FMI), Basel, Switzerland

AUTRES MEMBRES DU JURY :
Dr. SUMARA Isabela DR, IGBMC, Strasbourg, France

INVITÉS :
Pr. BYSTRICKY Kerstin Professor, Centre de Biologie Intégrative (CBI), Toulouse, France

“I don’t know where I’m going from here,
but I promise it won’t be boring”
David Bowie

To my mother..

Table of Contents

Acknowledgments

There are several people who have contributed to the conception and execution of my thesis and I would like to thank them.

First I wish to thank my supervisor, Tom Sexton, for accepting me in his laboratory and giving me the chance to work on such a challenging project. His motivation, crazy ideas and enthusiasm were an inspiration for me. Thank you for the guidance throughout this journey, for believing in me and helping me improve as a scientist. Except from a great scientist, Tom is a very cool and nice person. He has always been there when things went bad, helping me overcome several challenges and reminding me that the priority is to take care of my health and my well-being. I would do it all over again and I am extremely grateful for that!

Besides Tom, I would like to thank the rest of my committee members for agreeing to evaluate my work and attend my thesis presentation in the middle of summer. Pr. Wendy Bickmore I still cannot believe that you accepted to be a part of my jury. You have been an inspiration to me and it will be a great honor to discuss my work with you. Dr. Luca Giorgetti I had the chance to meet you in several conferences and I was always impressed after our short talks during poster sessions. I am looking forward to discuss my results with you. Dr. Isabela Sumara thank you for being so kind and postponing your holidays in order to attend my presentation. I am looking forward for your comments. Pr. Kerstin Bystricky thank you for discovering the ANCHOR system, for all the valuable advice that you gave me since the very beginning of my thesis and for being part of my TAC committee. Last, I would also like to thank Dr. Evi Soutoglou, who was also part of my TAC, and for all the great discussions that we had over the years.

I am grateful to all the Sexton lab members, former and current. Cathie, ma collègue d'amour, thank you for doing your magics with your golden hands during the last and most critical years of my thesis. You are adorable and it was such a pleasure working with you. Nezih, we went through this thesis together and my days wouldn't have been the same without you. Thank you for all the great discussions, the 4C brainstorming and the delicious Turkish dishes that you shared with me! A special mention to Gui for his patience and contribution to the project. Thank you for all the effort that you put on developing the GP-Tool and analyzing my data. Natalia,

my module and Sox2 mate, thank you for sharing your ideas with me and for bringing breakfast and chocolates to cheer us up. Dominique and Manon you did an excellent job while establishing the ANCHOR system in the lab and I am very grateful for that. Anne, Audrey, Yousra and Sanjay thank you for welcoming me in the team the first months of the Ph.D. Of course, the fresh blood in the lab Kubra and Thomas, thank you for reading and correcting my French summaries. I hope that we will have the opportunity to work together soon! Moreover, I wish to thank all the people from the IGBMC facilities, and especially Marion and Amelie that took care of my cell lines, as well as Erwan, that spent hours with me at the microscope trying to see the ANCHOR spots and helping me setup my experiments. Last, a special thanks goes to Tineke Lenstra and Marit de Kort for sharing their expertise on the MS2 system and to Jennifer Mitchell for all the great advice and reagents that shared with me throughout the years.

To the amazing group of talented girls that started their Ph.D. with me, thank you so much for everything! I consider you as truly good friends. A special mention goes to Sirine, who became one of my favorite people. I cannot imagine how my life would have been without you. Thank you for your support (inside and outside the lab), kindness, love, energy and all your DIY recipes. You are an amazing and inspiring woman. Karima, my Mediterranean friend, thank you for all the discussions about spiritual life, for being there when I needed you, for the delicious dishes and for sharing with me your tisane expertise. Another special thanks goes to Diana and Kamar. You are two talented goddesses and I am really happy to met you. Keep being so amazing!

I feel extremely fortunate that during these years I had the chance to meet some amazing people outside the lab. Lina, ζουζούνα, I lived some of my favorite moments ever with you. You are one of the best bar-, concert-, party-, running-, coffee- mates. Thank you for being you, simply. Katerina, τσιτσίνι, thank you for being so sweet and supporting me during all these years. Charlie, my first French friend, thank you for all the franco-grec discussions and the great time that we had together. I could not neglect thanking my dear friends Julien and Florine for all the moments that we lived together, for hosting me at their beautiful house and for preparing my birthday cakes the last two years. Special thanks goes to Gab, who never said no to the overloaded plates that I served him, as well as Marie, Stephane, Nico, Yiannis and Liza for all the soirées and the wonderful time that we spent together.

Now, I wish to thank several people that are by my side for many years and without whom this thesis would not have happened. I could write a full page expressing my gratitude for each one individually but I have to keep it short. Faye, I am so proud to be your best friend. Thank you for being next to me at every single step of my life. You have the best smile and best sense of humor. Vaso, I feel so lucky for having you in my life since day zero – literally. Your love and your support gave me strength at the most difficult moments. Adam, what can I say for you? My life during the university and Ph.D. years wouldn't have been the same without you. I wouldn't exchange you for 1 million camels! Thank you for all the holidays that we spent together, your critical comments on my project and for being so amazing. Giorgo (aka Kavvadia), you are the brightest star. Thank you for always being there, the right moment with the right thing to say. I cannot find the right words to express my gratitude to Konstantina, Katerina, Elina, Dimitra and Ioanna. Since 2010, we have lived EVERYTHING together. Thank you for being so wonderful, funny, openminded and supportive. Thank you for your true love and for accepting me as I am, with my good and bad moods.

One of the most important people in my life the last years is Julien. You showed me a whole new side of life that I had never imagined before and I am grateful for that. Thank you for being next to me, loving and supporting me every single day (yes, during the writing period as well!). Thank you for all the things and places that you made me discover. The best is yet to come! Of course, I want to thank Simone, Marc and Loic, for welcoming me into their family, for all the great diners that have shared with me and for their support.

I would like to express my gratitude to my parents, Eugenia and Antonis as well as my brother Stratos and my sister Maura. Thank you for your unconditional love, for all the great moments that you offered me from my childhood until today and for all the sacrifices that you have done for me. Without you I wouldn't have been where and who I am today. Mother, you are the strongest and sweetest woman in the world. This thesis project and all that are about to come are dedicated to you! Antonis, I am so grateful for having you in my life. Thank you for your love and for being next to me during the most happy, important, difficult times of my life. Stratos and Maura, my beloved siblings, I cannot imagine how my life would have been without you. You are the biggest loves of my life. We might be living apart during the last years, but every time that the three of us are together is as if we haven't spent one single day separately.

I would also like to acknowledge my extended family, my aunts (Sofia and Gioula), uncles (Vangelis and Thanasis) and of course my beloved cousins (Dionisis, Vasilis, Athina, Angeliki, Dimitra, Mirto) for their love, mentality, energy and support. I am extremely happy and proud for being part of this big and noisy Greek family. Last but not least, I wish to thank my grandparents who do not speak English, so the next lines are written in Greek just for them: Γιαγιάκα μου και παππούλη μου σας λατρεύω και μου λείπετε πολύ. Χωρίς εσάς δίπλα μου δεν θα κατάφερνα τίποτα. Ανυπομονώ να γυρίσω πίσω και να σας δω από κοντά.

Abbreviations

3C	Chromosome conformation capture
4C	Circular chromosome conformation capture
Cas13	CRISPR associated protein 13
Cas9	CRISPR associated protein 9
CDK9	Cyclin-dependent kinase 9
ChIP	Chromatin immunoprecipitation
CRISPR	Clustered regularly interspaced short palindromic repeats
CRM	*cis*-regulatory modules
CTD	C-terminal domain of RNA Pol II
dCas9	Catalytically dead Cas9
DRB	5,6-dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB sensitivity-inducing factor
eRNA	Enhancer RNA
ERα	Estrogen receptor α
FCS	Fluorescence correlation spectroscopy
FISH	Fluorescence *in situ* hybridization
FRAP	Fluorescence recovery after photobleaching
GTF	General transcription factor
HA	Homology arm
Hi-C	High-throughput chromosome conformation capture
HMG	High-mobility group
IDR	Intrinsically disordered regions
LCR	Locus control region
LIF	Leukemia inhibitory factor
LLPS	Liquid-liquid phase-separation

mESCs	Mouse embryonic stem cells
mRNA	Messenger RNA
NELF	Negative elongation factor
NPC	Neural progenitor cell
P-TEFb	Positive transcription elongation factor b
PAM	Protospacer adjacent motif
PIC	Pre-initiation complex
Pol II	RNA Polymerase II
PRS	Pierre Robin sequence
RF	Regulatory factor
SCR	*Sox2* control region
SDHA	Succinate dehydrogenase complex flavoprotein subunit A
SE	Super-enhancer
siRNA	Short interfering RNA
smFISH	Single-molecule Fluorescence *in situ* hybridization
SMT	Single-molecule tracking
SNP	Single nucleotide polymorphism
SRR	*Sox2* regulatory region
SRY	Sex-determining Region of Y chromosome
SV40	Simian virus 40
TAD	Topologically associating domain
TF	Transcription factor
TSS	Transcription start site
ZPA	Zone of polarizing activity
ZRS	ZPA regulatory sequence

List of figures

List of tables

Summary

Genomes must be highly compacted within cell nuclei, but in a manner ensuring accessibility of appropriate genes to the transcriptional machinery, and accommodating marked gene program changes during development and cell reprogramming. While we have a deep view of genome organization in space from microscopy and biochemical approaches on large populations of fixed cells, fundamental questions about chromatin dynamics in living cells are completely unaddressed, despite growing appreciation that dynamics likely play key roles in the regulation of gene expression. In this study I monitored enhancer-promoter communication in real time, coupled to MS2-labeled nascent mRNA, and characterized chromatin dynamics at these loci. To address these questions, I incorporated the *parS*-ParB (also termed ANCHOR) system, originally developed in yeast, into mouse embryonic stem cells (ESCs) to label the endogenous *Sox2* promoter and its distal (>100 kb downstream) enhancer (SCR, for *Sox2* control region) for single-cell live tracking.

The real time-behavior of these elements was dissected in wild-type ESCs, during early differentiation stages when *Sox2* expression is lost, in mutants of critical regulatory elements within the SCR, and in ESCs in response to pharmacological inhibition of transcriptional initiation or elongation. I found that *Sox2* and SCR are frequently in close proximity, and significantly closer than control regions, but also observed an apparent uncoupling of separation distances with transcriptional firing. Rather, transcriptional activity associated with changes in local chromatin dynamics. Both the gene and the enhancer are more constrained than the control regions. Further, perturbation of *Sox2* expression upon differentiation, SCR mutation and inhibition of transcriptional initiation by RNA Pol II resulted in decreased apparent diffusive speed at *Sox2* promoter. Although alterations in local chromatin dynamics correlated with gene activity, loci participating in active transcriptional bursts were not found to have different chromatin dynamics to non-transcribing alleles within the same experiment. I propose that “poised” loci have similar dynamics to transcribing loci, and that alterations in mobility accompany a transition from a refractory state to a poised, transcriptionally competent state, to be assessed in future experiments. Collectively, these findings bolster the importance of assessing chromatin dynamics at transcription sites as a means of understanding the nuclear microenvironment and regulatory principles.

Summary in French

Les génomes doivent être très compactés à l'intérieur des noyaux cellulaires, mais d'une manière qui puisse garantir l'accessibilité des gènes cibles à la machinerie transcriptionnelle et qui permette des changements importants du programme génétique au cours du développement et de la reprogrammation cellulaire. La topologie de la chromatine présente un intérêt particulier pour le contrôle de l'expression des gènes, car de nombreux gènes sont régulés par des éléments activateurs, appelés amplificateurs ou enhancers. Les enhancers peuvent être localisés à des distances de l'ordre de la mégabase de leurs gènes cibles ou dans les introns de gènes dont ils ne régulent pas l'expression (Shen *et al*., 2012; Thurman *et al*., 2012). Des variantes de la technique de capture de la conformation des chromosomes (3C) (Dekker, 2006; Denker & De Laat, 2016) ont montré que les enhancers se trouvent à proximité physique directe de leurs gènes cibles, vraisemblablement grâce à la formation de boucles chromatiniennes. Bien que l'on sache peu de choses sur la manière dont les enhancers régulent l'expression des gènes, il est suggéré que ces boucles augmentent la concentration locale de facteurs de transcription liés aux séquences promotrices et aux enhancers, fournissant ainsi un environnement permettant la transcription (Tolhuis, Palstra, Splinter, Grosveld, & De Laat, 2002). Ainsi, des études récentes suggèrent que jusqu'à la moitié des gènes ne dépendent pas de leur enhancer le plus proche (Lettice *et al*., 2003). De manière plus frappante, un grand nombre de contacts entre enhancers et promoteurs semblent être établis avant l'activation du gène, suggérant qu'un signal inconnu serait nécessaire dans certains contextes pour permettre la formation d'une boucle chromatinienne favorisant la transcription (Palstra *et al*., 2003). Actuellement, nous disposons d'une vue approfondie de l'organisation du génome dans l'espace grâce à la microscopie et aux approches biochimiques sur de grandes populations de cellules fixées. Néanmoins, alors qu'il est de plus en plus suggéré que la dynamique de la chromatine joue des rôles clés dans la régulation de l'expression des gènes, les questions fondamentales sur la dynamique de la chromatine dans les cellules vivantes sont encore très peu étudiées.

Le gène *Sox2* est impliqué dans le maintien de la pluripotence et dans l'autorenouvèlement des cellules souches embryonnaires (CSE), ainsi que dans la détermination du destin cellulaire, la reprogrammation cellulaire, la régénération des tissus adultes et le cancer dans de multiples tissus tels que le poumon et le sein (Takahashi & Yamanaka, 2006). Une étude récente sur le

locus *Sox2* a montré qu'une région distale (région de contrôle de *Sox2*, SCR) située >100 kb en aval de *Sox2*, forme une grande boucle chromatinienne spécifiquement requise pour la transcription de *Sox2* dans les CSE (Zhou *et al*., 2014). Néanmoins, dans le cadre de cette boucle chromatinienne, aucune information n'est disponible quant à la dynamique de la chromatine, la variabilité de cellule à cellule, la façon dont la formation de cette boucle peut influencer la régulation de *Sox2*.

Dans cette étude, j'ai suivi en temps réel la communication enhancer-promoteur et caractérisé la dynamique de la chromatine à ces loci. Pour réaliser cela, j'ai incorporé le système parS-ParB (également appelé ANCHOR) (Saad *et al*., 2014), initialement développé chez la levure, dans des CSE de souris afin de marquer le promoteur endogène de *Sox2* et son enhancer distal (SCR) pour suivre en direct leur localisation. En parallèle, j'ai construit une lignée témoin dans laquelle les séquences ANCHOR sont décalées de 60 kb afin qu'elles puissent conserver la même distance génomique sans toutefois former de boucle chromatinienne.

Après avoir généré la lignée CSE ANCHOR à double insertion, j'ai montré que le système ANCHOR ne perturbe ni l'expression de *Sox2*, ni la formation de la boucle chromatinienne. Plus précisément, j'ai confirmé entre autres par RT-qPCR en ciblant des allèles spécifiques, que la transcription de *Sox2*, la pluripotence et le potentiel de différenciation des CSE ne sont affectés ni par l'insertion des séquences parS, ni par le recrutement des protéines fluorescentes ParB. De plus, j'ai confirmé par 4C (Simonis *et al*., 2006; Van De Werken *et al*., 2012) que la boucle enhancer-promoteur n'est pas affectée par ces outils dans les CSE et que, comme pour les cellules de type sauvage, la boucle est perdue lors de la différenciation neuronale précoce. Les délétions alléliques spécifiques de différentes sous parties de la SCR ont montré leur contribution différentielle à l'expression de *Sox2* (laboratoire Mitchell, communication personnelle). J'ai donc créé deux lignées dérivées de ma lignée cellulaire ANCHOR de type sauvage en supprimant différents composants de l'enhancer. De la même façon que pour la lignée cellulaire ANCHOR de type sauvage, j'ai caractérisé des clones dans lesquels différentes sous parties de la SCR sont absentes et j'ai effectué des expériences 4C spécifiques à l'allèle (en ciblant SCR) pour évaluer la capacité de ces cellules à former une boucle chromatinienne indispensable à l'expression de *Sox2*. Ces expériences ont montré que des délétions situées dans l'enhancer entraînent une faible perte d'interaction entre *Sox2* et SCR. Afin de tester directement la relation entre le statut de l'ARN Pol II (ARN polymérase II) et le mouvement

de la chromatine dans des cellules vivantes, nous avons utilisé des petites molécules inhibitrices pour bloquer le cycle de transcription à deux étapes différentes : à l'initiation, avec le triptolide, ainsi qu'à l'élongation de l'ARN Pol II, avec le flavopiridol (Bensaude, 2011).

Afin de comprendre la relation entre les interactions promoteur-enhancer et la mobilité de la chromatine sur les paramètres d'activation de la transcription, j'ai utilisé le système MS2/MCP (Bertrand *et al*., 1998), qui permet la localisation d'ARNm uniques. J'ai inséré dans le génome de la lignée ANCHOR 24 fois la séquence MS2 et incorporé du MCP marqué par fluorescence pour l'imagerie simultanée en direct du statut transcriptionnel de *Sox2*. Après avoir établi et testé ces lignées cellulaires, j'ai obtenu et analysé un grand nombre de films courts de cellules individuelles pour examiner la dynamique de la chromatine dans des conditions de type sauvage dans les CSE, au cours des premiers stades de différenciation lorsque l'expression de *Sox2* est perdue, dans les sous-délétions de l'enhancer qui créent des perturbations de l'expression de *Sox2* et ont un effet léger sur le bouclage et, enfin, dans les CSE en réponse à l'inhibition pharmacologique de l'initiation ou de l'élongation transcriptionnelle. Cela a permis d'évaluer la variabilité d'une cellule à l'autre et de mesurer les paramètres de diffusion. En collaboration avec l'équipe de Nacho Molina (IGBMC, Strasbourg), nous avons développé une méthode pour mesurer précisément les paramètres de diffusion de la chromatine à partir de la microscopie en direct. Plus précisément, en effectuant une analyse de suivi de point unique, nous avons extrait deux paramètres : le coefficient de diffusion (D), qui reflète la vitesse de la diffusion; et le paramètre de diffusion anormale α, qui pour les petites valeurs ($\alpha<1$) indique que la diffusion est limitée à des volumes plus confinés.

J'ai démontré que *Sox2* et SCR sont fréquemment à proximité, et significativement plus proches que les régions de contrôle. Néanmoins, je n'ai pas mis en évidence de corrélation directe entre les distances de séparation entre *Sox2* et SCR et l'activité transcriptionnelle. En effet, l'activité transcriptionnelle est plutôt associée à des changements dans la dynamique locale de la chromatine. Le gène et l'enhancer subissent quant à eux des contraintes plus importantes que les régions de contrôles. La perturbation de l'expression de *Sox2* lors de la différenciation, la mutation du SCR et l'inhibition de l'initiation transcriptionnelle par l'ARN Pol II ont entraîné une diminution de la vitesse apparente de diffusion au niveau du promoteur de *Sox2*. Bien que les altérations de la dynamique locale de la chromatine soient en corrélation avec l'activité transcriptionelle des gènes, les loci transcriptionnellement actifs ne présentaient pas une

dynamique de la chromatine différente de celle des loci transcriptionnellement inactifs au cours de la même expérience. Je propose que les loci « en pause » aient une dynamique similaire à celle des loci transcriptionnellement actifs, et que les altérations de la mobilité accompagnent la transition d'un état réfractaire à un état transcriptionnellement compétent, à évaluer dans des expériences futures. Collectivement, ces résultats renforcent l'importance de l'évaluation de la dynamique de la chromatine aux sites de transcription comme moyen de comprendre le microenvironnement nucléaire et les règles qui régissent son organisation.

Revisite des modèles de boucle d'enhancer

Un grand nombre de recherches suggèrent que le bouclage enhancer-promoteur est nécessaire à l'expression des gènes (Bartman *et al*., 2016; H. Chen *et al*., 2018b; Deng *et al*., 2012, 2014; Palstra *et al*., 2003). Cependant, des expériences récentes d'imagerie ont démontré, au moins pour certains loci spécifiques, des enhancers à de grandes distances des gènes cibles activés (Alexander *et al*., 2019 ; Benabdallah *et al*., 2019 ; Lim *et al*., 2018), découplant ainsi la proximité enhancer-promoteur de l'expression génique. Dans cette étude, je montre que les distances *Sox2*-SCR sont significativement plus proches (médiane=146 nm) que les régions témoins de séparation génomique équivalente (médiane=195 nm), en accord avec les données Hi-C (Bonev *et al*., 2017). Néanmoins, les paires *Sox2*-SCR et de contrôle ont été trouvées à proximité immédiate, comme on pourrait s'y attendre pour des séparations génomiques de ~115 kb. L'imagerie à haut débit récente sur des cellules fixées a révélé un degré remarquable d'hétérogénéité et de variabilité entre les cellules dans l'organisation spatiale du génome (Bintu *et al*., 2018; Cattoni *et al*., 2017; Finn *et al*., 2019; Giorgetti *et al*., 2014). Bien que ma configuration expérimentale actuelle ne permette pas la visualisation simultanée des quatre régions, il est fort probable qu'il y ait des cellules individuelles où les régions témoins sont plus proches que *Sox2*-SCR. Des travaux futurs sont nécessaires pour solidifier cette hypothèse. Une expérience très ambitieuse consiste à réaliser un ANCHOR à triple marquage afin de déterminer si, et à quelle fréquence, les paires *Sox2*-SCR sont réellement plus proches que les paires *Sox2*-Inter contrôle et SCR-Inter contrôle.

Outre la caractérisation de la topologie de la chromatine à l'état pluripotent, où *Sox2* est exprimé, cette étude a également permis de mieux comprendre les distances de séparation entre les promoteurs et les enhancers aux premiers stades de la différenciation *in vitro* des CSE, où

Sox2 est réduit au silence. En accord avec les résultats de l'étude 4C, les distances moyennes entre *Sox2* et SCR ont augmenté de manière significative (médiane=205 nm) trois jours après la différenciation des CSE induite par l'acide rétinoïque (AR), mais les sondes étaient encore relativement proches dans l'espace, et beaucoup plus proches que les distances moyennes séparant *Sox2* et SCR dans les CSE proposées par une étude précédente (Alexander *et al*., 2019). Pour examiner directement si et comment la topologie de la chromatine est liée à l'activation des gènes, j'ai marqué l'ARNm *Sox2* natif de l'allèle *musculus* des CSE marquées ANCHOR avec des boucles MS2 (Bertrand *et al*., 1998), fournissant une lecture directe de la transcription du locus *Sox2* endogène. L'imagerie en direct par triple marquage a permis de mesurer simultanément la proximité de *Sox2* et de SCR et la transcription du gène.

Bien que l'imagerie d'un plus grand nombre de cellules soit nécessaire pour confirmer ces résultats préliminaires, je constate ici que *Sox2* et SCR sont légèrement plus proches pendant la transcription active et que le sous-ensemble de loci *Sox2* en transcription active présente des distances *Sox2*-SCR significativement plus proches par rapport à la distribution totale des expériences en double marquage. Ensemble, ces résultats correspondent à un modèle dans lequel la proximité physique entre l'enhancer et le promoteur est en quelque sorte corrélée à l'activation de la transcription. Néanmoins, lorsque j'ai évalué le lien entre la transcription de *Sox2* et l'architecture locale de la chromatine par d'autres moyens, en particulier après l'inhibition de l'initiation de la transcription lors du traitement au triptolide, qui réprime l'expression de *Sox2* à des niveaux presque indétectables, et après les délétions de sites critiques au sein du SCR, qui réduisent fortement l'expression de *Sox2* (>4 fois), je n'ai observé aucun effet sur la proximité *Sox2*-SCR. Dans l'ensemble, ces résultats suggèrent que le gène *Sox2* et son enhancer, sont fréquemment à proximité, mais que la distance de séparation exacte n'est pas nécessairement en corrélation avec la transcription. À la lumière de ces résultats et de ceux d'autres chercheurs (Benabdallah *et al*., 2019 ; Lim *et al*., 2018), le modèle classique de bouclage doit être révisé. La juxtaposition physique directe des promoteurs et des enhancers ne semble pas être requise (ni suffisante) pour que la régulation transcriptionnelle se déroule; au contraire, au moins pour *Sox2*, la communication entre les éléments semble être possible lorsqu'ils se trouvent dans un certain rayon, dont la taille reste à déterminer.

La dynamique chromatinienne de *Sox2* est cohérente avec les hubs et/ou condensats transcriptionnelles

Des recherches récentes ont mis en évidence l'existence de hubs transcriptionnelles qui créent des microenvironnements locaux, dans lesquels les enhancers et les promoteurs peuvent partager des clusters communs de facteurs Pol II et de facteurs de régulation (Lim & Levine, 2021). En tenant compte de mes résultats sur la distance de séparation enhancer-promoteur et les paramètres de diffusion de la chromatine, il est tentant de spéculer qu'un tel hub transcriptionnel pourrait englober *Sox2* et SCR, contraignant ainsi leur mobilité par rapport aux régions de contrôle. Puisque la séquence contrôle est également moins contrainte, elle peut sortir du hub transcriptionnel et ainsi être plus libre d'explorer le nucléoplasme. Mes résultats sont donc globalement cohérents avec un modèle dans lequel les promoteurs et les enhancers résident, sans nécessairement se juxtaposer directement, dans le même hub transcriptionnel, qui semble contraindre le mouvement local de la chromatine. Cependant, mes résultats ne soutiennent pas un modèle simple selon lequel les hubs contiennent exclusivement des loci transcriptionnellement actifs. Les contraintes chromatiniennes sont largement inchangées lors du traitement par des médicaments inhibant la transcription et, ce qui est plus révélateur, il n'y a pas de différences évidentes dans la dynamique des allèles transcriptionnellement actifs ou inactifs au sein des CSE de type sauvage. Les liens, s'ils existent, entre les hubs transcriptionnelles et les vitesses de diffusion apparentes semblent encore plus obscurs.

Le chargement de la Pol II et un effet plus indirect de la transcription sur la dynamique de la chromatine ?

À première vue, mes résultats semblent contenir une grande contradiction. Les résultats du double marquage après divers moyens de perturber l'expression de *Sox2* (différenciation, traitement avec des inhibiteurs de transcription, délétion des régions de l'enhancer) indiquent tous un lien direct entre la transcription et la dynamique de la chromatine, l'expression perturbée entraînant une réduction générale de la contrainte au niveau de l'enhancer et du promoteur, et une réduction plus spécifique de la vitesse apparente de diffusion au niveau du promoteur. Cependant, lors du marquage simultané de la transcription naissante de *Sox2*, il n'y a pas de différences apparentes dans les propriétés diffusives moyennes entre les allèles transcriptionnellement actifs ou inactifs, que ce soit au niveau du promoteur ou de l'enhancer.

Une explication possible est que dans les CSE de type sauvage, presque tous les allèles du locus *Sox2* fortement exprimé sont "prêts" à être activés transcriptionnellement dans le même microenvironnement nucléaire permissif, qu'ils synthétisent ou non l'ARNm au moment de l'acquisition de l'image. La dynamique de la chromatine est probablement déterminée par l'environnement local et ne devrait donc pas différer. Pour preuve, le promoteur de *Sox2* et le début du gène contiennent des niveaux élevés de Pol II en pause dans les CSE, une caractéristique connue des gènes en pause et rapidement inductibles (Williams *et al*., 2015). Les perturbations qui convertissent les loci *Sox2* en un état plus réfractaire peuvent le faire en modifiant le microenvironnement nucléaire, avec des effets ultérieurs sur la dynamique de la chromatine. Par exemple, la différenciation entraîne des changements de masse dans la charge de Pol II au niveau du promoteur, ainsi qu'une perte de liaison TF (facteur de transcription) à la fois au niveau du promoteur et du SCR; ces changements peuvent réduire directement les contraintes de mobilité de la chromatine locale, ou indirectement via la dissociation des hubs transcriptionnelles. Des expériences de triple marquage réalisées sur des périodes beaucoup plus longues pourraient également permettre de découvrir si des loci de type sauvage font la transition entre les états "réfractaire" et "en pause" proposés, déterminés par des altérations à plus long terme des paramètres de diffusion locaux. Il est intéressant de noter que la dynamique de la chromatine n'a pratiquement pas été modifiée par le traitement avec un inhibiteur de l'élongation de la transcription, le flavopiridol, mais que la dynamique du promoteur a été sensible au traitement avec un inhibiteur de l'initiation de la transcription, le triptolide, ce qui suggère que le recrutement de Pol II et/ou la fusion du promoteur, mais pas l'élongation processive, sont les étapes déterminantes de la mobilité locale de la chromatine.

Comparaison avec les résultats précédents

Pendant ma thèse, un autre laboratoire a publié ses travaux, dans lesquels ils ont interrogé l'organisation spatiale et l'activité de *Sox2* et SCR dans des CSE F1 (Alexander *et al*., 2019). En utilisant des opérateurs cuO et tetO pour marquer le promoteur et l'enhancer de l'allèle 129, les auteurs ont démontré que *Sox2* et SCR sont séparés par des distances beaucoup plus grandes (moyenne ~340 nm), par rapport à celles rapportées dans mon étude, dans le noyau des CSE. J'ai directement évalué cette divergence en introduisant ANCHOR dans les mêmes emplacements génomiques exacts de cette étude, et j'ai constaté que ces régions étaient aussi

fréquemment proximales. Je propose donc que les grandes distances de séparation rapportées par les auteurs (Alexander *et al*., 2019) soient dues à des différences techniques, peut-être causées par les grands opérateurs répétitifs tet/cu. En effet, les auteurs ont inclus une lignée témoin, où un fragment de 111 kb a été supprimé entre les paires cuO et tetO, laissant une attache de 14 kb entre elles (Alexander *et al*., 2019). Curieusement, la distance de séparation moyenne observée était de ~250 nm et jamais inférieure à 100 nm, comme cela a été précédemment rapporté pour des paires de sondes DNA FISH séparées par quelques dizaines (~30-60) de kilobases (Giorgetti *et al*., 2014). Je suppose que les grandes distances de séparation rapportées par Alexander *et al*. (Alexander *et al*., 2019) sont dues à l'intégration de copies multiples de séquences répétitives d'opérateurs et à la liaison stable des répresseurs marqués par fluorescence.

En outre, dans cette étude, en utilisant un système *Sox2*-MS2, nous avons mesuré des fréquences transcriptionnelles beaucoup plus élevées par rapport à ce qui a été rapporté précédemment (Alexander *et al*., 2019). Ces différences pourraient être causées par le marqueur de sélection de résistance qui était maintenu dans le locus dans l'autre étude (excisé dans mes cellules Anchor_*Sox2*-MS2), et pourrait perturber davantage la transcription de *Sox2* et la stabilité de l'ARNm. En effet, j'ai observé que la lignée conservant le gène marqueur sélectionnable, présentait des niveaux réduits d'ARNm naissant et de protéines, mesurés par smFISH et western blot. Ces résultats confèrent une valeur supplémentaire à mes systèmes de marquage ANCHOR et MS2-MCP. Cependant, malgré l'écart important dans les mesures de distance, les deux études remettent en question le modèle classique de boucle enhancer-promoteur.

Alors que j'ai observé une plus grande séparation entre *Sox2* et SCR au cours des tout premiers stades de la différenciation in vitro des CSE, Alexander *et al*. ont signalé une plus grande proximité entre *Sox2* et SCR après la différenciation complète en précurseurs neuronaux. De plus, des distances de séparation accrues ont été observées entre *Shh* et les enhancers cérébraux *Shh* pendant la différenciation neuronale, où *Shh* est actif (Benabdallah *et al*., 2019). Alors que ces derniers pourraient être attribués au contexte génomique différent, les résultats opposés de *Sox2*-SCR pourraient provenir des différents stades de différenciation suivis dans chaque étude. Alexander *et al*. ont proposé que l'ensemble de la région *Sox2* adopte une conformation plus compacte lors de la différenciation des CSE. Des expériences de différenciation plus longues

de mes CSE Anchor_Sox2-SCR seront nécessaires pour confirmer si le locus *Sox2* présente une structure compacte, si oui, comment cela affecte la proximité promoteur-enhancer et la dynamique locale.

Les paramètres de diffusion locale comme "mesures" de l'activité régulatrice ?

À ma connaissance, ce travail est la première étude d'imagerie en direct qui décrit les distances de séparation et les propriétés de diffusion d'une paire enhancer-promoteur, simultanément à la transcription du gène, et en réponse à des inhibiteurs de la transcription. Des études précédentes ont également abordé le lien entre la dynamique de la chromatine et l'expression des gènes (Germier *et al*., 2017 ; Gu *et al*., 2018), atteignant des résultats apparemment contradictoires. Mes résultats confirment et contredisent à la fois ces résultats précédents dans une certaine mesure. Tout d'abord, comme décrit ci-dessus pour mes propres résultats, le paramètre de diffusion anormale (α) semble lié à "l'activité", avec une plus grande contrainte dans les régions actives, en accord avec un rapport récent, où il a été démontré que la transcription coïncide avec le confinement d'un gène producteur d'ARNm dans des volumes nucléaires plus petits (Germier *et al*., 2017).

Plus précisément, j'ai également constaté que le promoteur de *Sox2*, mais pas le SCR, diffuse plus rapidement (D_{app} plus élevé) lors de l'activation. Ce résultat concorde partiellement avec un rapport récent: un gène et un enhancer activés lors de la différenciation se déplacent plus rapidement (Gu *et al*., 2018), interprété dans cette étude comme une augmentation de l'énergie thermique via le processus de la transcription active. Une explication alternative (voir également ci-dessus) est que les premières étapes du cycle de la transcription, et en particulier le remodelage de la chromatine et l'élimination des nucléosomes lorsque le PIC est recruté et/ou la fusion du promoteur après l'assemblage du PIC, aboutissent en fait à une chromatine moins dense qui peut finalement se déplacer plus rapidement. La mobilité réduite de *Sox2*, telle que montrée dans cette étude lors d'un traitement au triptolide, un inhibiteur connu de l'activité hélicase de TFIIH (Bensaude, 2011), est cohérente avec cette hypothèse. En revanche, contrairement à Gu *et al*., je n'ai pas observé de lien entre la vitesse de diffusion de l'enhancer et l'activation du gène. Les données RNA-seq disponibles publiquement provenant de CSE Dunham *et al*., 2012) suggèrent que le SCR ne produit aucun eRNA. Bien que nous n'ayons

pas accès à des données analogues provenant d'EpiLCs, où l'enhancer *Fgf5* évalué est actif (Gu *et al*., 2018), son activité transcriptionnelle pourrait expliquer pourquoi il se comporte comme les promoteurs actifs de *Fgf5* et *Sox2*, et non comme le SCR. De plus, contrairement à Gu *et al*., j'ai conclu que l'inhibition de l'élongation transcriptionnelle n'a aucun effet sur la mobilité chromatinienne de l'un ou l'autre des éléments régulateurs. Il reste à voir si cela peut être expliqué par des différences techniques ou des contextes spécifiques aux gènes.

Mes résultats, ainsi que d'autres (Germier *et al*., 2017 ; Gu *et al*., 2018), soulignent l'importance d'évaluer la dynamique de la chromatine aux sites de transcription comme moyen de comprendre le microenvironnement nucléaire et les principes de régulation. Néanmoins, D_{app} et α sont des paramètres diffusifs assez généraux. Par exemple, un α réduit peut s'expliquer par de multiples mécanismes, tels que le confinement à un volume réduit ou "l'attraction" vers des régions spécifiques, qu'il est impossible de distinguer. J'attends avec impatience les progrès de la théorie physique permettant de définir des paramètres diffusifs inférés plus précis, ce qui pourrait donner un meilleur aperçu et permettre une analyse encore plus précise de la dynamique locale de la chromatine. Si l'on ajoute à cela des outils d'imagerie plus sophistiqués et de meilleures méthodes de manipulation du noyau (par exemple, l'ablation de protéines régulatrices clés à l'aide du système de dégradation de l'auxine (N. Q. Liu *et al*., 2021; P. Nora *et al*., 2017; Rao *et al*., 2017), ou la manipulation des propriétés de séparation des phases liquide-liquide (Y. Shin *et al*., 2018), la compréhension globale des aspects quadridimensionnels de la régulation transcriptionnelle devient de plus en plus réalisable.

Introduction

Transcription

Eukaryotic genomes are much more than linear sequences of DNA. The DNA helix must be condensed in the form of chromatin to fit within the cell nucleus, which has a diameter only about one-tenth that of a human hair. The fundamental unit of chromatin is the nucleosome core particle, which consists of 147 base pairs (bp) of DNA wrapped 1.7 turns around a histone octamer. Each histone octamer is composed of two copies of H2A and H2B (forming two H2A/H2B dimers) and two copies of H3 and H4 (forming one H3/H4 tetramer)(Kornberg, 1974; Kornberg & Thomas, 1973; Luger, Mäder, Richmond, Sargent, & Richmond, 1997; Olins & Olins, 1974) (**Figure 1**). Further packaging of DNA is brought about by formation of the chromatosome core particle , composed of a linker histone H1 bound to the nucleosome (Simpson, 1978). The spatial organization of chromatin not only serves as a way to compact DNA but also affects nuclear processes involving DNA, including transcription, DNA replication and DNA repair.

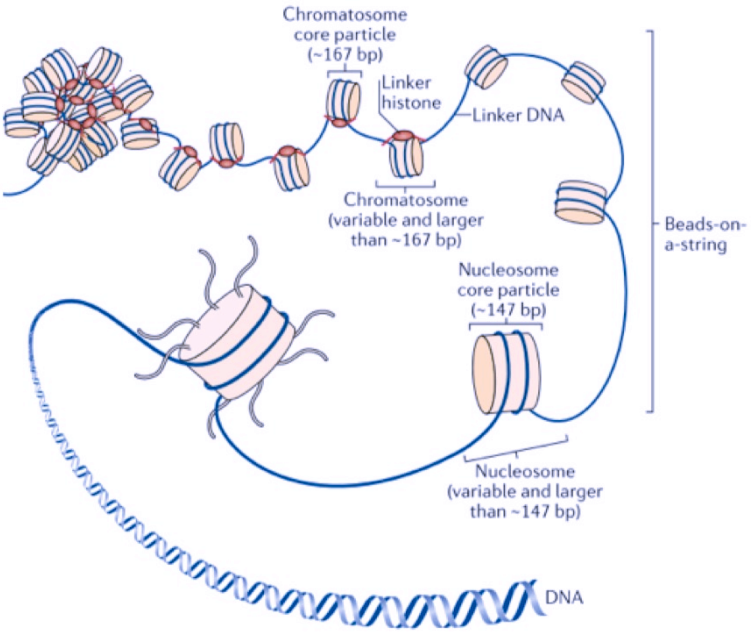


Figure 1: Compaction of DNA into chromatin within the interphase nucleus. Adapted from (Fyodorov, Zhou, Skoultchi, & Bai, 2018).

Transcription cycle

Transcription is the process by which the information from a DNA template is copied into a new RNA molecule. In eukaryotes, RNA Polymerase II (Pol II) transcribes messenger RNAs (mRNAs) from protein-coding genes. Access to key *cis*-regulatory binding sites is restricted by nucleosomes, thus Pol II collaborates with transcriptional regulators - also known as activators or repressors -, general transcription factors (GTFs) including TATA-binding protein (TBP), TFII-A, -B, -D, -E, -F, and -H and co-activators, like the Mediator complex (Esnault et al., 2008; Y. J. Kim, Björklund, Li, Sayre, & Kornberg, 1994; Plaschka et al., 2015; Workman & Kingston, 1992), chromatin modifiers and remodelers, in order to get access to the promoter and form the preinitiation complex (PIC) (Lorch, Zhang, & Kornberg, 1999; Mizuguchi et al., 2004; Narlikar, Fan, & Kingston, 2002; Zhang, Roberts, & Cairns, 2005). Eukaryotic transcription is a complex and highly dynamic event, exhibiting regulation at multiple steps. Regulation of transcriptional output requires the concerted assembly of trans-acting chromatin factors and transcription factors (TFs) on proximal and distal regulatory DNA elements, termed "promoters" and "enhancers", respectively (discussed in detail later on in the Introduction). Both promoters and enhancers contain binding sites for gene-specific TFs that determine when a gene will be activated throughout development (Lambert et al., 2018; Shlyueva, Stampfel, & Stark, 2014).

The transcription cycle can be divided in eight major steps: chromatin opening, pre-initiation complex (PIC) formation, initiation, promoter-proximal pausing, pause release, productive elongation, termination and recycling (**Figure 2**). The tight packaging of DNA into chromatin can present an obstacle to transcription by rendering recognition sites inaccessible. Hence, nucleosome displacement or ejection seems to be a critical event in the process of transcription (Mueller et al., 2017; Petesch & Lis, 2008). Such changes in chromatin accessibility are believed to be mediated by a combination of regulatory proteins, including a special class of TFs called pioneer factors (Mayran & Drouin, 2018; Zaret & Carroll, 2011), histone acetylation, chromatin-remodeling complexes, and perhaps the histone variant H2A.Z (Esnault et al., 2008; Y. J. Kim et al., 1994; Lorch et al., 1999; Mizuguchi et al., 2004; Narlikar et al., 2002; Plaschka et al., 2015; Workman & Kingston, 1992; Zhang et al., 2005) (**Figure 3**).

Binding of these regulatory factors serves as a mechanism of selection between the cell-type specific genes that can be transcribed out of a large pool of regulatory DNA elements that are available in the genome.

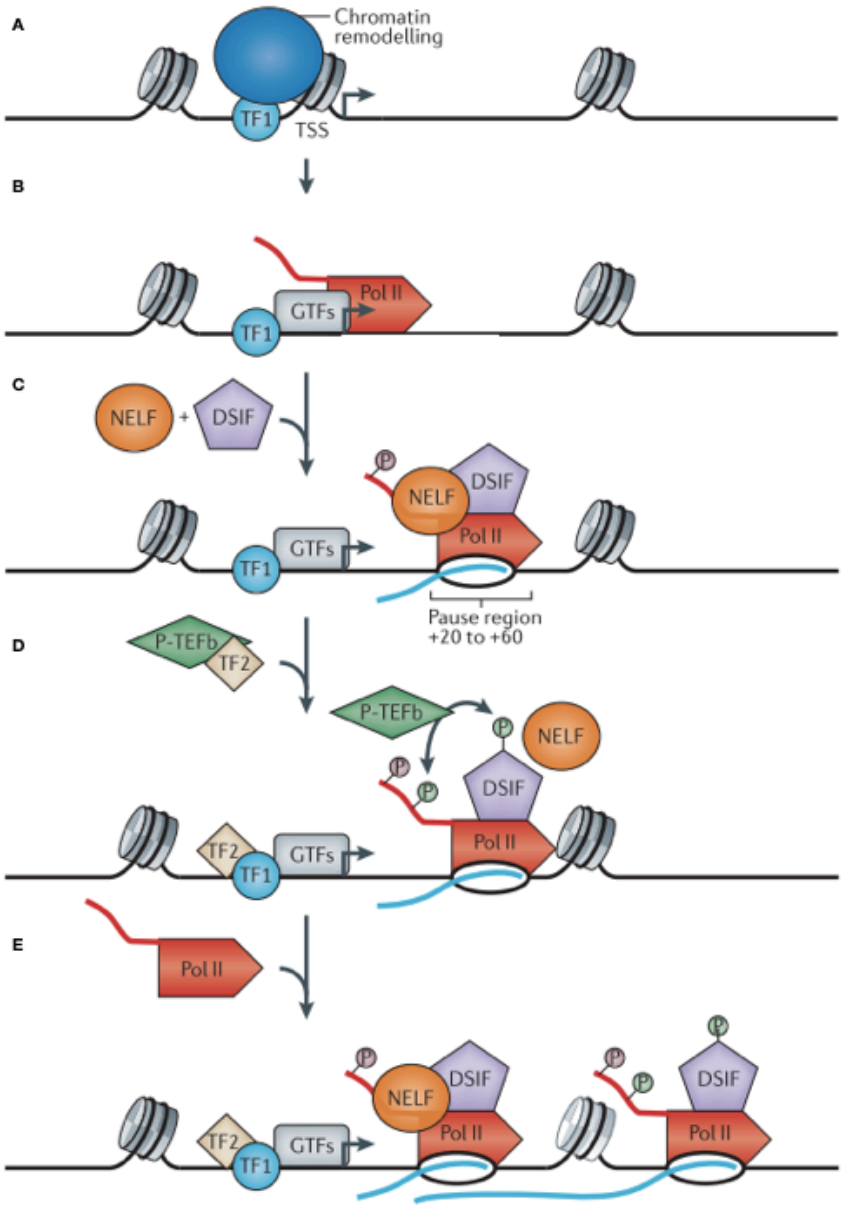


Figure 2: Schematic illustration of the transcription cycle. (A) Promoter opening at the core promoter (B) Pre-initiation complex assembly. (C) Promoter-proximal pausing. (D) Pause release. (E) Productive elongation and filling of paused region. Adapted from (Adelman & Lis, 2012).

After this series of modifications, the recognition sites that were buried in chromatin are exposed to TFs and RNA Polymerase II. Gene-specific TFs associate with sequence-specific sites on the promoter region (Lambert et al., 2018). This binding triggers the recruitment of GTFs and RNA Pol II, resulting in the proper docking of a functional pre-initiation complex (Haberle & Stark, 2018). After the assembly of the PIC, engaged Pol II unwinds locally the DNA to form the transcription bubble and initiates RNA synthesis. As the RNA extends further, contacts with the promoter and the promoter-bound GTFs are broken, resulting in promoter escape and transition to elongation (Y. J. Kim et al., 1994; Y. Liu et al., 2004). In many eukaryotic genes, RNA Pol II undergoes promoter-proximal pausing within the first 20-60 bp before it starts productive elongation (Krumm, Meulia, Brunvand, & Groudine, 1992; Leighton J. Core, Waterfall, & Lis, 2008; Muse et al., 2007; Rougvie & Lis, 1988; Zeitlinger et al., 2007).

The phenomenon of Pol II pausing is widespread across metazoans and is mediated by two pausing-inducing factors: the negative elongation factor (NELF) and the DRB sensitivity-inducing factor (DSIF) (Muse et al., 2007; Vos, Farnung, Urlaub, & Cramer, 2018; Wada, Takagi, Yamaguchi, Ferdous, et al., 1998; C. H. Wu et al., 2003; Yamaguchi et al., 1999). Paused Pol II is fully competent to resume elongation, remaining stably engaged with the nascent RNA. However, productive transcription requires escape of paused Pol II from the promoter and transition to rapid elongation of the transcript. Pause-release of RNA Poll II is triggered by the recruitment of the positive transcription elongation factor b (P-TEFb) -which consists of CDK9 and cyclin T1- (N. F. Marshall & Price, 1995; Ni et al., 2008; Rahl et al., 2010); treatment of cells with the P-TEFb inhibitor flavopiridol dramatically decreases global transcription (Chao & Price, 2001; Jonkers, Kwak, & Lis, 2014). Cyclin-dependent kinase 9 (CDK9) subunit of P-TEFb phosphorylates the C-terminal domain (CTD) of RNA Pol II and the DSIF–NELF complex, causing NELF to dissociate from Pol II and transforming DSIF to a positive association factor that associates with Pol II that traverses through the gene body (Harlen & Churchman, 2017; Lu et al., 2018; Wada, Takagi, Yamaguchi, Watanabe, & Handa, 1998).

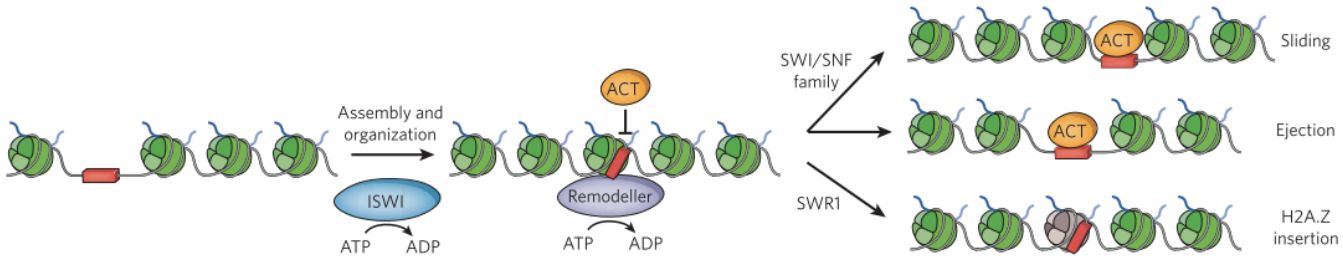


Figure 3: Function of chromatin remodelers in nucleosome dynamics. Remodelers use the energy from ATP hydrolysis to alter promoter architecture and expose binding sites in the promoter. Most remodelers in the ISWI family help conduct chromatin assembly and nucleosome organization which often promotes repression. SWI/SNF-family remodelers can slide and eject nucleosomes and positively stimulate transcription. Remodelers of the SWR1 family reconstruct nucleosomes by replacing canonical H2A with the histone variant H2A.Z. H2A.Z nucleosomes are less stable than H2A-nucleosomes and therefore might be more easily ejected from promoters in certain temporal contexts. Adapted from (Cairns, 2009).

As previously described, folding promoter DNA around histones to form nucleosomes renders important recognition sites inaccessible. Thus, promoters undergo nucleosome removal to allow the recruitment of the transcription machinery (Lorch et al., 1999; Mizuguchi et al., 2004; Narlikar et al., 2002; Zhang et al., 2005) (**Figure 3**). Genes with paused Pol II have been shown to remove nucleosomes and open promoters before gene activation (Costlow & Lis, 1984; C. Wu, 1980). In addition, it has been demonstrated that paused genes possess low levels of nucleosome occupancy that depends on the presence of promoter-paused polymerase (Gilchrist et al., 2010, 2008). The relationship between paused polymerase and promoter-proximal nucleosome organization has been resolved in *D. melanogaster* (Gilchrist et al., 2010): depletion of NELF largely reduced promoter-proximal Pol II occupancy at highly regulated genes and led to an increase in nucleosome occupancy downstream of the TSS at these genes. These findings suggested that pausing could serve as a mechanism to maintain open and regulatory factor-accessible promoters, hence facilitating the binding of the transcription machinery and future gene expression. Besides the chromatin opening function of paused Pol II, another potential role of pausing is to ensure rapid and/or synchronous gene activation. The presence of a preloaded Pol II already bound by transcriptional activators and co-activators would allow a rapid switch into productive elongation. In support of this idea, pausing has been

identified at highly inducible genes and a number of genes that are implicated in early embryonic development at *Drosophila melanogaster* (Boettiger & Levine, 2009; Hendrix, Hong, Zeitlinger, Rokhsar, & Levine, 2008; Muse et al., 2007; Zeitlinger et al., 2007). However, it is important to mention that pausing before activation is not a feature of all rapidly inducible genes (Kininis et al., 2007).

Finally, after pause release, at the 3' end of the gene, the nascent RNA is released, cleaved and polyadenylated (Proudfoot, 2016). The process of polyadenylation is important for later nuclear export, translation and stability of the mRNA transcript. After cleavage, Pol II terminates transcription, dissociates from the template, and eventually can be recycled to start a new round of the transcriptional cycle (Leighton J. Core et al., 2008; Nojima et al., 2015; Yudkovsky, Ranish, & Hahn, 2000).

Transcriptional bursting

As noted earlier, gene expression is a highly regulated process. At the same time, single-cell analysis has revealed that gene expression is a stochastic process and that stable phenotypes in a population are derived from variable single-cell gene expression patterns (Blake, Kærn, Cantor, & Collins, 2003; Gregor, Tank, Wieschaus, & Bialek, 2007). Traditional biochemical technologies, such as chromatin immunoprecipitation (ChIP), have been used to elucidate the steps of transcription in millions of cells simultaneously (Rhee & Pugh, 2011; Shang, Hu, Direnzo, Lazar, & Brown, 2000). However, the behavior measured on the population level does not reflect accurately the behavior in individual cells (Lickwar, Mueller, Hanlon, McNally, & Lieb, 2012; Stenoien et al., 2001).

Visualization of transcription in individual cells provided evidence that genes are often transcribed in discontinuous bursts, with periods of gene activity followed by periods of inactivity (Boettiger & Levine, 2009; Chubb, Trcek, Shenoy, & Singer, 2006; Golding, Paulsson, Zawilski, & Cox, 2005; Karpova et al., 2008; Larson, Zenklusen, Wu, Chao, & Singer, 2011; Raj, Peskin, Tranchina, Vargas, & Tyagi, 2006; Suter et al., 2011; Yunger, Rosenfeld, Garini, & Shav-Tal, 2010). These bursts were first described in the late 1980s where McKnight et al. (McKnight & Miller, 1979) observed some "fiber-free gaps" located within

transcription units in Drosophila embryos. Since then, bursting has been demonstrated to be ubiquitous from bacteria to human cells (Boettiger & Levine, 2009; Chubb et al., 2006; Golding et al., 2005; Suter et al., 2011; Yunger et al., 2010). Live-cell recordings (Chubb et al., 2006; Fukaya, Lim, & Levine, 2016; Suter et al., 2011) and single-molecule fluorescence *in situ* hybridization (smFISH) (Bartman, Hsu, Hsiung, Raj, & Blobel, 2016; Fukaya et al., 2016; Raj et al., 2006) have been used to characterize transcriptional bursting. All these observations suggested that any aspect of gene bursting can be regulated, including burst duration, amplitude and frequency (number of bursts per unit time) (Bartman et al., 2016; Dar et al., 2012; Fukaya et al., 2016; Suter et al., 2011) (**Figure 4**).

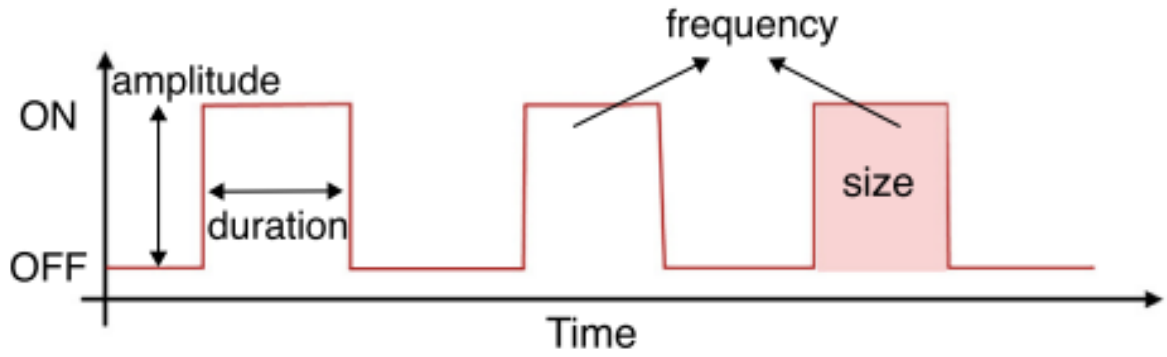


Figure 4: Description of bursting parameters. Adapted from (Lim, 2018).

Bursting is often described using a two-state model, where a promoter switches between two states, on and off, and transcription can occur only at the on state. Despite being widely applied in modeling, several studies on mammalian cells reported the presence of a refractory period between bursts: after responding to a stimulus, the template is no longer permissive for transcription (Harper et al., 2011; C. Li, Cesbron, Oehler, Brunner, & Höfer, 2018; Molina et al., 2013; Suter et al., 2011). The presence of this reset time is of great importance for several genes since they acquire a "memory" in the reactivation process (Harper et al., 2011; Suter et al., 2011). This observation suggests the existence of two major subpopulations across a cell population, in which a fraction of cells are poised for induction while another fraction are in a refractory state. This heterogeneity in the cell state allows the population to have both a rapid and a longer maintained response. Furthermore, having a refractory period indicates that several biochemical reactions are involved in order to reactivate a previously fired gene (Harper et al., 2011; Suter et al., 2011).

Several seminal studies have given valuable insight that points toward a role for chromatin, transcription factors and promoter-enhancer communications in regulating bursting. A proposed mechanism underlying gene bursting is the state of chromatin (Becskei, Kaufmann, & Van Oudenaarden, 2005; Raj et al., 2006; Raser & O'Shea, 2004). In this model, the promoter's active and inactive states may reflect different chromatin conformations, of which only some are permissive for transcription, thus controlling the timing of gene firing and inactivation (Hager, McNally, & Misteli, 2009). Another model suggests that bursting properties are influenced by the kinetics of transcription-factor binding (J. Chen et al., 2014; Karpova, Chen, Sprague, & McNally, 2004; Yihan Lin, Sohn, Dalal, Cai, & Elowitz, 2015; Loffreda et al., 2017; Mir et al., 2017; Senecal et al., 2014; Swinstead et al., 2016; Thomas A. Johnson & Cem Elbi, Bhavin S. Parekh,§ Gordon L. Hager, 2008).

Recent advances in *in vivo* fluorescence microscopy, such as fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) (Scalettar, Hearst, & Klein, 1989), have made it possible to resolve the binding and the diffusion of molecules in the nucleus of living cells: transcription factors were found to bind to DNA in a dynamic manner and have short residence times (Karpova et al., 2004; Phair et al., 2004; Swinstead et al., 2016; Thomas A. Johnson & Cem Elbi, Bhavin S. Parekh,§ Gordon L. Hager, 2008), while the core histones (H2A, H2B, H3 and H4) display a more stable binding (Kimura & Cook, 2001; Phair et al., 2004). Since TFs show high turnover, it has been assumed that the activity of target genes depends on the TF on-rate and thus local TF concentration. Consistently, by utilizing a combination of single molecule observations of RNA in fixed and living cells with computational modeling, it has been demonstrated that TF concentration regulates the burst frequency (Mir et al., 2017; Senecal et al., 2014). A supplementary way of regulating gene bursting is through modulation of the TF dwell time (J. Chen et al., 2014; Lickwar et al., 2012; Yihan Lin et al., 2015; Loffreda et al., 2017; Swinstead et al., 2016). Transcription factor acetylation (Loffreda et al., 2017), affinity of the TF binding site (Suter et al., 2011) or cooperative binding of other regulators (J. Chen et al., 2014; Yihan Lin et al., 2015; Swinstead et al., 2016) determine the residence time of TFs. Long dwell time of a transcription factor might increase the number of polymerases that are recruited to the target gene, thereby increasing the burst duration of the gene. These models of transcriptional bursting have focused on biochemical activities at gene promoters, and are further confounded by their complex interplay with distal regulatory elements. The models proposed above may operate

simultaneously on the same gene, providing cells the ability to regulate different bursting parameters independently, hence tuning the levels and the heterogeneity of gene expression between cells.

Distal regulatory elements

Enhancers

Early transgenic studies have revealed that gene expression during development is weak in the absence of additional, often distal, *cis*-acting regulatory elements (Banerji, Olson, & Schaffner, 1983; Banerji, Rusconi, & Schaffner, 1981; Benoist & Chambon, 1981; Gillies, Morrison, Oi, & Tonegawa, 1983; Moreau et al., 1981). These elements, named enhancers, were first described in the Xenopus oocyte where a remote upstream sequence, originally termed modulator, positively affected the expression of the sea urchin H2A histone gene (Grosschedl & Birnstiel, 1980). Deletion of this element resulted in a 15- to 20-fold decrease of H2A gene expression. Shortly after this initial observation, plasmid-based assays revealed that the tandem 72 bp simian virus 40 (SV40) DNA repeat located 150 bp upstream from the cap site of the SV40 early gene was essential for SV40 expression (Benoist & Chambon, 1981). Deletion of this upstream sequence reduced early gene expression of T antigen to undetectable levels and concomitantly abolished virus viability. Likewise, the same 72 bp repeated DNA element was shown to increase the expression of a β-globin gene in HeLa cells by more than two orders of magnitude even when it was located 3300 bp downstream from the promoter (Banerji et al., 1981). The first endogenous enhancer found in mammalian genomes was the immunoglobulin (Ig) heavy-chain enhancer (Banerji et al., 1983; Gillies et al., 1983; Mercola, Wang, Olsen, & Calame, 1983; Neuberger, 1983). Notably, the Ig enhancer is the first example of an enhancer that functions in a tissue- or cell-type specific manner. Of the various cell-lines tested, Ig enhancer activity was observed only in B-lymphocyte-derived myeloma cells (Banerji et al., 1983; Gillies et al., 1983).

Properties and classification

Enhancers appear to increase or activate the rate of transcription in a position- and orientation-independent manner with respect to their target genes (Banerji et al., 1983; Benoist & Chambon, 1981; Fromm & Berg, 1983; Moreau et al., 1981). This flexibility is the defining hallmark of enhancers. In many cases, enhancers are found in the non-coding part of the

genome, at great genomic distances from their target genes (sometimes more than hundreds of kilobases) and can bypass more proximal genes to regulate their targets (Amano et al., 2009; Mifsud et al., 2015; Sanyal, Lajoie, Jain, & Dekker, 2012; Schoenfelder et al., 2015) (**Figure 5**). One of the most characteristic examples reported so far, is a mouse limb bud enhancer of the Sonic Hedgehog (Shh) gene, called ZRS (ZPA regulatory sequence), which is found in the intron of another gene, approximately 1 megabase (Mb) upstream of the Shh promoter (Amano et al., 2009; Lettice et al., 2003; Sagai, Hosoya, Mizushina, Tamura, & Shiroishi, 2005).

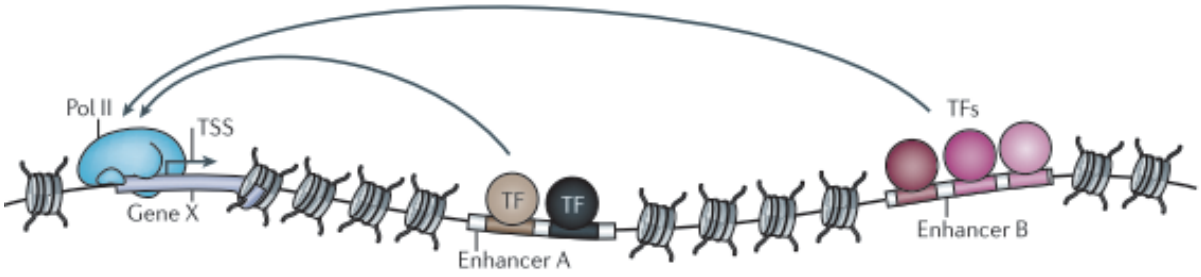


Figure 5 : Schematic representation of enhancers located distally from their target gene and upregulating its expression. Adapted from (Shlyueva et al., 2014).

Similar to promoters, enhancers contain DNA recognition motifs that attract sequence-specific TFs, RNA Pol II and GTFs (Koch et al., 2011; Lambert et al., 2018). These proteins recruit co-activators and co-repressors, such that the combination of all bound factors will ultimately determine the activity of the enhancer in response to developmental cues. Moreover, active enhancers exhibit DNase I hypersensitivity (HS), which reflects an "open" chromatin state - depleted of nucleosomes - as a result of the binding of several transcription factors (Boyle et al., 2008; Roadmap Epigenomics Consortium et al., 2015; Thurman et al., 2012). Nucleosomes flanking the enhancers show characteristic post-translational histone modifications (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner, Tesar, & Scacheri, 2011). Depending on these modifications, enhancers can be divided into several states: active, inactive/poised and intermediate/primed. Active enhancers display high levels of histone H3 lysine 4 mono-methylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac), a major substrate for the histone acetyltransferase CBP/p300 (Creyghton et al., 2010; Jin et al., 2011; Rada-Iglesias et al., 2011; Tie et al., 2009) (**Figure 6A**), as well as the "alternative" marks histone H3 lysine 64 acetylation (H3K64ac) and histone H3 lysine 122 acetylation (H3K122ac)

recently identified in mouse embryonic stem cells (mESCs) (Pradeepa et al., 2016). Inactive or poised enhancers are marked by H3K4me1 and H3K27me3, a mark enriched in Polycomb (PcG)-associated and transcriptionally repressed regions (Creyghton et al., 2010; Rada-Iglesias et al., 2011) (**Figure 6B**). In addition, enhancers that have not yet been activated but are primed for activation either at a later developmental stage or in response to external stimuli can be pre-marked by H3K4me1 (Zentner et al., 2011) (**Figure 6C**).

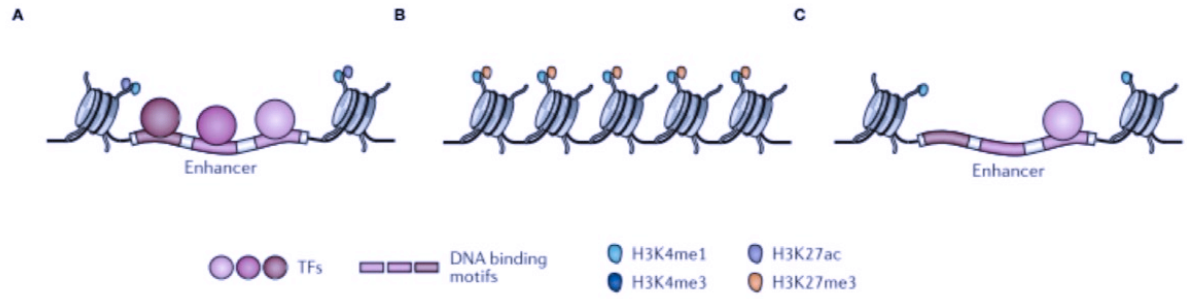


Figure 6 : Chromatin accessibility and histone marks at enhancers. (A) Active enhancers often bear the H3K27ac and H3K4me1 marks. (B) Inactive or closed enhancer regions are often marked by the active H3K4me1 and the repressive H3K27me3 marks. (C) Primed enhancers can be marked by H3K4me1. Adapted from (Shlyueva et al., 2014).

Identification and validation

The advent of high-throughput and deep sequencing methods has enabled the systematic study of potential regulatory elements: more than one million putative enhancers have been documented in several mouse and human tissues and cell types (Altshuler et al., 2012; Dunham et al., 2012; Roadmap Epigenomics Consortium et al., 2015; Shen et al., 2012; Thurman et al., 2012). Several approaches have been used in order to predict and functionally assess transcriptional enhancers, including the analysis of sequence motifs, chromatin structure and histone modifications, the binding of TFs and cofactors, as well as parallel and/or genome-wide screens of enhancer activity.

As mentioned above, enhancers are bound by a large range of transcription factors and cofactors. Given this association, computational analysis of TF binding motifs combined with

the assessment of evolutionary DNA sequence conservation was used as an approach for predicting putative enhancer sequences (Berman et al., 2002; Del Bene et al., 2007; Kheradpour, Stark, Roy, & Kellis, 2007; Pennacchio et al., 2006). However, even though these types of approaches provide a first overview of the regulatory elements across the genome, they seem to identify only a proportion of enhancers. For instance, it has been demonstrated that only a small fraction of the conserved motif matches are bound by TFs *in vivo* in a given tissue or stage and that transcription factor binding can be context-dependent and rely on the formation of complexes with other proteins (Birney et al., 2007; Z. Liu et al., 2014; Slattery, 2014; Yáñez-Cuna, Dinh, Kvon, Shlyueva, & Stark, 2012). As the role of motifs is to recruit TFs, chromatin immunoprecipitation followed by either microarray analysis (ChIP-on-chip) or deep sequencing (ChIP–seq) enables the determination of *in vivo* TF binding sites in an unbiased, genome-wide and systematic manner (Johnson, Mortazavi, & Myers, 2007; Robertson et al., 2007) (**Figure 7A**). Alternatively, ChIP-seq has been commonly used as a method for identifying the *in vivo* binding sites of transcriptional cofactors, such as the histone acetyltransferase p300 (May et al., 2012; Visel, Blow, et al., 2009; Visel et al., 2013) (**Figure 7C**). Binding of p300 has enabled the genome-wide prediction of enhancers in mice and humans (Heintzman et al., 2007; Rada-Iglesias et al., 2011; Visel, Blow, et al., 2009), while a great amount of the p300-binding sites tested using episomal mouse reporter assays demonstrated reproducible enhancer activity (Blow, 2010; May et al., 2012; Visel, Blow, et al., 2009; Visel et al., 2013).

A complementary method in identifying enhancers takes advantage of their chromatin accessibility. Active enhancers have been shown to persist in a nucleosome-deprived chromatin state and display hypersensitivity to DNase I digestion (Boyle et al., 2008; Thurman et al., 2012). Next-generation sequence-based techniques, such as DNase-seq (Boyle et al., 2008), have been utilized to determine enhancers independently of any given TF binding motif or TF binding (**Figure 7B**). As described above, nucleosomes flanking active enhancers show a characteristic pattern of post-translational histone modifications (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Zentner et al., 2011). Genome-wide prediction of enhancers using epigenetic marks has been used extensively in a plethora of biological contexts to predict regulatory elements (Bonn et al., 2012; Heintzman et al., 2009, 2007; Rada-Iglesias et al., 2011; Roh, Cuddapah, & Zhao, 2005) and it agrees well with enhancer activity assays (Arnold et al., 2013; Bonn et al., 2012; Heintzman et al., 2007). However, none of the known histone marks

correlates perfectly with enhancer activity (Arnold et al., 2013; Bonn et al., 2012). Tissue-specific analysis of chromatin state showed that several active mesodermal enhancers in D. melanogaster embryos displayed heterogeneous histone modifications and were not predicted by the presence of H3K27ac (Bonn et al., 2012). Furthermore, it is becoming increasingly apparent that regulatory elements can be devoid of classical enhancer chromatin marks but still display enhancer functions (Pradeepa et al., 2016; Rajagopal et al., 2016). In particular, a new class of active functional enhancers was found in mouse ESCs that is marked by H3K122ac, but lacking H3K27ac (Pradeepa et al., 2016).

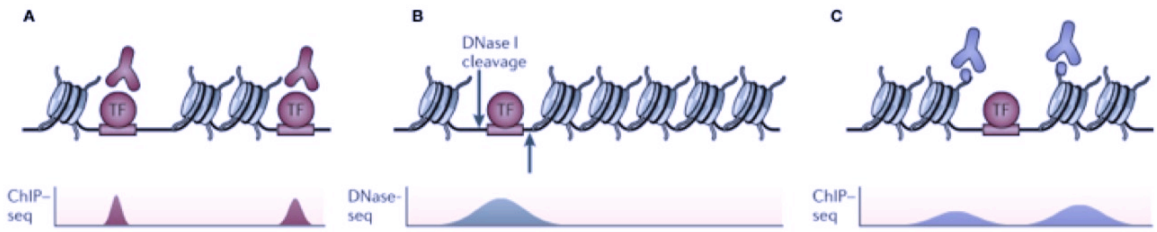


Figure 7 : Genomic methods for predicting enhancer elements. (A) ChIP–seq using antibodies against transcription factors (TF) is used in order to determine TF binding sites throughout the genome. However, TF binding sites do not always correspond to functional enhancers. (B) DNase-seq is utilized to identify active enhancers based on DNA accessibility after nucleosome removal and independently of TF binding. (C) Active enhancers are flanked by nucleosomes bearing characteristic histone marks that can be detected by ChIP–seq. Adapted from (Shlyueva et al., 2014).

Despite the fact that a genomic region has been suggested by DNase-seq or ChIP-seq to act as an enhancer, its physiological relevance needs to be established in the appropriate developmental context. Predicted enhancers should be further validated in functional assays. Enhancer sequences are frequently tested for their ability to activate transcription by simply cloning the candidate enhancer sequence upstream of a minimal promoter driving the expression of a reporter gene (such as lacZ, GFP or luciferase) and measuring reporter gene expression in cell lines or transgenic animal models. This activity is the hallmark property of enhancers (Banerji et al., 1981; Benoist & Chambon, 1981) and is broadly used to evaluate predicted enhancers (Heintzman et al., 2007; Rada-Iglesias et al., 2011; Visel, Blow, et al., 2009; Zinzen, Girardot, Gagneur, Braun, & Furlong, 2009) in transgenic D. melanogaster and

mouse embryos (Visel, Minovitsky, Dubchak, & Pennacchio, 2007). However, such classical plasmid-based approaches are mainly low throughput and are mainly used to test previously predicted enhancers (Heintzman et al., 2007; Rada-Iglesias et al., 2011; Visel, Blow, et al., 2009). Self-transcribing active regulatory region sequencing (STARR- seq) is a plasmid- based approach, whereby enhancer sequences mediate their own transcription (Arnold et al., 2013). This method exploits the fact that enhancers can exert control in a position- and orientation-independent manner (Banerji et al., 1983; Benoist & Chambon, 1981; Fromm & Berg, 1983; Moreau et al., 1981) and inserts the candidate elements downstream of a reporter gene driven by a minimal promoter (Arnold et al., 2013). STARR- seq is of special importance since it allows genome-wide screens of enhancer activity. More recently, CRISPR–Cas9-based approaches have been used to examine enhancer function (Fulco et al., 2016; Klann, Black, Chellappan, Safi, & Song, 2017). For example, targeting KRAB-dCas9 (KRAB effector domain fused to catalytically dead Cas9) (Thakore et al., 2015) to the MYC locus in K562 cells, identified seven putative MYC enhancers.

Despite this enormous progress, several seminal questions remain unanswered: notably, what percentage of enhancers predicted by sequence- and chromatin- based methods can truly function as enhancers *in vivo*. By using ChIP-STARR-seq in human ESCs, an assay that combines antibodies against TFs or histone marks with STARR-seq plasmid, it has been shown that only 12% of H3K27ac ChIP-seq peaks displayed enhancer activity (Barakat et al., 2018), which indicates that genome-wide assays might have overvalued the number of functional enhancers. In addition, a high-throughput CRISPR-Cas9–based approach was used to identify the regulatory elements controlling the expression of the ESC–specific genes, Nanog, Rpp25, Tdgf1 and Zfp42 (Rajagopal et al., 2016). Strikingly, the authors identified a novel class of cis-regulatory elements that did not coincide with any known markers of regulatory activity, including H3K27ac and H3K4me1. These findings suggest that, to identify enhancers, a more comprehensive analysis of histone marks is required than has previously been considered.

Notably, the same study (Rajagopal et al., 2016), and a CRISPR-Cas9-mediated deletion screening for *cis*-regulatory elements in the human *POU5F1* locus (Diao et al., 2017), showed that promoters can function as enhancers for neighboring genes. Actually, enhancers and promoters are highly related and show important similarities in structure and function (Boyle

et al., 2008; Koch et al., 2011; Thurman et al., 2012). Both these regulatory elements exhibit DNase I hypersensitivity, which results from the depletion of nucleosomes (Boyle et al., 2008; Thurman et al., 2012). Similar to promoters, enhancers are bound by general transcription factors and; cofactors, recruit RNAPII and initiate transcription (see below) (T. K. Kim et al., 2010; Koch et al., 2011). Nevertheless, the fact the promoters can behave as enhancer-like elements for neighboring genes (Dao et al., 2017; Diao et al., 2017; Rajagopal et al., 2016) adds additional complexity to the distinction of these two regulatory elements and argues that the classical definitions of the promoter and the enhancer might need to be revised.

Apart from that, the functional roles for most of the chromatin modifications associated with active enhancers are not fully understood. There is no clear evidence that histone marks such as H3K4me1 or H3K27ac are necessary or even mechanistically involved in transcription (Pengelly, 2013). Interestingly, loss of H3K4me1 from enhancers in a series of MLL3 and MLL4 (MLL3/MLL4) histone methyltransferase catalytic mutant mouse ESC lines leads to a partial reduction of H3K27ac from enhancers but has only a minor effect on transcription from either enhancers or promoters, implying that H3K4me1 is dispensable for maintenance of gene expression programs in mouse ESCs (Dorighi et al., 2017).

Enhancer RNAs

A great number of studies has demonstrated that enhancers become activated as transcribed units, adding a new layer of complexity to genome regulation. Evidence of enhancer-derived transcription was first documented in human erythroid cells, in the locus control region (LCR) that regulates the transcription of the downstream beta-globin gene clusters, located 10-50 kb away (Collis, Antoniou, & Grosveld, 1990; Ling et al., 2004; D. Tuan, Kong, & Hu, 1992). The LCR was shown to comprise five erythroid-specific DNase I hypersensitivity sites (HS) and most importantly, transcriptional initiation sites were found inside these regions (Collis et al., 1990; D. Tuan et al., 1992). A few years later, separate studies reported direct RNA Pol II recruitment and production of enhancer RNAs (eRNAs) from functionally active enhancers in several mammalian cell types, including cortical neurons and T cells (de Santa et al., 2010; T. K. Kim et al., 2010; Koch et al., 2011). RNA Pol II is recruited to cell-specific enhancers (Koch et al., 2011) and responds dynamically upon stimulus-induced activation (for example by

membrane depolarization or endotoxin activation) (de Santa et al., 2010; T. K. Kim et al., 2010).

As large numbers of eRNA transcripts have been detected, more comprehensive insights into their features and regulation have been gained. eRNA-producing enhancers display high levels of H3K4me1, H3K27ac, H3K79me2, low levels of H3K4me3, and are depleted for H3K27me3 (Djebali et al., 2012; Heintzman et al., 2009). eRNAs are generally not spliced and the majority, although not all, lack polyadenylated tails (de Santa et al., 2010; Djebali et al., 2012; T. K. Kim et al., 2010; Koch et al., 2011). By using genome-wide binding profiles, it has been demonstrated that eRNA-producing enhancers are extensively occupied by TFs, such as p53 (Melo et al., 2013) or (E2)-bound estrogen receptor α (ERα) in MCF-7 human breast cancer cells (W. Li et al., 2013), and that the expression levels of the produced eRNAs dynamically change in response to signal transduction events (de Santa et al., 2010; Hah et al., 2011; T. K. Kim et al., 2010; Lam, 2013; W. Li et al., 2013; Melo et al., 2013; Mousavi et al., 2013; D. Wang et al., 2011). Most importantly, signal-dependent changes in eRNA expression levels at enhancers are strongly correlated with changes at the level of mRNA synthesis at nearby genes (T. K. Kim et al., 2010; Lam, 2013; W. Li et al., 2013), suggesting that, at least for some enhancers, a stimulus-induced eRNA is linked to proper gene activation of neighboring coding genes. Collectively, enhancer transcription is a genome-wide phenomenon across multiple cell types which does not appear to be a random process resulting from the irregular/accidental binding of RNA Pol II to the open chromatin of enhancers (Hah et al., 2011; T. K. Kim et al., 2010; Rada-Iglesias et al., 2011). When considering the functional significance of enhancer transcription, two possible scenarios are emerging: first, the act of eRNA transcription, not the features of the eRNA transcript itself, have a specific activating function; second, the eRNA *per se* plays a functional role in regulating gene expression. In support of the latter model, artificial recruitment of eRNA to promoters (Melo et al., 2013) or enhancers (Li et al., 2013) stimulated reporter transcription, and reporter activity is reduced when the eRNA-producing sequence is inverted, altering the sequence of the eRNA without affecting binding sites for TFs (Lam et al., 2013). However, these experimental strategies have only been applied to reporter systems, and the relative contribution of endogenous enhancer transcription and/or eRNA binding at promoters to transcriptional regulation is poorly characterized and may be context-dependent.

Super-Enhancers

As previously noted, mammalian genomes contain a great amount of active transcriptional enhancers that can be detected based on enhancer-specific chromatin features like chromatin accessibility, the binding of cell-type specific transcription factors and the accumulation of active histone marks, notably H3K27ac (Long, Prescott, & Wysocka, 2016). Such studies resulted in the identification of a new class of regulatory elements, the so-called "super-enhancers" (Hah et al., 2015; Hay et al., 2016; Hnisz et al., 2013; Lovén et al., 2013; H. Y. Shin et al., 2016; Whyte et al., 2013; Zhou et al., 2014). Super-enhancers (SEs) share many features with typical enhancers but in a considerably larger scale: they are typically an order of magnitude larger than typical enhancers and can be distinguished by the very dense accumulation of master regulators, the Mediator co-activator complex, p300 and H3K27ac (Hnisz et al., 2013; Kagey et al., 2010; Lovén et al., 2013; Whyte et al., 2013) (**Figure 8**). Super-enhancers exhibit high levels of RNA Pol II binding and produce quite high amounts of eRNAs compared to typical enhancers (Hah et al., 2015; Hnisz et al., 2013, 2015) which are dynamically regulated upon cellular signaling (Hah et al., 2015). Recent studies suggest that eRNAs may be involved or required for gene activation (Hah, Murakami, Nagari, Danko, & Lee Kraus, 2013; W. Li et al., 2013; Mousavi et al., 2013), therefore high-levels of SE transcripts, in concert with master TFs, could contribute to the transcriptional activation of their associated genes and the establishment of a cell-type specific regulatory network.

Figure 8 : Schematic representation of a typical vs. super-enhancer. The higher accumulation of activators at super-enhancers is believed to contribute to higher transcriptional activity (TA). Adapted from (Hnisz, Shrinivas, Young, Chakraborty, & Sharp, 2017).

Various studies in a wide variety of mammalian cell types, have reported the association of super-enhancers with highly cell-type specific genes known to play major roles in their biological identities (Hay et al., 2016; Hnisz et al., 2013; Lovén et al., 2013; H. Y. Shin et al., 2016; Vahedi et al., 2015; Whyte et al., 2013). Given the importance of super-enhancers in the control of cell identity, it is not surprising that key oncogenes and other important genes for tumorigenesis reside in enhancer-dense genomic regions (Hnisz et al., 2013, 2015; Lovén et al., 2013; Mansour et al., 2014) and disease-associated sequence variation has been observed within super-enhancers of disease-relevant cell types (Hnisz et al., 2013). It is also interesting to mention that although the definition of SEs is recent, locus control regions (LCRs) characterized in early studies, such as the LCR of β-globin (Grosveld, van Assendelft, Greaves, & Kollias, 1987) and the enhancer element of the immunoglobin heavy chain (Banerji et al., 1983; Gillies et al., 1983; Mercola et al., 1983), also correspond to this class of regulatory element. Collectively, these results suggest that super-enhancers have evolved to robustly control genes that play critical roles in cellular physiology and disease (Hay et al., 2016; Hnisz et al., 2013, 2015; Lovén et al., 2013; Mansour et al., 2014; Whyte et al., 2013). Despite the fact that super-enhancers have been identified in a plethora of different cell types, there are conflicting results as to whether SEs represent a new category of regulatory elements, different from conventional enhancers, and on the importance of individual elements within them (Hay et al., 2016; Moorthy et al., 2017; H. Y. Shin et al., 2016; Thomas et al., 2021). At some loci, each constituent element contributes independently and in an additive rather than synergistic fashion to gene expression (Hay et al., 2016). At other loci, some constituent enhancers were shown to be more essential than others and to control the activation of other elements within the same SE in order to secure full expression of the target genes (Hnisz et al., 2015; H. Y. Shin et al., 2016; Thomas et al., 2021).

Multiple enhancers promote developmental robustness

With our understanding of the complexity of genome organization, it is becoming evident that many enhancers do not function in isolation. Numerous studies have demonstrated that developmental genes are controlled by multiple enhancers with both overlapping and separate spatial and temporal activities (Frankel et al., 2010; Hay et al., 2016; Hong, Hendrix, & Levine, 2008; Marinić, Aktas, Ruf, & Spitz, 2013; Osterwalder et al., 2018). This finding raises two

important questions: What are the relationships between simultaneously active enhancers? What is the purpose of multiple-enhancer regulation of single genes?

Multiple enhancers can cooperate in the same tissue, where they produce additive, synergistic or redundant interactions to drive high levels of expression of their shared target gene (Bender et al., 2001; Frankel et al., 2010; Hay et al., 2016; Hong et al., 2008; Long et al., 2020; Marinić et al., 2013; Perry, Boettiger, Bothma, & Levine, 2010; H. Y. Shin et al., 2016) (**Figure 9A**). Some characteristic examples in this category are the α- and β-globin genes (Bender et al., 2001; Hay et al., 2016). Interestingly, genome-wide profiling of promoter contacts in human and mouse cells with high-resolution promoter-capture HiC, showed that the number of promoter-interacting enhancers is positively related with the level of gene expression (Mifsud et al., 2015; Schoenfelder et al., 2015), which supports the idea that additive enhancers drive higher levels of expression compared to individually acting enhancers. Another mode of function in gene activation is enhancer redundancy, as has been described for some "shadow enhancers" in Drosophila melanogaster (Frankel et al., 2010; Hong et al., 2008; Perry et al., 2010; Perry, Boettiger, & Levine, 2011). By performing whole-genome ChIP-chip assays in the early Drosophila embryo (Hong et al., 2008), Mike Levine and colleagues discovered some remote "secondary" enhancers that appear to produce similar patterns of expression to these generated by "primary" enhancers. The overlapping function of primary and shadow enhancers buffers the expression of critical patterning genes and confers robustness against environmental and genetic variation, which is a common feature of natural populations (Frankel et al., 2010; Perry et al., 2010). Importantly, redundancy between enhancers could explain why deletions of some *cis*-regulatory elements lead to dramatic phenotypes (Bahr et al., 2018; Gonen et al., 2018; Moorthy et al., 2017; Sagai et al., 2005; Zhou et al., 2014), while others show weak or no effects (Moorthy et al., 2017; Osterwalder et al., 2018).

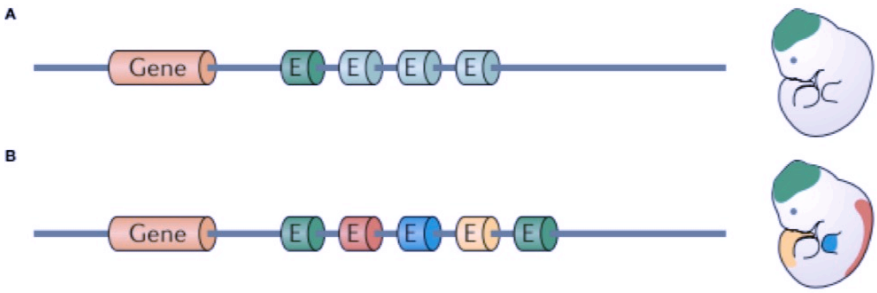


Figure 9 : Gene expression regulation by multiple enhancers. (A) Multiple enhancers cooperating in one tissue to increase the expression of a shared gene. (B) Combinations of two or more enhancers (depicted by different colors) control the expression of a shared key developmental gene in a tissue-specific manner. Adapted from (Krijger & De Laat, 2016).

Alternatively, genes with a complex tissue-specific developmental expression pattern, such as *Shh* (Jeong, El-Jaick, Roessler, Muenke, & Epstein, 2006; Lettice et al., 2003; Sagai et al., 2009), *Myc* (Uslu et al., 2014) and the *Hox* gene clusters (Andrey et al., 2013; Montavon et al., 2011), are controlled by multiple enhancers with non-redundant spatiotemporal activities spread over tens to hundreds of kilobases upstream or downstream of the target gene, facilitating vast combinatorial complexity of gene-expression programs (**Figure 9B**). Together, these studies highlight a key role of multiple enhancers in guaranteeing precise fine-tuning of gene expression during development.

Role of enhancers in disease

The release of the human genome sequence in combination with genome-wide association studies (GWAS) and chromatin signatures of enhancers have revealed that many disease- and trait- genetic variants systematically lie within enhancers (Dunham et al., 2012; Maurano et al., 2012; Roadmap Epigenomics Consortium et al., 2015). Indeed, it has been demonstrated that non-coding mutations and genome rearrangements that perturb enhancer-promoter communication can underlie disease susceptibility and developmental abnormalities (Bahr et al., 2018; Benko et al., 2009; Bhatia et al., 2013; Franke et al., 2016; Gonen et al., 2018; Kioussis, Vanin, Delange, Flavell, & Grosveld, 1983; Lettice et al., 2003, 2012; Long et al., 2020; Lupiáñez, Kraft, Visel, & Mundlos, 2015; Mansour et al., 2014; Sagai et al., 2005; Uslu et al., 2014). Mutations in enhancers and non-coding DNA can provoke severe Mendelian and common genetic disease, a class of conditions referred as “enhanceropathies” (Smith & Shilatifard, 2014). The deleterious effects of disruptions in enhancers can result from deletions, duplications, rearrangements, inversions or point mutations. In the present section, I discuss several examples in which distant-acting gene enhancers were directly shown to play a causative role in disease (**Table 1**).

SOX9 is a pivotal regulator of morphogenesis in the developing embryo, particularly in craniofacial development, chondrogenesis and sex determination (Lee & Saint-Jeannet, 2011). Heterozygous loss-of-function mutations in the SOX9 gene cause campomelic dysplasia (CD), a severe, often lethal, congenital dysplasia characterized by bowing of the long bones of the legs, sex reversal and craniofacial defects (Wagner et al., 1994). Pierre Robin sequence (PRS) is a congenital craniofacial disorder associated with micrognathia, glossoptosis and upper airway obstruction (Robin, 1994). Because of the phenotypic overlap between the craniofacial malformations of campomelic dysplasia and Pierre Robin sequence, and the existence of several non-coding mutations, including translocations and microdeletions, within the 2-Mb gene desert upstream of SOX9, it has been hypothesized that PRS may be caused by dysregulation of SOX9 (Amarillo, Dipple, & Quintero-Rivera, 2013; Bagheri-Fam et al., 2006). Indeed, it has been subsequently shown that Pierre Robin sequence is caused by developmental misexpression of SOX9 due to deletion of extreme long-range tissue-specific enhancers, located 1.25 and 1.45 Mb centromeric of SOX9 (Benko et al., 2009; Long et al., 2020).

As mentioned earlier, key developmental genes might have functions in more than one tissue and be surrounded and regulated by multiple tissue-specific cis-regulatory elements, spread over hundreds of kilobases from the TSS. In mammals, *Sox9* is a direct target of SRY (Sex-determining Region of Y chromosome), a pivotal factor for testis and subsequent male development (Capel, 2017). An *in vivo* screen for gonad enhancers in the 2-Mb gene desert upstream of *Sox9,* uncovered a regulatory element, the enhancer 13 (Enh13), located 565 Kb upstream of the TSS which appeared to be essential for initiating mouse testis development (Gonen et al., 2018). Surprisingly, while XY embryos carrying a heterozygous deletion of Enh13 (Enh13$^{+/-}$) undergo normal testis development, XY Enh13$^{-/-}$ embryos produce fully sex-reversed ovaries identical to those of XX wild-type embryos, demonstrating that deletion of a single distal-enhancer results in sex reversal (Gonen et al., 2018).

Deletions of key tissue-specific enhancers have been shown to be causative for several other disorders, including: cleft lip with or without cleft palate (CL/P) craniofacial malformation resulting from deletion of a distal enhancer controlling *Myc* expression in the future upper lip (Uslu et al., 2014); truncated limbs in mice resulting from deletion of a limb-specific *Shh* enhancer (Sagai et al., 2005) (**Figure 10**) and β-thalassemia due to the inability of red blood cells to produce mature globin after deletion of the hemoglobin subunit beta (HBB) LCR (Kioussis et al., 1983; D. Y. H. Tuan, Solomon, London, & Lee, 1989).

Besides the pathological consequences of the removal of enhancer elements, repositioning of distal enhancers can lead to human disorders. A representative example is α-thalassaemia, a haemoglobinopathy -like β-thalassemia- caused by imbalances in the ratio of α-globin in erythroid cells. While many patients diagnosed with α-thalassaemia were carrying globin-coding mutations (Kan et al., 1975), a subset of individuals did not exhibit any abnormalities in the protein-coding sequence. It was eventually identified that a single-nucleotide polymorphism (SNP) in the non-coding region between the α-globin genes and their cognate enhancer, creates a novel binding site for the erythroid TF GATA1, forming a cryptic promoter (De Gobbi et al., 2006). This new element interferes with normal activation of α-globin, by reallocating the enhancer away from α-globin genes.

Point mutations in enhancers can also produce disease, considering that enhancers are the elements where tissue-specific TFs bind (Lambert et al., 2018). Among others, aniridia (Bhatia et al., 2013; Lauderdale, Wilensky, Oliver, Walton, & Glaser, 2000) and preaxial polydactyly (Lettice et al., 2003, 2012) are of great importance. Aniridia is a congenital eye disorder caused by haploinsufficiency of the developmental regulator PAX6, predominantly through loss-of-function mutations or gene deletions and in some cases due to disruption of the genomic region downstream of *PAX6* (Lauderdale et al., 2000). A *de novo* single point mutation (chr11: 31,685,945 G->T) in an ultra-conserved enhancer (SIMO) located 150 kb telomeric of *PAX6* was identified in an affected individual (Bhatia et al., 2013). SIMO was shown to be critical for *PAX6* expression in developing ocular tissues as the single point mutation in this element was sufficient to disrupt an autoregulatory PAX6 binding site, cause loss of enhancer activity and deficient maintenance of PAX6 expression (Bhatia et al., 2013). In the same line, point mutations residing within the limb-specific enhancer ZRS, located approximately 850 kb away from the *Shh,* lead to a congenital limb malformation called preaxial polydactyly (PPD) (Lettice et al., 2003, 2012) (**Figure 10**). These point mutations cause aberrant expression of *Shh*, normally expressed in the zone of polarizing activity (ZPA) posteriorly in the limb bud, in an additional ectopic site at the anterior margin of the limb.

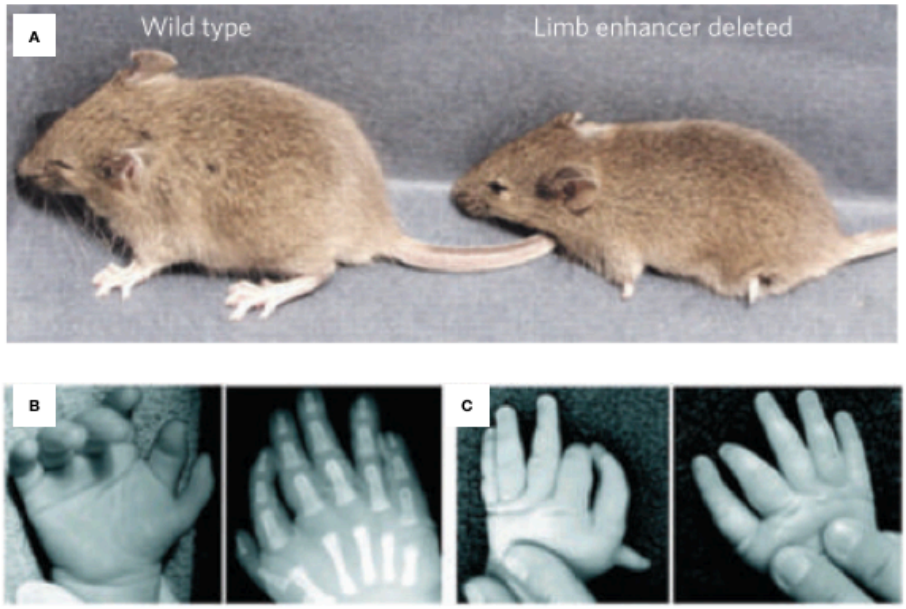


Figure 10: Deletion and mutation of the limb enhancer of *Shh* (ZRS) lead to limb malformations. (A) Mice with deletion of the ZRS have gravely truncated limbs (Sagai et al., 2005). (B-C) Hands of two patients with point mutations in the ZRS (Lettice et al., 2003). Adapted from (Visel, Rubin, & Pennacchio, 2009).

Structural rearrangements and duplications of enhancers can also provoke disease. Chromosomal rearrangement of a distal GATA2 enhancer causes dysregulation of two unrelated genes, *EVI1* and *GATA2*, with acute myeloid leukemia (AML) as the outcome. Removal of the distal enhancer from its genomic origin caused aberrant activation of *EVI1* and simultaneous transcriptional impairment of *GATA2* (Gröschel et al., 2014). Lastly, chromosomal duplications of a T cell-specific distal MYC enhancer, named N-Me (from NOTCH *MYC* enhancer) located 1.47 Mb downstream from *MYC* TSS, are directly implicated in the pathogenesis of human T cell acute lymphoblastic leukemia (T-ALL) (Herranz et al., 2014). Together, these results underscore the essential role of enhancer elements in development and further demonstrate that disruption of a single cis-element is sufficient to cause aberrant gene expression and disease phenotypes.

Table 1: Improper enhancer-promoter wiring causes disease.

Disease	Disruption	Affected gene	Enhancer	Reference
Pierre Robin sequence (PRS)	Enhancer deletion	*SOX9*	3' ~1.5 Mb enhancers	Benko *et al.*, 2009; Long *et al.*, 2020
Sex reversal	Enhancer deletion	*Sox9*	Enhancer 13	Gonen *et al.*, 2018
Facial dysmorphology	Enhancer deletion	*Myc*	Medionasal enhancer (MNE) region	Uslu *et al.*, 2014
Truncated limbs	Enhancer deletion	*Shh*	ZRS (referred as MFCS1)	Sagai *et al.*, 2005
β-thalassemia	Enhancer deletion	β-globin genes	LCR	Kioussis *et al.*, 1983; Tuan *et al.*, 1989
α-thalassaemia	Promoter introduction	α -globin genes	α -globin enhancers	De Gobbi *et al.*, 2006
Aniridia	Point mutation	*PAX6*	SIMO	Bhatia *et al.*, 2013;
Preaxial polydactyly (PPD)	Point mutation	*Shh*	ZRS	Lettice *et al.*, 2003, 2012
Acute myeloid leukemia (AML)	Enhancer Rearrangement	*EVI1* and *GATA2*	GATA2 enhancer	Groschel et *al.*, 2014
T cell acute lymphoblastic leukemia (T-ALL)	Enhancer duplication	*MYC*	N-Me (NOTCH *MYC*) enhancer	Herranz et *al.*, 2014

Other distal regulatory elements

Enhancers are the most extensively studied example of distal regulatory elements (DREs), however other types of *cis*-regulatory modules (CRMs) have also been identified. Silencers can act in an orientation-independent manner, similarly to enhancers, however they downregulate, rather than activate, target gene expression (Brand, Breeden, Abraham, Sternglanz, & Nasmyth, 1985), probably via binding transcriptional repressors (Shore, Stillman, Brand, & Nasmyth, 1987). Although examples of silencers have been reported in the literature for more than 30 years (Brand et al., 1985; Cao, Gutman, Dave, & Schechter, 1989), their mechanism of function and characteristic sequence and chromatin features, if any, are not well understood. Recent highly parallel reporter-based assays have augmented the number of identified putative silencer elements, both in mammals and in Drosophila (Doni Jayavelu, Jajodia, Mishra, & Hawkins, 2020; Gisselbrecht et al., 2020). Interestingly, despite the fact that enhancers and silencers are being treated as two distinct groups of CRMs, bifunctional

elements that can act as both enhancers and silencers have been discovered (Gisselbrecht et al., 2020; Jiang, Cai, Zhou, & Levine, 1993; Rafael Galupa et al., 2020). In some cases, this behavior has been shown to depend on the cellular context (Gisselbrecht et al., 2020; Jiang et al., 1993), while in others it is dependent on the TAD (Topologically associating domain) (see Chapter 4) in which the CRM resides (Rafael Galupa et al., 2020). Additional studies are required to uncover the mechanistic basis of silencers, and if and to what extent they coordinate with enhancers. Another type of *cis*-regulatory modules are insulators which are covered in a later section of the introduction.

Enhancer – promoter communication

Studies of several loci suggest that transcription in higher eukaryotes is often regulated by remote enhancer sequences, that can be positioned from a few up to hundreds of kilobases away from promoters (Banerji et al., 1981; Bender, Bulger, Close, & Groudine, 2000; Benko et al., 2009; Grosveld et al., 1987; Uslu et al., 2014). Thus, communication between these distal elements is requisite for transcriptional regulation. Two basic models have been put forward to describe how enhancers and promoters communicate: (A) the tracking model and (B) the looping model (**Figure 11**).

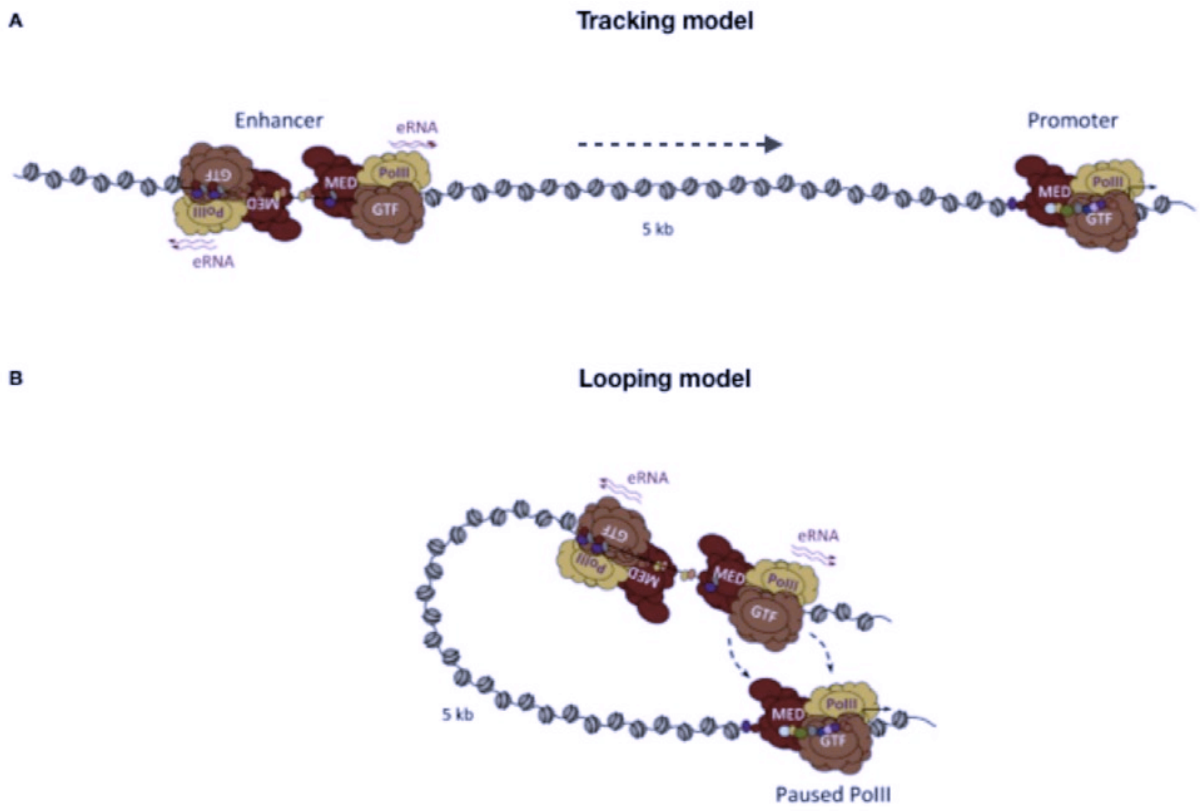


Figure 11: Models of enhancer-promoter communication. (A) Tracking and (B) Looping model. Adapted from (Vernimmen & Bickmore, 2015).

The tracking model states that enhancers act as nucleation points for factors which move progressively towards the promoter, through the intervening DNA sequence separating the two regulatory elements. On the other hand, the looping model proposes a direct, physical interaction between distant enhancers and promoters, with the intervening DNA sequence

looping out. For example, it has been shown in *in vitro* experimental systems that enhancers are capable of stimulating transcription in *trans* when brought into close proximity to a promoter-containing plasmid (De Bruin, Zaman, Liberatore, & Ptashne, 2001; Müller, Sogo, & Schaffner, 1989) or that the GAGA factor can form a protein link when bound to separate DNA molecules in *trans* (Mahmoudi, Katsani, & Verrijzer, 2002). Carter et *al*. (Carter, Chakalova, Osborne, Dai, & Fraser, 2002) developed an *in situ* technique termed, RNA TRAP (tagging and recovery of associated proteins), to tag and recover chromatin in the close vicinity of a transcriptionally active gene. By applying this method at the *Hbb* locus (encoding β-globin) in mouse erythroid fetal liver cells, Carter and colleagues demonstrated that the LCR of the β-globin gene cluster is in close physical proximity to the actively transcribing β-globin gene, located over 50 kb away, that it regulates *in vivo*. These results represented the first direct evidence of long-range enhancer-promoter communication and provided support for the looping model. Shortly afterwards, Tolhuis et *al*. (Tolhuis et al., 2002) used 3C (Capturing chromosome conformation) analysis (Dekker, Rippe, Dekker, & Kleckner, 2002) to demonstrate that this LCR comes in close spatial proximity with the active β-globin genes in the expressing erythroid cells, while a linear type of structure was found in non-expressing brain cells. It is worth mentioning that the communication between distal enhancers and promoters through the formation of chromatin loops does not exclude a scenario in which enhancers are communicating through a tracking mechanism (Hatzis & Talianidis, 2002; Vernimmen & Bickmore, 2015; Q. Wang, Carroll, & Brown, 2005).

Enhancer action through chromatin loops

As discussed above, strong evidence supports the idea that spatial proximity between enhancers and promoters can increase promoter activity (Carter et al., 2002; Tolhuis et al., 2002). Ever since, a body of data supports loop models, using microscopy techniques, such as fluorescent *in situ* hybridization (FISH), or molecular biology techniques, such as proximity ligation (3C-derived) methods (Bartman et al., 2016; Bonev et al., 2017; H. Chen et al., 2018a; Deng et al., 2012, 2014; Dixon et al., 2015; Fulco et al., 2019; Gasperini et al., 2019; Ghavi-Helm et al., 2014; Palstra et al., 2003; Phillips-Cremins et al., 2013; Zhou et al., 2014). Key developmentally regulated genes, especially those that are expressed at high levels, appear to interact with enhancers specifically in the corresponding cell type (Palstra et al., 2003; Phillips-

Cremins et al., 2013). A representative example is the *Sox2* gene where a distal enhancer region (Sox2 control region, SCR) located >100 kb downstream of *Sox2* is essential for *cis*-regulation of *Sox2* specifically in ES cells (Zhou et al., 2014). 3C and 4C analysis revealed that the distal SCR contacts the promoter through the formation of a chromatin loop (Ben Zouari, Platania, Molitor, & Sexton, 2020; Zhou et al., 2014). Interestingly, homozygous Cas9-mediated deletions of this SCR resulted in significant reduction in Sox2 mRNA and protein levels, loss of ES colony features, changes in gene expression on the genome-wide level and decreased neuroectodermal formation upon differentiation. Notably, 5C (Phillips-Cremins et al., 2013) and ultra-high resolution HiC (Bonev et al., 2017) data observed a loss of the ES-specific *Sox2*-SCR looping interaction upon neural progenitor cell (NPC) differentiation and gain of a new contact between the *Sox2* promoter and a more distal (~450 kb downstream) NPC-specific enhancer.

Direct proof of the functional importance of enhancer-promoter contacts also comes from studies that manipulated the chromatin conformation at a native environment (Bartman et al., 2016; Deng et al., 2012, 2014). The chromatin loop between the LCR and β-globin (Hbb) promoter requires the erythroid-specific TF GATA1 and the adaptor protein Ldb1 (LIM domain binding protein 1), which is recruited to DNA via GATA1 and the transcription factors TAL1, LMO2 and E2A. In a seminal study, Deng *et al*. (Deng et al., 2012) used GATA1-null G1E erythroid cells in which an LCR-globin loop is absent and globin genes are inactive. In the absence of GATA1, Ldb1 is no longer recruited to the β-globin promoter but it can still bind to the LCR through the other TFs. Artificial zinc finger (ZF) tethering of the self-association domain (SA) of Ldb1 to the β-globin promoter in G1E cells induced a chromatin loop which was sufficient to induce robust transcriptional activation of β-globin, even in the absence of GATA1, directly demonstrating that enhancer-promoter juxtaposition can activate transcription (**Figure 12A**).

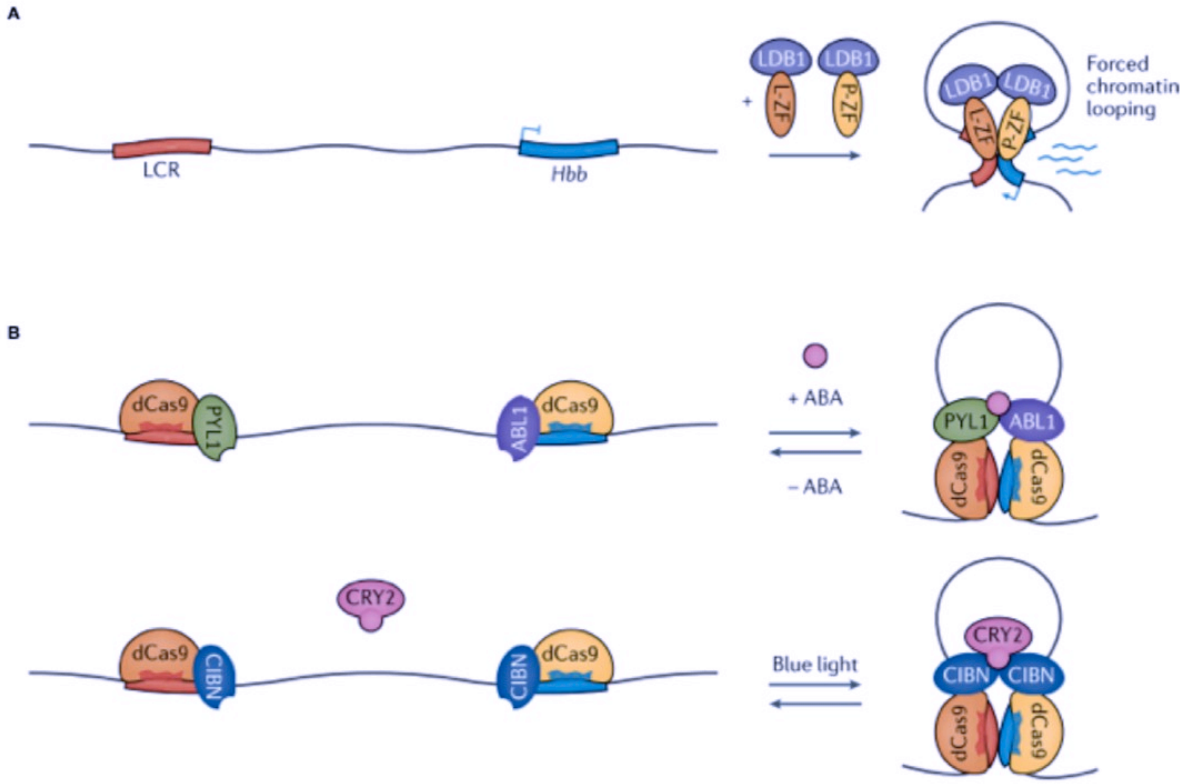


Figure 12: Enhancer-promoter looping in transcriptional control. (A) Forced looping between the β- globin (Hbb) promoter and its enhancer (LCR) via zinc finger (ZF) tethering of Ldb1 (LIM domain binding protein 1) to the Hbb promoter was sufficient to increase β-globin expression, even in the absence of the erythroid-specific transcription factor GATA1 (Deng et al., 2012). (B) Variants of forced chromatin looping use catalytically dead Cas9 (dCas9) fusion proteins to engineer inducible and reversible contacts between enhancers and promoters, which led to increased gene expression (J. H. Kim et al., 2019; Morgan et al., 2017). Adapted from (Schoenfelder & Fraser, 2019).

Follow-up studies showed that forced chromatin looping between the LCR and a developmentally silenced embryonic globin gene (γ-globin) in adult erythroid cells leads to strong transcriptional activation at this locus by increasing the burst fraction (Bartman et al., 2016; Deng et al., 2014). Additionally, two recent strategies of forced chromatin looping direct dCas9 (catalytically dead Cas9) fusion proteins to target genomic loci, which can be brought into close proximity either by small-molecule (Morgan et al., 2017) or light-induced (J. H. Kim et al., 2019) (**Figure 12B**) dimerization to reversibly establish chromatin loops and induce gene expression. Recently, enhancers have also been implicated in the control of transcriptional bursting parameters (Bartman et al., 2016; H. Chen et al., 2018b; Fukaya et al., 2016; Lim,

Heist, Levine, & Fukaya, 2018). In particular, recruitment of the LCR to a chosen β-type globin gene promoter, increased β-globin burst frequency but not burst size (Bartman et al., 2016). In the same line, live-cell visualization of enhancer-promoter interactions in *Drosophila* embryos revealed that sustained enhancer-promoter association is necessary for the initiation of transcriptional bursts (H. Chen et al., 2018b). Moreover, other studies reported that a shared enhancer is able to simultaneously activate linked reporter genes in *cis* and in *trans*, resulting in coordinated bursting of these genes (Fukaya et al., 2016; Lim et al., 2018). Overall, these studies fit with a model in which physical contacts are necessary for enhancers to activate transcription.

Potential role of eRNAs in regulating gene expression via chromatin loops

Chromatin interaction studies held in several cell lines showed a much greater abundance of eRNAs at enhancers that interacted with promoters of protein-coding genes compared to those that did not (Y. C. Lin et al., 2012; Sanyal et al., 2012). In support of that, Li *et al*. (W. Li et al., 2013) used both short interfering RNAs (siRNAs) and locked nucleic acid antisense oligonucleotides (LNAs) against eRNAs produced by (E2)-bound estrogen receptor α (ERα) enhancers, in order to explore their potential roles on gene expression events. Interestingly, siRNA/LNA-mediated knockdown of eRNAs caused a significant inhibition of gene expression and enhancer–promoter specific interactions induced by E2. Furthermore, knockdown of these eRNAs resulted in a decrease of cohesin recruitment to several ERα-bound enhancers, while knockdown of RAD21 (a subunit of cohesin) inhibited both gene induction and the observed induced enhancer–promoter loops. These data suggest that besides cohesin, eRNAs are also likely to have important functions in the formation or the stabilization of enhancer-promoter loops.

By contrast, when MCF7 cells were treated with flavopiridol, a drug that blocks transcription elongation by inhibiting the CDK9 kinase of the P-TEFb, Hah et al. (Hah et al., 2013) found that although the production of eRNA and gene transcripts was reduced, the estradiol-induced enhancer-promoter looping was not affected. This difference could be due to the different experimental approach used (eRNA knockdown (W. Li et al., 2013)/ inhibition of PolII

elongation (Hah et al., 2013)), or it may indicate that eRNAs are functionally important for chromatin looping only at specific gene loci.

Transcription hubs

Although population-average "C" methods support a model of chromatin loops bringing individual enhancers into contact with their cognate promoter (Palstra et al., 2003), this is challenged by recent findings. Recent imaging experiments have identified, at least for some specific loci, that enhancer-promoter proximity can be decoupled from, or even negatively correlate with gene expression (Alexander et al., 2019; Benabdallah et al., 2019; Espinola et al., 2021; Heist, Fukaya, & Levine, 2019). So how do enhancers regulate specific genes from a distance? Large multi-enhancer/promoter hubs have been discovered (Allahyar et al., 2018; Oudelaar et al., 2018), which sometimes (but not always) co-associate with nuclear foci - or clusters - of transcription factors (Dufourt et al., 2018; Jieru Li et al., 2020). Binding of TFs on DNA depends on TF's affinity for target sequences, the concentration, as well as the localization of the TF in the nuclear space (Lambert et al., 2018). Such hubs have been proposed to create local microenvironments, perhaps via the formation of liquid-liquid phase-separated (LLPS) condensates (Sabari et al., 2018), whereby regulatory factors can be rapidly exchanged within a high local concentration without direct juxtaposition of promoters and enhancers (Mir et al., 2018; Tsai et al., 2017) (**Figure 13**). Recent analysis of the activity of linked reporter genes using quantitative live-imaging further strengthens this revised hub model, by showing that a single enhancer is able to simultaneously coactivate two promoters linked in *cis* and in *trans* (Fukaya et al., 2016; Lim et al., 2018). These results oppose the classic enhancer-promoter looping model and suggest that target promoters can share common clusters of TFs and Pol II (Lim & Levine, 2021).

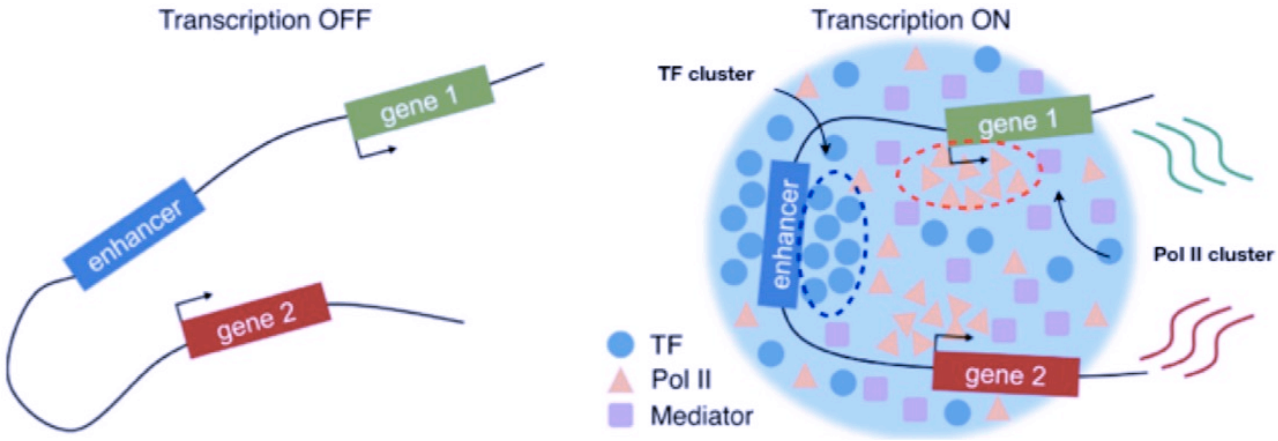


Figure 13: Formation of transcription hub as a possible mechanism for robust regulation of gene expression. TFs, Pol II, Mediators and other regulatory factors form clusters, perhaps via liquid-liquid phase-separation (LLPS), that facilitate enhancer-promoter interactions for transcriptional activation without direct juxtaposition of enhancers and promoters. Adapted from (Lim & Levine, 2021).

Transcription factor and RNA Pol II clusters

Transcription factors (TFs) are capable of modifying gene-expression levels through recognition and binding to specific DNA sequences at enhancer and promoter elements (Lambert et al., 2018). Early FRAP experiments showed that transcription factors display fast recovery after photobleaching, suggesting that TFs diffuse rapidly throughout the nucleus, allowing them to sample the genome for specific binding sites (McNally, Müller, Walker, Wolford, & Hager, 2000; Stenoien et al., 2001). Single-molecule tracking (SMT) experiments permitted to study the binding and diffusion dynamics of TFs in the nucleus, directly demonstrating that TF-chromatin interactions are characterized by high off-rates, with most TFs dynamically binding and dissociating from their target regions (J. Chen et al., 2014; S. Kim & Shendure, 2019; Z. Liu & Tjian, 2018; Mir et al., 2017; Swinstead et al., 2016). For example, Sox2 was found to spend most of its time (>95%) in a stochastic diffusive state or to be involved in non-specific DNA interactions. Only a small fraction of Sox2 molecules at any time was bound to specific recognition sites and even then dwell-times are on the scale of few (~12) seconds (J. Chen et al., 2014). Since transcription factors show high turnover, a growing

view is that temporal occupancy at the target site is sustained by high local concentrations -clusters- of TFs (J. Chen et al., 2014; Lim & Levine, 2021) (**Figure 13**). Such TF clusters have been identified in several systems, including Drosophila embryos (Bicoid, Zelda) (Dufourt et al., 2018; Mir et al., 2017; Yamada et al., 2019), yeast (Mig1)(Wollman et al., 2017) and ESCs (Oct4, Sox2) (Lim & Levine, 2021; Z. Liu et al., 2014).

Similarly, recent studies using super-resolution imaging on living cells revealed that RNA Pol II forms transient clusters (**Figure 13**), with an average lifetime of ~8 seconds (Cho et al., 2016; Cisse et al., 2013), which contradicts earlier fixed cell studies supporting the existence of Pol II clusters as statically pre-assembled architectures, named "transcription factories" (Iborra, Pombo, Jackson, & Cook, 1996). These dynamics were considerably faster than the period needed to complete transcription of a mammalian gene. Indeed, Pol II clusters were found to be more pronounced upon treatment with flavopiridol, a known P-TEFb inhibitor (Chao & Price, 2001), suggesting a role of these clusters for transcription initiation rather than productive elongation (Cisse et al., 2013). A follow up study demonstrated that transient Pol II clusters positively correlate with the number of nascent mRNAs subsequently synthesized (Cho et al., 2016). Treatment of cells with DRB, another P-TEFb inhibitor (Chodosh, Fire, Samuels, & Sharp, 1989) induced stable Pol II clusters similar to those observed upon flavopiridol treatment. Interestingly, upon removal of DRB transcriptional bursting resumed, suggesting that Pol II clusters are loaded on gene promoters to form and regulate transcriptional bursts (Cho et al., 2016). Pol II contains an intrinsically disordered C-terminal domain (CTD) (Portz et al., 2017). In some cases, the CTD has been found to undergo liquid-liquid phase separation (Boehning et al., 2018) while in others it does not (Lu et al., 2018). However, CTD was found to interact with kinase cyclin T1, a subunit of P-TEFb, droplets (Lu et al., 2018) and the Mediator complex (Chong et al., 2018), forming large clusters at sites of active transcription. Additionally, phosphorylation of CTD liberates Pol II from hubs and enables the transition into active elongation (Boehning et al., 2018; Lu et al., 2018). However, additional studies are necessary to get more insights into this process.

Enhancer hubs

The recently developed Multi-contact 4C (MC-4C)(Allahyar et al., 2018) and Tri-C (Oudelaar et al., 2018) methods were used to study multi-way contacts at the β-globin locus and were able to identify interaction patterns suggestive of the formation of enhancer hubs. Highly interacting enhancers that share binding by specific sets of transcription factors have been identified in several other loci (de Wit et al., 2013; Denholtz et al., 2013; Petrovic et al., 2019), colocalizing in hubs that present high occupancy by those factors (e.g., Oct4-Nanog) (de Wit et al., 2013). How is gene regulation achieved by the local environment? It is tempting to speculate that, inside these hubs specialized nuclear foci -or clusters- with a characteristic composition of regulatory and transcription factors are formed, whereby individual *cis*-regulatory elements aggregate and switch on multiple genes simultaneously, in a cooperative rather than competitive manner, to establish robust gene expression patterns. Whether enhancer hubs represent actively transcribing regions remains largely obscure, but colocalization between enhancers, regulatory factors (RFs), Pol II and nascent transcription would be consistent with such a scenario (S. Kim & Shendure, 2019; Lim & Levine, 2021).

Initial supportive evidence comes from single-molecule live imaging of clustering of the TF Bicoid in early Drosophila embryos (Mir et al., 2017), however such ideas have mostly been hypothetical since clustering could not be assigned to specific genes. More direct evidence comes from single-gene and single-molecule optical nanoscopy techniques which enabled simultaneous imaging and tracking of regulatory factors (e.g., Sox2, Brd4 and Mediator), Pol II, nascent transcription and *cis*-regulatory elements at target key pluripotency genes (*Pou5f1*, *Sox2* and *Nanog)* in mESCs (J. Li et al., 2019; Jieru Li et al., 2020). These imaging experiments uncovered frequent proximity of enhancers to the target gene within ~100- to 200-nm-sized clusters, composed of ~10-30 RF molecules, within the vicinity of the TSS, thus linking multi-enhancer clusters with RF accumulation and transcriptional activation: Enhancer hubs locally concentrate RNA Pol II RFs and create local microenvironments that can be sampled by multiple gene promoters (J. Li et al., 2019; Jieru Li et al., 2020). Surprisingly, when another imaging-based technique, called Hi-M (Cardozo Gizzi et al., 2019), was applied to two clusters of key developmental genes in *Drosophila* embryos, multi-enhancer hubs were found to frequently arise, independently of transcriptional activity, and with rare engagement of promoters (Espinola et al., 2021). Formation of these developmental hubs is partly dictated by

the pioneer TF Zelda. Further studies will be required to see whether multi-enhancer hubs are ubiquitous features of transcriptional control, and how they coordinate with promoters to regulate gene expression.

What are the mechanisms of clustering?

Two models have been proposed to explain the mechanism of clustering, depending on how molecules are interacting and recruited into clusters. The first model postulates formation of transcription condensates that exhibit properties of liquid–liquid phase separation (LLPS) (Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018) (**Figure 13**), as has been proposed for the formation of membraneless organelles, such as the nucleolus (Feric et al., 2016), inside cells. Phase separation is characterized by multiple low-affinity interactions between proteins and nucleic acids that drive demixing of particular combinations of factors into different "phases" (analogous to demixing of an oil/water emulsion), such that compatible factors form distinct physical droplets or condensates (Y. Shin & Brangwynne, 2017). The activation domains of TFs are particularly enriched for low complexity intrinsically disordered regions (IDRs) (Hnisz et al., 2017; S. Kim & Shendure, 2019). Recent results from *in vitro* and *in vivo* experiments, have indicated that weak and dynamically multivalent IDR-IDR interactions result in the formation of phase-separated condensates (Chong et al., 2018; Sabari et al., 2018). Pol II and the co-activator Mediator have been found to colocalize in clusters, which associate with chromatin and have properties of phase-separated condensates (Cho et al., 2018). Furthermore, MED1 subunit of the Mediator along with the co-factor BRD4 have been found to form nuclear bodies at super-enhancers (Sabari et al., 2018). Interestingly, disruption of MED1 and BRD4 condensates upon 1,6-hexanediol treatment, which is known to disrupt liquid-like condensates (Yi Lin et al., 2016), was accompanied by reduced MED1 and BRD4 occupancy at super-enhancers and loss of RNA Pol II occupancy at SE-associated genes, implying that loss of these condensates can impact transcription. In the context of transcription, IDR-IDR interactions between different molecules of the transcriptional machinery -including Pol II- might occur, thus facilitating the loading of the cluster on specific genomic loci.

The second model proposes that clustering is dominated by binding of multiple regulatory factors to a scaffold of cognate DNA and chromatin binding sites (S. Kim & Shendure, 2019;

J. Li et al., 2019; Jieru Li et al., 2020). Evidence for this model comes from systematic characterization of the effects of mutations that abolish specific recognition of DNA and chromatin targets or IDR-IDR interactions of Sox2 and Brd4, at specific active-gene loci (Jieru Li et al., 2020). While IDR-deletion mutants incorporated into clusters, with similar efficiencies as those of the respective wild-type proteins, Sox2 with deleted or mutated HMG DNA-binding domain and Brd4 with mutated bromodomains involved in acetyl-lysine recognition did not. These results demonstrate that Sox2 and Brd4 clustering in mESCs nuclei relies on specific recognition of multiple cognate DNA and chromatin binding sites, rather than IDR-IDR interactions, indicating spatial clustering of distal *cis*-regulatory elements and target genes.

Together, these results point out that different mechanisms may regulate enhancer enhancer-promoter communications. Whether phase separation is essential for 3D genome organization is under intense debate, however increasing evidence supports a role for these condensates (S. Kim & Shendure, 2019; Lim & Levine, 2021). Nevertheless, tools to study phase separation *in vivo* are still in their infancy, and have been primarily focused on live imaging experiments on tagged protein components; the dynamics of specific gene loci relative to their residence in nuclear microenvironments is largely unknown. Elegant optogenetic clustering of IDRs can modulate phase separation (Y. Shin et al., 2017), and this has been combined with dCas9-targeting (CasDrop) to repetitive sequences to induce droplets consistent with transcriptional activation to heterochromatin (Y. Shin et al., 2018) but such experiments have yet to be realized on single-copy gene loci, let alone determination of their effects on specific gene expression.

Is enhancer-promoter proximity even required?

Although several studies support that enhancer control of gene expression requires direct promoter contact, recent imaging experiments have shown enhancers at large distances from activated genes (Alexander et al., 2019; Benabdallah et al., 2019; Lim et al., 2018). A case in point comes from the aforementioned ZRS-*Shh* communication, whereby all *Shh* enhancers are found within the same ~960 kb TAD (Anderson & Hill, 2014; Williamson, Lettic, Hill, & Bickmore, 2016), of which one of the most distal, ZRS, resides within the intron of the unrelated Lmbr1 gene, and is specific to developing limb bud (**Figure 14**).

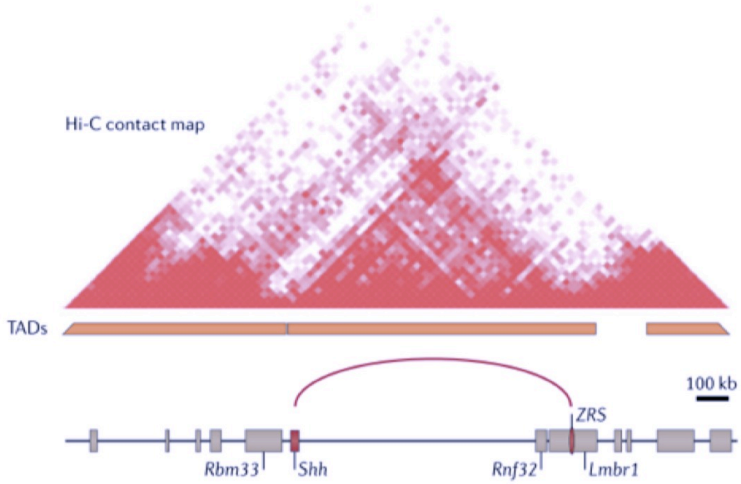


Figure 14: Genomic organization of the mouse Shh locus. In developing limb buds, the ZRS enhancer controls the expression of *Shh* located 850 kb away (Lettice et al., 2003). Adapted from (Schoenfelder & Fraser, 2019).

Using 3D-FISH and structured illumination microscopy, Williamson and colleagues (Williamson et al., 2016) reported close proximity between *Shh* and ZRS in the nucleus of all cell types examined, regardless of *Shh* activity status. Nevertheless, high levels of colocalization (<200 nm) between *Shh* and the ZRS were observed only in the zone of polarizing activity (ZPA) of the distal posterior limb bud where *Shh* is expressed, consistent with a specific enhancer-promoter loop contact. This finding contradicts a previously published work that identified similar rates of *Shh*-ZRS colocalization between the *Shh*-expressing ZPA and anterior limb buds, in which Shh is not expressed under normal conditions (Amano et al., 2009).

The same group used the CRISPR-Cas9 technology to delete major CTCF sites of the Shh TAD in mouse ESCs and *in vivo* (Williamson et al., 2019). Loss of individual CTCF sites (ΔCTCF1, 2, 3) at the Shh TAD boundaries disrupted chromatin structure and the spatial proximity between *Shh* and ZRS in both ESCs and E11.5 limb bud tissue. Strikingly, CTCF-site deletion had no effect on Shh expression pattern or development and mice carrying homozygous CTCF site deletions were viable, without any deleterious phenotypes, arguing that the previously identified loop between *Shh* and ZRS in the posterior limb bud is not functionally important. These results are compatible with evidence showing increased Shh

neuronal enhancer-promoter distance upon transcription activation (Benabdallah et al., 2019). In the developing mouse brain, *Shh* expression is driven by a set of enhancers, the Shh-Brain-Enhancers (SBEs), located from 100 up to 780 kb upstream of Shh (Benabdallah et al., 2016; Jeong et al., 2006; Yao et al., 2016). Using super-resolution 3D-FISH and 5C assays, Benabdallah *et al*. observed greater rather than decreased separation between *Shh* and its enhancers SBE6, SBE4, and SBE2/3 during differentiation of ESCs to neural progenitors and *in vivo*. Activation from a distance, using TALE-VP16 targeting either the *Shh* promoter or the distal enhancers, recapitulated the decreased enhancer-promoter proximity identified during ESC differentiation.

Consistently, interrogation of the spatial organization and the activity of *Sox2* and its distal enhancer (SCR) in living mouse ESCs, showed no association between *Sox2*-SCR spatial proximity and *Sox2* transcription (Alexander et al., 2019). *Sox2*-SCR separation distances were ranging between 200–400 nm (mean ~340 nm) in the nucleus of mouse ESCs, while the same label pairs displayed greater proximity in differentiated cells, where the SCR is inactivated, compared to ESCs. Incorporation of an MS2 reporter (Bertrand et al., 1998) into *Sox2,* for direct visualization of nascent transcription in real-time, showed no correlation between *Sox2*-SCR distances and transcriptional activity, neither prior to nor during transcriptional bursts. This locus is the focus of my thesis work, and direct comparison and discussion of my results relative to this study (which was published during my thesis work) is detailed in later sections (see Results and Discussion).

Together, these results provide strong evidence that direct molecular contact is not essential for enhancer activation of its target gene. All these observations could be explained by the "hub", "cluster" or "condensates" hypothesis. However, a follow up study on *Pou5f1* and *Sox2* loci demonstrated that distal enhancers are frequently in close distance (<200 nm for *Pou5f1*; <180 nm for *Sox2*) around the transcription site. Altogether, despite remarkable progress over the past years, much requires to be understood about the mechanisms of enhancer-promoter communication.

Chromatin architecture and topology

During my thesis I wrote a book chapter (Platania & Sexton, 2020), giving an overview of how chromatin folding is linked to cell fate decisions via transcriptional regulation, focusing on pluripotent and somatic cells, as well as particular cases (e.g. mitosis) where chromatin folding appears very different.

CHAPTER

Chromatin architecture and topology in pluripotent stem cells

4

Angeliki Platania[a,b,c,d] and Tom Sexton[a,b,c,d]

Institute of Genetics and Molecular and Cellular Biology (IGBMC), Illkirch, France[a] CNRS UMR7104, Illkirch, France[b] INSERM U1258, Illkirch, France[c] University of Strasbourg, Illkirch, France[d]

Introduction

Genomes are more than linear sequences of DNA. Two meters of mammalian DNA is compacted by an order of ~ 10^5 to be contained within each cell nucleus, implying that chromosomes must be highly folded structures. Since early descriptions of chromatin-dense heterochromatin and more open euchromatin, it has been appreciated that chromosome structure is heterogeneous yet highly organized. Crystallographic and electron microscopy studies have given detailed structural insight into the primary folding of DNA into 10-nm nucleosome fibers, although the nature of any higher-order folding was difficult to elucidate [1]. Recent advances showed that nucleosome fibers do not appear to form higher-order "30-nm filaments" in vivo; instead the different compaction states between interphase euchromatin and heterochromatin and even mitotic chromosomes are brought about by different local concentrations of 10-nm nucleosome fibers [2, 3]. The advent of the chromosome conformation capture method [4] and its higher-throughput variants [5–11] allowed higher-order chromosome folding to be inferred by identifying spatial proximity between distal genomic sequences (see Ref. [12] for overview of the technical differences between the methods). Collectively, these built up a hierarchical model of chromosome folding, whereby architectural features are present at multiple scales, each *correlating* with transcriptional control (Fig. 1), although questions remain as to whether chromatin topology is a driver or passenger of genome function [13]. Firstly, distal regulatory elements such as enhancers are brought into physical proximity with their cognate gene promoters via the formation of chromatin loops. Second, chromosomes appear to be organized into discretely folded modules, termed topologically associated domains (TADs), which may delimit autonomously regulated chromatin regions. Third, genes or TADs tend to form long-range interaction networks with regions of similar transcriptional activity or chromatin state, which has been proposed to allow coordinated control of gene expression programs. Lastly, chromosomes occupy distinct territories within the interphase nucleus, with nonrandom radial positions linked to gene density and activity. This chapter will give an overview of our understanding of all these scales of chromosome folding.

Pluripotent stem cells (PSCs) derived from the blastocyst inner cell mass are indefinitely self-renewing and can be differentiated into any of the three germ layers [14, 15]. All lineage-specific genes need to be capable of activation under appropriate developmental cues. As a consequence, PSCs have widespread low-level transcription across the whole genome [16] and have hallmarks of chromatin that

Stem Cell Epigenetics. https://doi.org/10.1016/B978-0-12-814085-7.00004-0

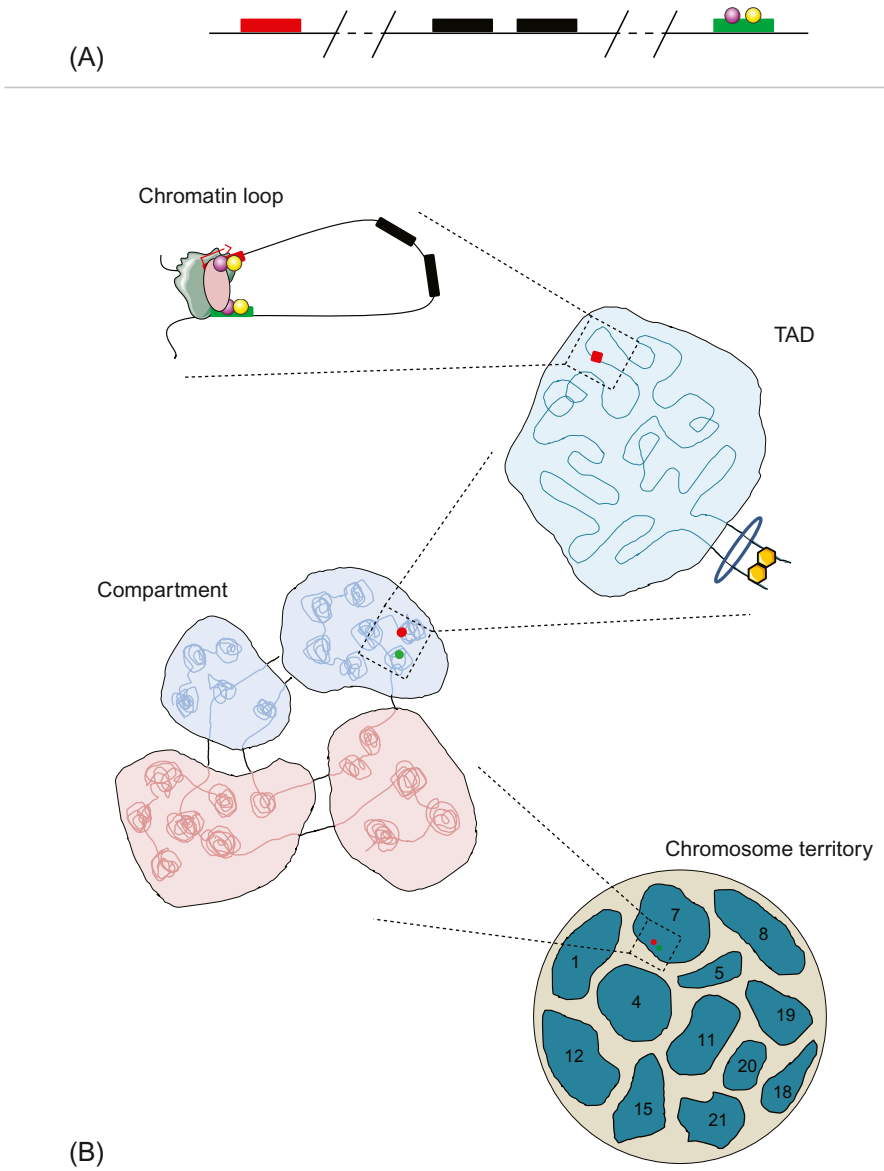


FIG. 1

Hierarchical spatial organization of the genome. (A) A linear view of genome control, whereby a regulated gene (red) is located far from a regulatory enhancer (green), bound by specific transcription factors (purple and yellow spheres), with unregulated genes (black) located in the intervening sequence. (B) A three-dimensional view of the same regulated gene. First, chromatin loops bring genes and their regulatory elements into direct physical proximity, often bridged by bound transcription factors (spheres) and the engaged RNA polymerase complex (pink oval and sea green shape). These loops are organized into discrete TAD modules, believed to be in turn defined by cohesin rings (blue) and barrier elements bound by CTCF (yellow hexagons). Specific gene coassociations (e.g., between the red and green gene indicated in the figure) occur within and between TADs within compartments of chromatin sharing the same epigenetic features, denoted by blue and red compartment colors. Compartments are in turn organized into discrete chromosome territories.

is more open than in differentiated cells: a reduced amount of heterochromatin, and looser binding of architectural proteins such as linker histones [17]. Importantly, establishment of this chromatin state appears to be an epigenetic barrier to all methods of reprogramming differentiated cells to induced PSCs (iPSCs) [18]. Repressed yet rapidly inducible lineage-specific genes are kept in a poised state, characterized by a "bivalent" chromatin mark containing both trimethylation of histone H3 lysine-27 (H3K27me3), deposited by the Polycomb group (PcG) protein repressors, and trimethylation of histone H3 lysine-4 (H3K4me3), associated with active promoters. This chromatin state is not unique to PSCs, but is more prevalent in less differentiated cells, and believed to confer developmental plasticity on the transcriptional fate of marked genes (reviewed in Ref. [19]). Despite the special chromatin configuration of PSCs, the underlying principles of their chromosome folding appear to be largely conserved. Recent studies have identified and characterized the major players in PSC chromatin organization, some constitutive factors and others unique to pluripotency, which will be discussed in more detail in this chapter.

Chromatin loops: Remote control of gene transcription

Since early transgenic studies, it has been appreciated that promoter sequences are insufficient for full activation of most developmental genes, nor to confer faithful spatiotemporal expression patterns. Most metazoan genes are also under the control of distal regulatory elements, which can act over megabase distances and reside within nonregulated genes [20]. Seminal studies of the beta-globin locus demonstrated that a cluster of erythroid enhancers, the locus control region (LCR), comes into closer physical proximity with the activated globin gene promoter than the intervening sequence [21], implying the formation of chromatin loops. Such loops linking enhancers to their cognate genes were subsequently identified in many other tissues, including PSCs [22]. The resulting "active chromatin hub" bringing together the regulatory factors bound to enhancer and promoter sequences is proposed to facilitate transcription. Extensive epigenomic profiling of PSCs and various cell and tissue types suggested that enhancers impart most control over gene expression programs and hence cell identity, since they are much more developmentally plastic than promoters [23]. Among more conventional punctate enhancers, LCR-like clusters or "super-enhancers" were found enriched near genes controlling pluripotency in PSCs [24] and lineage-specific genes in differentiated cells [25], and it has been proposed that they are somehow a distinct, more important class of regulatory element controlling cell identity. However, more precise genetic dissection of "super-enhancers" in PSCs or erythrocytes suggests that they are merely clusters of conventional enhancers, with additive rather than synergistic effects on transcription [26, 27].

Despite their genome-wide identification and importance in epigenetic regulation, it has been difficult to unambiguously assign gene targets to enhancers, as they may not regulate the closest genes on the linear sequence [20]. A significant effort has been made with different 3C variants to identify all promoter-centered looping interactions and thus identify the enhancers which physically contact, and presumably regulate, each promoter [28–30]. The most extensive promoter "interactomes" have been determined with the very recently developed Promoter Capture Hi-C (PCHi-C) method, which in essence entails high-throughput sequencing of 3C interactions (as for Hi-C) after preenrichment by sequence capture of the 3C material with complementary oligonucleotides covering all promoters

of a genome [7, 31, 32]. Collectively, all of these studies varied in resolution and coverage and were applied to different cell types but arrived at the same general conclusions. First, promoter-interacting regions are enriched in hallmarks for enhancers, such as monomethylation of histone H3 lysine-4 (H3K4me1) and acetylation of histone H3 lysine-27 (H3K27ac), and this enrichment is much more pronounced for active genes. Second, many enhancers do not interact with the genes that are closest in terms of linear sequence, demonstrating that the previously described "long-range" enhancers are not exceptional cases. Third, many promoter-enhancer interactions bypass intervening sites for the insulator protein CTCF (CCCTC-binding factor), which were believed to prevent enhancers communicating with more distal genes [33]. These findings were also the case for PSC promoter interactomes [31, 32]. Interestingly, bivalent promoters were enriched in interactions with enhancers carrying the chromatin hallmarks of "poised" enhancers (H3K4me1 and H3K27me3, instead of H3K27ac). Perturbation of PcG proteins does not cause large-scale reduction of these interactions as the bivalent state is lost; rather the interacting enhancers adopt more active chromatin states, consistent with derepression of PcG-regulated genes [34].

Despite our growing appreciation of the 3-D promoter-enhancer interaction landscape, it is still unclear *how* a regulatory element locates its cognate promoter. Furthermore, questions remain as to whether their interaction is sufficient for transcriptional activation, and whether transcription (re-) initiation, elongation and/or another process is regulated. The LCR does not form distal interactions in nonerythroid cells and specifically contacts the transcribed globin gene in the developmental stage where it is expressed [21], suggesting that such enhancer-promoter loops can be instructive for transcription. In support of this, it was found that overexpressed pluripotency transcription factors bound to enhancer and promoter sequences of the *OCT4* locus in both human iPSCs and unreprogrammed cells, but that iPSCs uniquely had enhancer-promoter interactions and OCT4 transcription [35]. Furthermore, elegant studies showed that artificial induction of chromatin loops within the beta-globin locus can stimulate transcription [36, 37]. However, enhancer-promoter loops have been identified in other model systems, which form prior to transcriptional activation, suggesting that chromatin topology may be permissive for gene expression, but is not always sufficient. Some of these cases are for acute responses of differentiated cells to extracellular signals [38, 39], where preformed enhancer contacts would be advantageous for a rapid transcriptional induction. However in other cases, such as Drosophila embryogenesis or iPSC reprogramming, enhancer-promoter interactions were detected cell cycles and/or days before transcriptional induction, suggesting that stable chromatin loops are not restricted to acute biological processes [40–42]. Very recent PCHi-C studies of PSC [43] and other progenitor [44, 45] differentiation systems identified large-scale occurrences of stable and dynamic promoter-enhancer interactions, suggesting that both instructive and permissive looping models apply to transcriptional control. It is currently unclear how these two loop types are mechanistically distinguished. The "poised" interactions reported in Drosophila were enriched at promoters loaded with paused RNA polymerase [41]. Similarly, bivalent gene promoters in PSCs have PcG-dependent paused polymerase [46], and these tend to participate in permissive rather than instructive chromatin loops [34]. Coupled with independent findings that induced chromatin loops rescued transcription initiation but not elongation in the artificial globin system tested [36] and that enhancers may contact transcribed gene bodies in addition to promoters [47], it is interesting to speculate that instructive chromatin loops can stimulate transcription initiation and elongation, whereas permissive chromatin loops require extra regulatory input for efficient elongation (Fig. 2).

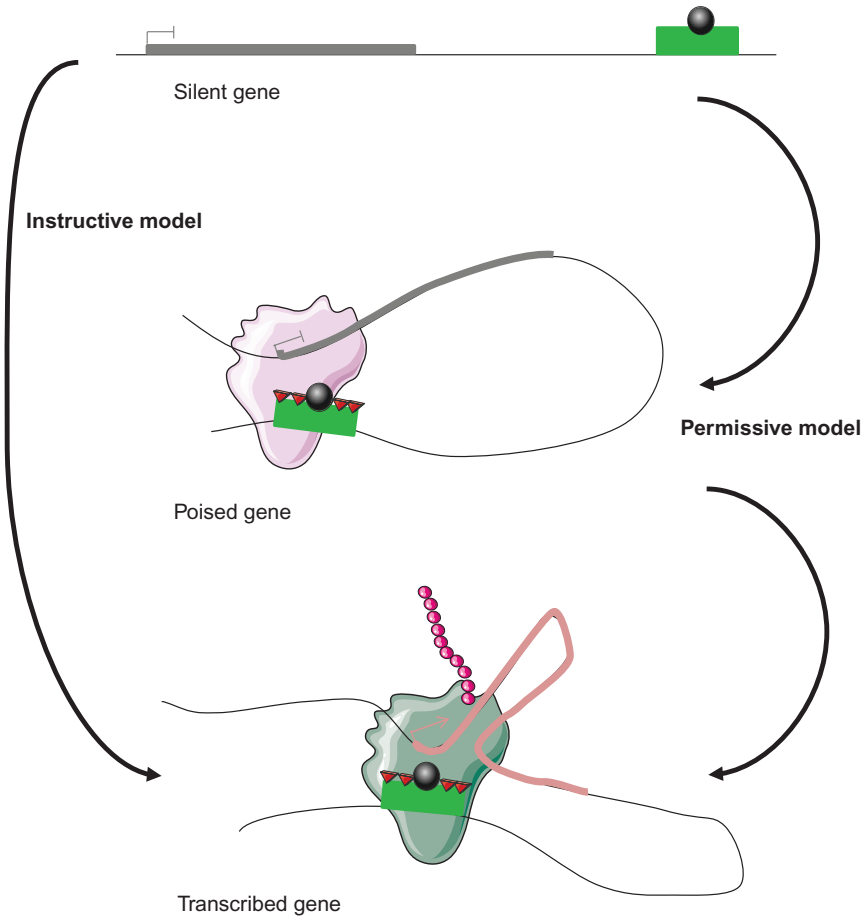


FIG. 2

Instructive and permissive models for enhancer-promoter chromatin loops. Top: The locus containing a silent gene (gray) and a distal enhancer (green) is in a linear conformation, with no enhancer-promoter contacts. Middle: The enhancer and promoter interact, presumably via binding of specific transcription factors (red triangles), and paused RNA polymerase (pink shape) is bound, but productive transcription is unable to occur. Bottom: The elongating RNA polymerase (sea green shape) now produces transcripts (magenta spheres), with the active gene (salmon pink) in a conformation allowing interactions between promoter, enhancer and the transcribed gene body. The instructive looping model allows a direct transition between the silent and transcribed states, whereas the permissive model passes through the intermediate poised state.

The factors in chromatin looping: Not just transcription

As the initial studies concentrated on enhancer interactions, the first factors to be implicated in chromatin looping were transcription factors, which cobind to both promoter and distal regulatory elements and are thus gene- and often cell type specific [48, 49]. In mouse PSCs, for example, Klf4 is required

for enhancer interactions within the *Oct4* locus and subsequent maintenance of pluripotency [42]. In most of these studies, it was difficult to decouple the effects of factor depletion on chromatin looping and transcription, although induced looping experiments in some loci [36] and observations of interaction loss preceding transcriptional perturbation in others [42] suggests that binding of transcription factor-mediated chromatin interactions is not just a by-product of transcriptional activation. Beyond gene-specific loops the general transcriptional coactivator Mediator is frequently found at transcribed promoters and enhancers and is implicated in most chromatin loops associated with gene activation [22, 50]. However, it remains unclear whether the transcription factors, Mediator, and/or some other factors are principally involved in chromatin looping, nor whether this is brought about by simple protein-protein interactions between factors at distal genomic sites, or involves a more complex process to deform the chromatin fiber or environment, such as phase-separated condensates mediated by coactivators [51].

Beyond promoter-centric interactions, the other factor implicated in chromatin loops by early studies was CTCF, the major insulator protein in mammals [33, 52]. Initially, these loops were considered to mediate insulation by keeping genes and distal regulatory elements physically separated [53], although the finding that many enhancers can bypass CTCF sites questions the universality of this model [30, 32]. Homotypic interactions between CTCF sites are present in the strongest interactions detected in genome-wide Hi-C studies [54], and although tissue-specific CTCF-mediated loops have been reported [55, 56], the majority appear to be constitutive, longer-range interactions, indicative of a more general chromosomal architectural role [30, 50]. Although CTCF tends to be depleted from enhancer-promoter interactions directly, CTCF-mediated loops interweave with and appear to reinforce them, since deletion of specific sites can actually perturb enhancer contacts [57]. Interestingly, such deletions only caused modest reductions in transcription levels, but greatly increased cell-to-cell variability, suggesting that CTCF-mediated loops confer robust expression control by context-dependent architectures and not solely by classical insulator activity. Curiously, CTCF loops seem to only occur between sites in convergent orientation [54, 58], a topological constraint which does not apply to transcription factor binding motifs at looping regions. In support, genome editing experiments to invert specific sites did not generally affect CTCF binding, but severely disrupted chromatin loops [59–61]. However, the inverted sites often did not participate in de novo interactions with CTCF sites now brought into a compatible orientation, suggesting that the motif alone is not sufficient to define chromatin topology.

A factor implicated in both of these seemingly distinct "transcriptional" and "architectural" chromatin loops is cohesin, a protein complex initially discovered to tether sister chromatids together after DNA replication [62]. Cohesin has been implicated in both enhancer-promoter interactions [22, 35, 42, 63] and CTCF-mediated loops [50, 64], although many interactions of both types have also been detected, which are not accompanied by bound cohesin. The cohesin complex tethers sister chromatids by the formation of a ring structure, which needs to be broken to release the chromatids at mitosis [62]. Although yet to be demonstrated, it is tempting to speculate that the same ring structure can stabilize chromatin loops, particularly since cohesin binding appeared to discriminate stable, poised enhancer-promoter interactions from acute loops brought about by specific transcription factors in a model of terminal differentiation [44]. Further, depletion of cohesin destabilized chromatin interactions required for maintenance of pluripotency in PSCs [35], although cell cycle effects from abrogation of sister chromatid cohesion could not be ruled out in this study. The mechanisms of postreplicative cohesin loading and unloading have been well studied (reviewed in Ref. [62]), although it is less clear how cohesin is recruited to specific genomic sites associated with chromatin loops. Cohesin proteins have

no sequence-specific binding activities, but have been shown to interact with CTCF [65], Mediator [22], and some transcription factors [42], sometimes in conjunction with the cohesin-loading factor Nipbl, implying that cohesin recruitment is secondary to binding of these sequence-specific binding factors. Future studies should elucidate what determines whether cohesin is brought to a chromatin interaction and whether or how this affects chromatin topology dynamics.

Overall, chromatin loops provide a first order of chromosome folding, modulating communication between gene promoters and distal regulatory elements. The core factors, such as Mediator, cohesin and CTCF are involved in presumably the same mechanisms in PSCs and differentiated cells; specific transcription factors may provide some of the looping specificity, including the pluripotency transcription factors in PSCs. Polycomb appears to play an additional role in permissive loops at bivalent gene promoters, but this may not be unique to stem cells.

Topologically associated domains as units of genome regulation

One of the first observations to come from Hi-C studies was the apparent organization of chromosomes into discretely folded modules, or TADs, of multikilobase to megabase size, whereby chromatin interactions are strong within a domain, but sharply reduced for interactions spanning a border between two TADs [66–68]. TAD demarcation correlates extremely well with the organization of "linear" epigenetic marks, such as histone modifications, transcriptional activity, lamin association and DNA replication timing [66–70], so has been proposed to reflect both a spatial and functional organization of chromosomes into autonomously regulated gene neighborhoods. TADs could facilitate regulation of gene programs in two, nonmutually exclusive ways. First, they can spatially delimit the effective operational range of distal regulatory elements such as enhancers, thus preventing their aberrant activation of neighboring genes. In support of this model, most promoter interactions detected by PCHi-C are confined to the same TAD [32], and high-throughput enhancer trap studies showed that TADs also operationally delimit reporter gene expression patterns [71]. Most strikingly, genetic diseases have now been attributed to aberrant enhancer-promoter contacts caused by deletions of TAD borders, inversions or translocations of TADs, or duplication events creating new hybrid TADs [72–74], although the phenotypic consequences seem to be context dependent [75] (Fig. 3). Conversely, efficient enhancer activity on target genes may be assured by limiting their three-dimensional search ranges to within TADs. In support of this, systematic genetic manipulation of the TAD containing the *Shh* locus showed that the limb-specific enhancer was relatively insensitive to distance from the gene, provided they were in the same domain [76]. Such a confined interaction search space has also been proposed to explain why TADs can contain coordinately regulated genes [13, 67, 69].

Comparisons between cell types, including between naïve and primed PSCs, and between PSCs and different derived lineages, revealed that the majority of TADs are invariant [66, 77, 78], even at syntenic chromosomal regions across different species [58, 79]. Super-resolution DNA fluorescent in situ hybridization (FISH) support this structural conservation further [80], suggesting that TADs are somehow genetically encoded and relatively insensitive to tissue-specific epigenomic profiles. At first, this view of invariant TADs appears at odds with their correlation with epigenetic marks [66, 68], which are in turn highly developmentally dynamic [23]. More focused studies of selected loci during PSC differentiation observed that "sub-TADs" and enhancer-promoter interactions were rewired within stable larger domains [50], leading to the suggestion that TADs may indeed be stable, and that cell reprogramming

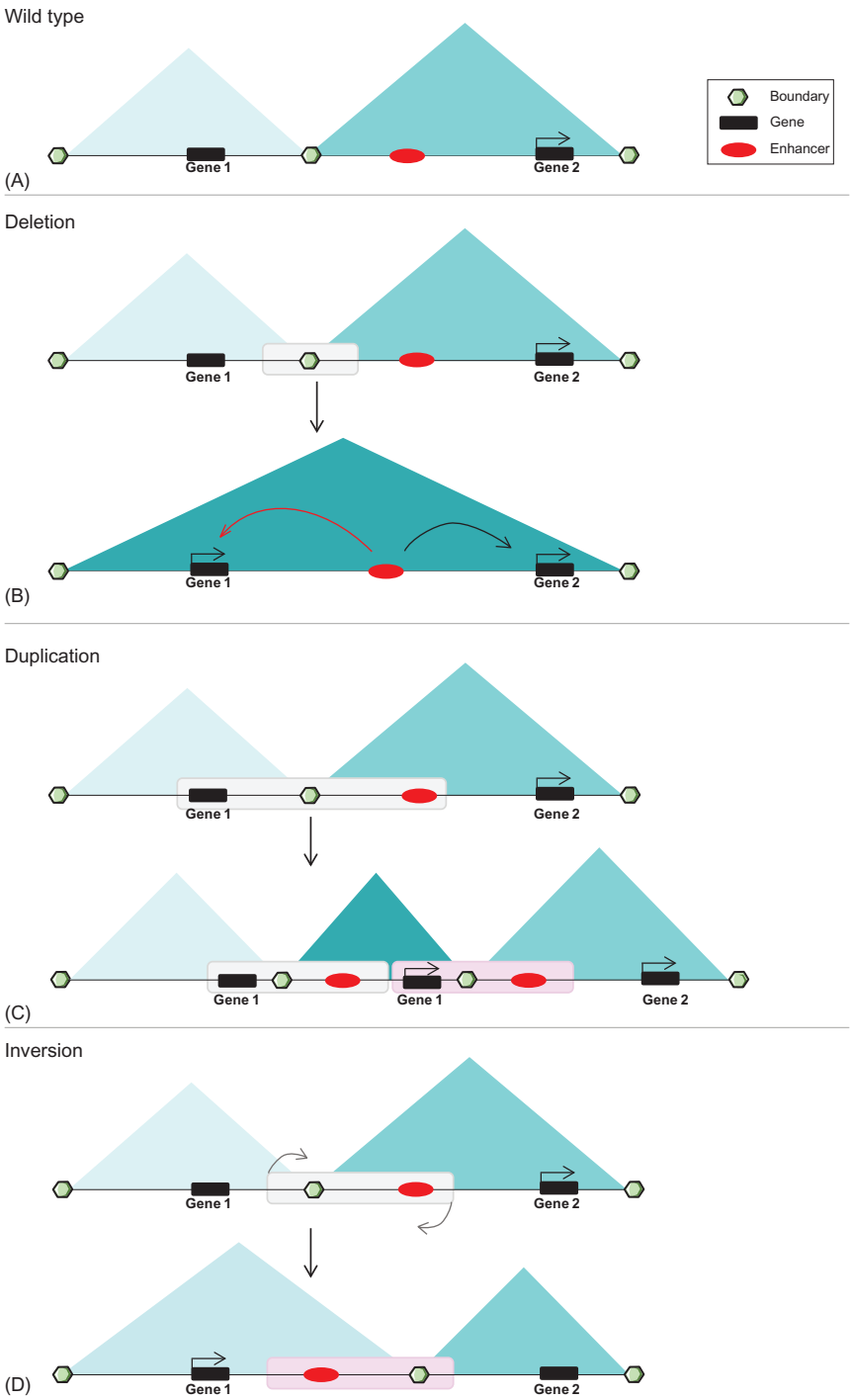


FIG. 3

See figure legend on opposite page.

is accompanied by finer-scale chromatin topology dynamics. Nevertheless, complete TAD remodeling has been observed during the collinear activation of Hox genes, whereby a single TAD encompassing the inactive gene cluster splits into two with a developmentally shifting border between the active and repressed Hox genes [81, 82]. More detailed higher-resolution studies are required to explore to what extent the distinction between a TAD and a "sub-TAD" is semantic, and how much chromatin domains can or need to be restructured to accommodate epigenetic and transcriptional changes during differentiation [83].

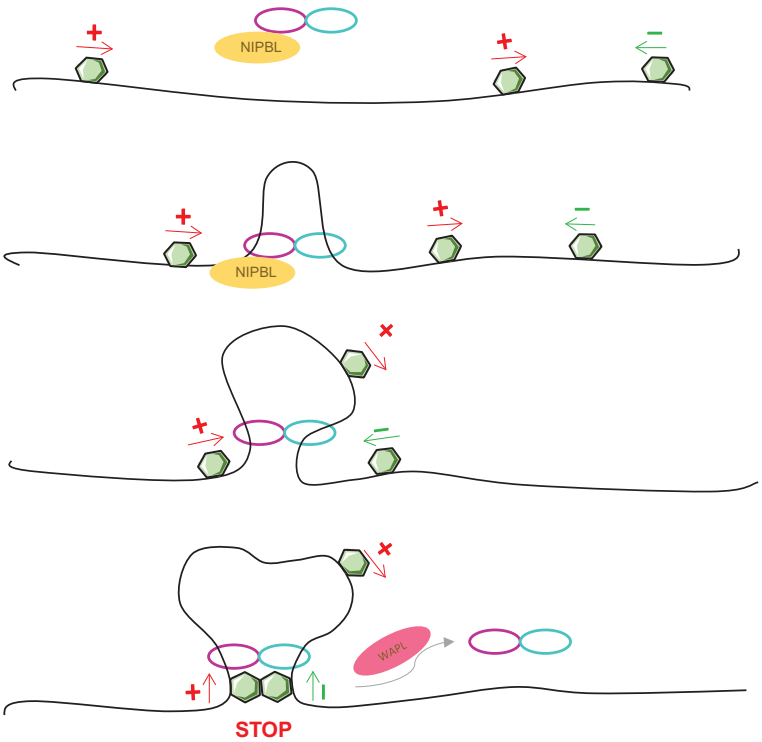
Despite recent intensive studies, it remains unclear exactly how TADs are formed or stabilized. Borders are enriched in active genes and hallmarks of transcription, such as RNA polymerase and H3K4me3, as well as the "architectural" proteins cohesin and CTCF [66, 68]. As these factors are the major players in chromatin loops, TADs could be a consequence of very strong interactions between TAD borders [54]. However, TAD borders represent only a small subset of genome-wide CTCF sites and active genes, so their presence alone is insufficient to define chromatin domains. Large genetic deletions at borders can cause TADs to fuse [74], or for preexisting intra-TAD interactions to reinforce as new, "secondary" borders [67, 84]. Precise deletions of single CTCF sites only caused mild effects on the overall TAD structure, but could nevertheless have important functional consequences from increased inappropriate enhancer interactions [85–87]. A recent study with tunable ablation of CTCF in PSCs revealed that the residual amounts of protein left after a traditional knockdown was sufficient to maintain TAD structure [88]. Complete ablation, on the other hand, severely disrupted ~80% of TADs, with widespread but mild misregulation of transcription. Even more dramatically, systematic TAD loss has been reported when cohesin is completely ablated, either by degradation or block of loading onto chromosomes [89–91]. Cohesin thus appears essential for generation or stability of loops and TADs, whereas CTCF plays a large but not complete role in defining the positions of TADs.

Building chromosomal domains: Looping principles extended a TAD

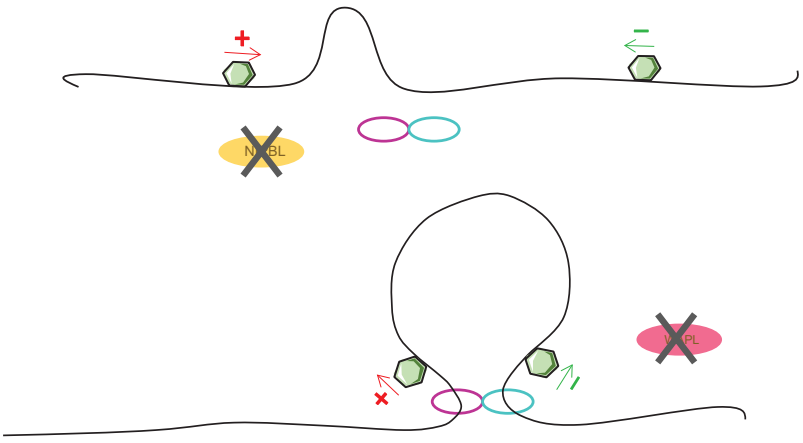
Since Hi-C pairwise interaction maps have been generated, physical models have attempted to explain some or all of the complex architectural features discovered [8, 61, 92, 93]. The loop extrusion model, shown in Fig. 4, provides an explanation for TAD structures containing convergent CTCF elements at the borders and relatively uniform interactions within the domains [61, 94, 95]. An extrusion factor or

FIG. 3

Genetic diseases can be mediated by pathological alterations to TAD structures. (A) Schematic TAD architecture around a typical wild-type locus, whereby two TADs (triangles of different blue shades) are demarcated by a boundary (green hexagon). The left TAD contains an inactive gene (black rectangle), and the right TAD contains an active gene under the control of an enhancer (red oval). (B) Deletion of the TAD boundary causes the unification of the two TADs, with aberrant activation of gene 1, which is now contained within the same domain as the enhancer. (C) Duplication of a genomic locus containing a gene, a TAD boundary, and an enhancer can create a new TAD structure. Although the initial copy of gene 1 is still repressed, the duplicated gene may be aberrantly activated if it shares the new TAD with the duplicated enhancer. (D) Inversion or translocation events may move enhancers to different TADs, potentially causing aberrant expression of gene 1 *and* reduced expression of gene 2.



(A)



(B)

FIG. 4

See figure legend on opposite page.

factors is proposed to bind at seemingly random sites within the genome, then physically extrudes a chromatin loop bidirectionally, with two components of the extruding factor translocating in opposite directions. An equilibrium between extrusion and factor dissociation from chromatin is set up as the extruded loop grows, particularly influenced by physical barriers to extrusion set up at TAD borders. Asymmetric barriers, such as CTCF sites, may be expected to be in a convergent orientation to appropriately meet the oncoming translocating chromatin fiber. Since cohesin forms ring structures to tether and organize sister chromatids [96], it is a major candidate for a loop extrusion factor. Further, TAD maintenance appears to be an active process, since TAD maintenance requires ATP and the cohesin complex ATPase domain [97]. Chromatin immunoprecipitation studies finding frequent cooccupancy of cohesin and CTCF, particularly at TAD borders, can therefore be interpreted as more stable, stalled loop extrusion complexes. This model predicts that altering the extrusion/dissociation equilibrium, for instance, by influencing extrusion factor residence time, will alter TAD size. In support of this and of cohesin being an important extrusion factor, deletion of cohesin release factor, WAPL, or loading factors such as Nipbl, increase or shrink average chromatin loop length, respectively, with a weakening of TAD structures [91, 98]. Notably, deletion of Nipbl in a different model system, mouse liver, had a more extreme phenotype, completely destroying TADs [90]. Depending on the cell type, extruding factors in addition to cohesin may operate, and/or extra factors may inefficiently load/disassemble cohesin from interphase chromosomes. The bulky RNA polymerase complex and associated coactivators could potentially form a CTCF-independent barrier to loop extrusion, explaining the prevalence of active genes at TAD borders [66, 68]. Cohesin-mediated enhancer-promoter interactions may thus be considered as metastable loop extrusion intermediates. However, many chromatin interactions are not associated with bound cohesin, again implying that chromatin loops may be stabilized by multiple protein-protein interactions, and/or that other factors can mediate loop extrusion.

Overall, seemingly autonomous regulatory domains are set up in PSCs and other somatic cells by largely evolutionarily and developmentally stable TADs. Their basic architectural principles, outlined in this chapter, may allow genes to be more precisely homed to their cognate regulatory elements, although major questions remain poorly addressed. For example, what is the interplay between bulk TAD organization and dynamic chromatin ultrastructures observed at the large transcriptional changes accompanying development [83]? What role do specific intra-TAD interactions play in TAD stability? We have much to learn about the "fine print" of TAD "architectural blueprints" and how this organization fits within chromatin regulation and other scales of chromosome folding.

FIG. 4

The loop extrusion model for TAD formation. (A) NIPBL (yellow oval) loads cohesin rings (purple and cyan circles) on to chromatin, which then extrudes a growing loop by translocating in opposite directions. The extrusion process is stalled when the base of the loop contains two barrier elements (green hexagons), usually CTCF sites, in convergent orientation, creating metastable loops in turn building TAD structures. Cohesin rings are also unloaded from chromatin by WAPL (pink oval), with extruded loop size determined by the equilibrium of cohesin loading and disassembly. (B) The loop extrusion model explains phenotypes of NIPBL or WAPL perturbation. Defects in cohesin loading on disruption of NIPBL drastically reduces extruded loops and hence TAD structures. Impaired cohesin disassembly by WAPL disruption shifts the equilibrium towards larger extruded loops, which may bypass CTCF-mediated borders.

Chromosome compartments: From a bipolar genome to specific functional networks

Before TADs were resolved with improved sequencing depth, the first Hi-C maps identified a curious property of chromatin interactions at the megabase scale: the genome appeared to be organized into large domains, or compartments, of two different types. "A" compartments, enriched in hallmarks of transcriptionally active genes, preferentially interacted with other "A" compartments, whereas inactive "B" chromatin compartments interacted with other "B" compartments [8]. Intermixing of the two compartments was significantly depleted, suggesting a spatial segregation of active and inactive chromatin within the nucleus. Extensive repetitive sequences are normally filtered out of Hi-C analyses, so B compartments are unlikely to be simply the classical heterochromatin observed in microscopy studies. It should also be noted that conventional Hi-C studies give the average interaction landscape of millions of fixed nuclei, so the presence of two compartment types should not be interpreted as there being two segregated "poles" of chromatin within the nucleus. Rather, FISH and single-cell Hi-C experiments suggest that a particular gene will tend to coassociate with other chromosomal regions bearing similar epigenetic marks, even though the exact interaction partner(s) can vary widely from cell to cell [99, 100]. Closer analysis of Hi-C interaction patterns revealed subcompartments of A and B types [54], which may even be extended to a continuous spectrum [99], characterized by different epigenetic signatures. The major principle of genome compartmentalization thus appears that, at a large scale, genes with similar activities tend to coassociate in the nuclear space. In agreement, whereas PSC differentiation is not accompanied by large-scale TAD structural changes, many genes switch compartments to interact with other regions matching their upregulated or repressed state [77].

Before elucidation of genomic compartments, specific networks of long-range interactions between coregulated genes had been identified in the context of developmental gene activation [10] or repression [101]. In these examples, genes spatially coassociate at nuclear foci of RNA polymerase [10, 102, 103] or PcG proteins [101, 104], presumably allowing shared access to regulatory factors and coordinate control of genes within the network. In support of this, specialized spatial networks have been identified between genes regulated by common transcription factors [10, 103]. Focused studies in (i)PSCs before and after differentiation or reprogramming suggest that PcG proteins and pluripotency transcription factors form major spatial networks in the pluripotent state [28, 34, 105]. Interestingly, perturbations of one gene can affect expression of others within their specific spatial network [106]. Although the universality of such "transcription factories" within metazoan nuclei is under debate [107], spatial gene networks would both explain and provide a functional framework for genome compartmentalization by a self-organization model [108, 109]. As the majority of DNA-bound factors have short residence times on their cognate sites [110], chance encounters between two loci create an increased local concentration of any factors that are bound at both genes. When a factor dissociates, it is thus more likely to be retrapped by a spatial cluster of nearby binding sites than to diffuse to a different location, thus reinforcing a growing, self-organizing network. The genome compartments deduced from global Hi-C maps could conceivably be an amalgamation of multiple such self-organized networks, although they have yet to be reliably deconvolved.

Since Drosophila inter-TAD interaction patterns mimic compartment organization [68] and mammalian TADs form subdomains [50], it was initially proposed that compartments just represent larger-scale TADs, subject to the same organizational principles. However, TAD and compartment architectures can be decoupled in mammals: CTCF ablation disrupts TADs with minimal changes to compartments [88],

and removal of cohesin severely affects loops and TADs but actually *reinforces* genomic compartments [89–91]. Thus competing architectural principles may actually influence chromosome folding, a concept taken further in a comparative Hi-C study, which reclassified TAD borders in active genes as small A compartments perturbing the organization of flanking B compartments [111]. A possible explanation for apparent TAD-compartment competition is that compartment self-organization allows *general* reinforcement of entire gene program regulation but creates too large a search space for efficient targeting of enhancers to *specific* genes. TAD organization may thus restrict the enhancer search space within smaller domains, perhaps ensuring transcriptional fidelity, potentially explaining why TAD disruption has widespread positive *and* negative, albeit *minor*, effects on gene expression. In summary, although the mechanistic details are likely to differ, homotypic interactions appear to organize chromosomes at different levels, from individual chromatin loops to whole megabase-sized compartments.

Chromosome territories and interchromosomal interactions

DNA FISH experiments show that interphase chromosomes predominantly occupy discrete regions of the nucleus, termed chromosome territories, but that chromosome intermingling is frequent [112]. Such a configuration presumably prevents tangling between chromosome strands and breakage during mitosis; conversely, physical models suggest that it can be the direct result of very slow equilibration after cell division [113]. Interestingly, all the different physical models explaining the behavior of Hi-C data provide a means for single chromosomes to be folded and unfolded without tangling as well [8, 61, 95]. Gene activity is correlated with nuclear location away from the bulk of chromosome territories, for example, when Hox genes are activated during PSC differentiation [114], where they are proposed to be decompacted and accessible to transcriptional machinery. In line with these findings, Hi-C experiments report much stronger intrachromosomal interactions, with a high enrichment for any interchromosomal interactions to be within the active A compartment [8, 68]. Specific cases of interchromosomal promoter-enhancer interactions have been reported [115, 116] but remain rare and controversial [117]. More likely a large repertoire of transient or infrequent interchromosomal interactions arises from the same self-organization principles proposed to build intrachromosomal compartments. An interesting corollary of this phenomenon is that the most recurrent oncogenic translocation partners correlate with preexisting interchromosomal interactions, which may be linked to their participation in spatial gene networks [118–120], although more study is needed to address whether this link is causal or correlative.

Breaking TADs: The exceptions to the rule

As the previous parts of this chapter have described, the genome is subject to the same organizational "rules" in PSCs and differentiated cells. TAD structures appear relatively stable throughout [77], and spatial gene networks linked to pluripotency transcription factors in PSCs [28, 34, 105] are replaced by different ones employing cell type-specific factors but similar principles (e.g., Klf1 [10]). However, three important cases have been discovered where chromosome structure changes drastically, each of which have a relevance in maintaining or leaving a pluripotent state.

The most dramatic genomic structural changes occur when chromosomes condense, align, and segregate during mitosis, accompanied by global transcriptional shutdown and dissociation of the majority

of regulatory factors from chromatin [121]. In concordance, bulk Hi-C studies in sorted mitotic cells revealed a complete loss of TADs and compartments, with the chromosomes forming linear rods composed of consecutive looped domains accompanied by a drastic reduction in long-range interactions [122]. Such a configuration facilitates metaphase alignment of the chromosomes but raises questions of how epigenetic information, such as that required to maintain pluripotency, is transmitted to daughter cells. Mitotic retention of "bookmarking" factors, such as Esrrb or Sox2 in PSCs [123, 124], and persistent histone modifications on chromatin likely play a role, but since even "stable" enhancer-promoter loops seem to be destroyed, it is unclear how interphase chromatin architecture is reestablished at each cell cycle. TAD organization is restored very early in G1, coincident with the moment when the replication timing of chromatin domains is established [99, 122, 125]. Very early G1 cells also experience a brief pulse of transcriptional hyperactivity [126], and it is interesting to speculate that engaged RNA polymerases facilitate chromatin architectural remodeling to the interphase state via the formation of active chromatin hubs and cotranscriptional recruitment of nucleosome remodelers and histone modifications. Recently a large catalog of single-cell Hi-C profiles allowed chromosome topology to be finely mapped according to cell cycle progression [99]. An intriguing finding was that whereas genome compartments steadily strengthened across the cell cycle, TADs were noticeably weakened at the onset of DNA replication and stayed so throughout G2. This further supports the idea that TADs and compartments are organized by fundamentally different mechanisms, and elucidating precisely what they are and how they are altered in the cell cycle will be extremely interesting. One possibility is that a limiting amount of cohesin complexes are diverted from G1 chromosome loop extrusion to G2 chromatid cohesion, but this has yet to be shown.

Before PSCs arise in the inner cell mass, major genome structural remodeling has already taken place. Sperm chromatin is uniquely compacted, with protamines replacing the majority of histones, and fundamentally different to that of the oocyte. On fertilization, these two genomes need to become more equivalent for coordinated genetic control of the zygote and subsequent developing embryo. Recent few- or single-cell Hi-C datasets have studied chromatin topology on the transitions from oocyte to zygote and early stage embryos, with some surprising results [127–129]. TADs appear largely absent in oocytes, with weak [128] to no [127, 129] presence in early zygotes, and are built up slowly on progression to the eight-cell stage [127, 129]. Curiously an appreciable depletion of interactions between the two parental genomes is observed until the eight-cell stage too, suggesting that the genomes remain spatially segregated, and that establishing their "equivalency" is somehow linked to TAD assembly. Curiously, compartments are specifically absent in the maternal genome within early zygotes [128], but it is unclear how or why. The onset of zygotic transcription would be a likely candidate to help build up TADs, in the same manner that has been proposed just after mitosis. However, transcriptional inhibition only had minor effects on TAD maturation in both mammalian [127, 129] and Drosophila [130] early embryos. Interestingly, DNA replication inhibition did seem to have an effect [129], seemingly in contrast to the finding that S phase weakens TADs in PSCs [99]. How these unique architectures are established and remodeled remain unclear; reprogramming the epigenetic landscape in early development has been an intensive area of study for many years, and we are beginning to have the tools to include chromosome topology in future research in this exciting field.

The third major chromosomal reorganization to have been uncovered is X chromosome inactivation. Both X chromosomes within female PSCs start off equivalent, but later in development, one copy must become predominantly heterochromatic and transcriptionally silent to balance gene dosage between the sex chromosomes and the autosomes. The epigenetic changes associated with choice of which

chromosome to silence and the onset and the maintenance of X inactivation are intensively studied areas [131]. Comparing chromatin topologies between the active (Xa) and the inactive (Xi) X chromosome revealed that TADs and compartments are largely absent from the Xi, although they are maintained around genes that escape X inactivation [132, 133]. Although the controversy about this topic has been described throughout the chapter, this finding suggests that ongoing transcription may have some role in TAD generation or stability. Despite the paucity of TADs, the Xi is not devoid of higher-order structure: it segregates into two "megadomains," the function of which is not entirely clear. These megadomains are dependent on the macrosatellite DXZ4 repeat, which forms the border between them, and the non-coding RNA Xist, which is essential for X inactivation onset and maintenance [132, 133]. Future research is likely to elucidate if and how megadomain structure facilitates chromosome-wide transcriptional shutdown and if they can be induced experimentally in other conditions.

Concluding remarks

Pluripotency requires the robust transcriptional silencing of all lineage-specific genes but in a manner allowing their efficient activation in response to differentiation signals. This challenge is met by a specialized epigenomic profile and chromatin architecture, maintaining developmental genes in a poised bivalent state. Nevertheless the fundamental principles of chromosome folding apply equally to PSCs as to differentiated cells. Gene promoters are regulated by looping interactions with distal regulatory elements, some instructive, such as the enhancer-promoter contacts at pluripotency gene loci [22], and some permissive, in particular the PcG-regulated interactions between bivalent promoters and poised promoters [34]. These loops appear to be regulated by constitutive factors, such as CTCF, cohesin, and Mediator, with specificity also conferred by pluripotency transcription factors. As for nearly all mammalian cells in interphase, PSC genomes are spatially organized into TADs, which may modulate the specificity and robustness of enhancer activity. Loop extrusion by cohesin complexes is the most likely and conserved mechanism for TAD generation and turnover [61, 95]. TADs and genes are further organized into genome compartments, which may contain spatial networks of coordinately regulated genes, such as that centered around the pluripotency genes [28]. A growing list of single-cell studies supports a self-organization model of gene compartmentalization, whereby the coassociation of similarly regulated genes is reinforced, but one particular gene pair need only coassociate in a small subset of the cell population. Future research to elucidate the mechanisms regulating chromatin topology and dynamics promise to give groundbreaking insight into how gene programs are regulated and will be extremely fruitful in understanding the pluripotent epigenetic state.

References

[1] Woodcock CL. Chromatin architecture. Curr Opin Struct Biol 2006;16(2):213–20.
[2] Fussner E, et al. Open and closed domains in the mouse genome are configured as 10-nm chromatin fibres. EMBO Rep 2012;13(11):992–6.
[3] Ou HD, et al. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. Science 2017;357(6349), eaag0025.
[4] Dekker J, et al. Capturing chromosome conformation. Science 2002;295(5558):1306–11.

[5] Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 2006;16(10):1299–309.
[6] Fullwood MJ, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature 2009;462(7269):58–64.
[7] Hughes JR, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. Nat Genet 2014;46(2):205–12.
[8] Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326(5950):289–93.
[9] Mifsud B, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 2015;47(6):598–606.
[10] Schoenfelder S, et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 2010;42(1):53–61.
[11] Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 2006;38(11):1348–54.
[12] Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 2016;30(12):1357–82.
[13] Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. Cell 2015;160(6):1049–59.
[14] Evans MJ, Kaufman MH. Establishment in culture of pluripotent cells from mouse embryos. Nature 1981;292(5819):154–6.
[15] Thomson JA, et al. Embryonic stem cell lines derived from human blastocysts. Science 1998;282(5391):1145–7.
[16] Efroni S, et al. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2008;2(5):437–47.
[17] Gaspar-Maia A, et al. Open chromatin in pluripotency and reprogramming. Nat Rev Mol Cell Biol 2011;12(1):36–47.
[18] Papp B, Plath K. Epigenetics of reprogramming to induced pluripotency. Cell 2013;152(6):1324–43.
[19] Voigt P, et al. A double take on bivalent promoters. Genes Dev 2013;27(12):1318–38.
[20] Amano T, et al. Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. Dev Cell 2009;16(1):47–57.
[21] Palstra RJ, et al. The beta-globin nuclear compartment in development and erythroid differentiation. Nat Genet 2003;35(2):190–4.
[22] Kagey MH, et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 2010;467(7314):430–5.
[23] Roadmap Epigenomics C, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518(7539):317–30.
[24] Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 2013;153(2):307–19.
[25] Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell 2013;155(4):934–47.
[26] Hay D, et al. Genetic dissection of the alpha-globin super-enhancer in vivo. Nat Genet 2016;48(8):895–903.
[27] Moorthy SD, Davidson S, Shchuka VM, Singh G, Malek-Gilani N, Langroudi L, et al. Enhancers and super-enhancers have an equivalent regulatory role in embryonic stem cells through regulation of single or multiple genes. Genome Res 2017;27(2):246–58.
[28] de Wit E, et al. The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature 2013;501(7466):227–31.
[29] Li G, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 2012;148(1–2):84–98.
[30] Sanyal A, et al. The long-range interaction landscape of gene promoters. Nature 2012;489(7414):109–13.

[31] Sahlen P, et al. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. Genome Biol 2015;16:156.
[32] Schoenfelder S, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res 2015;25(4):582–97.
[33] Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell 2009;137(7):1194–211.
[34] Schoenfelder S, et al. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat Genet 2015;47(10):1179–86.
[35] Zhang H, et al. Intrachromosomal looping is required for activation of endogenous pluripotency genes during reprogramming. Cell Stem Cell 2013;13(1):30–5.
[36] Deng W, et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell 2012;149(6):1233–44.
[37] Deng W, et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell 2014;158(4):849–60.
[38] Hakim O, et al. Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. Genome Res 2011;21(5):697–706.
[39] Jin F, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 2013;503(7475):290–4.
[40] Apostolou E, et al. Genome-wide chromatin interactions at the Nanog locus in pluripotency, differentiation, and reprogramming. Cell Stem Cell 2013;12(6):699–712.
[41] Ghavi-Helm Y, et al. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 2014;512:96–100.
[42] Wei Z, et al. Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. Cell Stem Cell 2013;13(1):36–47.
[43] Freire-Pritchett P, et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. Elife 2017;6, e21926.
[44] Rubin AJ, et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat Genet 2017;49:1522–8.
[45] Siersbaek R, et al. Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. Mol Cell 2017;66(3):420–435.e5.
[46] Stock JK, et al. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol 2007;9(12):1428–35.
[47] Lee K, et al. Dynamic enhancer-gene body contacts during transcription elongation. Genes Dev 2015;29(19):1992–7.
[48] Drissen R, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. Genes Dev 2004;18(20):2485–90.
[49] Jing H, et al. Exchange of GATA factors mediates transitions in looped chromatin organization at a developmentally regulated gene locus. Mol Cell 2008;29(2):232–42.
[50] Phillips-Cremins JE, et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell 2013;153(6):1281–95.
[51] Sabari BR, et al. Coactivator condensation at super-enhancers links phase separation and gene control. Science 2018;361, eaar3958.
[52] Splinter E, et al. CTCF mediates long-range chromatin looping and local histone modifications in the beta-globin locus. Genes Dev 2006;20(17):2349–54.
[53] Kurukuti S, et al. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc Natl Acad Sci U S A 2006;103(28):10684–9.
[54] Rao SS, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 2014;159(7):1665–80.

[55] Hanssen LLP, et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat Cell Biol 2017;19(8):952–61.
[56] Hou C, et al. Cell type specificity of chromatin organization mediated by CTCF and cohesin. Proc Natl Acad Sci U S A 2010;107(8):3651–6.
[57] Ren G, et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol Cell 2017;67(6). 1049–1058.e6.
[58] Vietri Rudan M, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep 2015;10(8):1297–309.
[59] de Wit E, et al. CTCF binding polarity determines chromatin looping. Mol Cell 2015;60(4):676–84.
[60] Guo Y, et al. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell 2015;162(4):900–10.
[61] Sanborn AL, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A 2015;112(47):E6456–65.
[62] Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. Annu Rev Genet 2009;43:525–58.
[63] Hadjur S, et al. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature 2009;460(7253):410–3.
[64] DeMare LE, et al. The genomic landscape of cohesin-associated chromatin interactions. Genome Res 2013;23(8):1224–34.
[65] Rubio ED, et al. CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A 2008;105(24):8309–14.
[66] Dixon JR, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012;485(7398):376–80.
[67] Nora EP, et al. Spatial partitioning of the regulatory landscape of the X-inactivation Centre. Nature 2012;485(7398):381–5.
[68] Sexton T, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 2012;148(3):458–72.
[69] Le Dily F, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. Genes Dev 2014;28(19):2151–62.
[70] Pope BD, et al. Topologically associated domains are stable units of replication-timing regulation. Nature 2014;515(7527):402–5.
[71] Symmons O, et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res 2014;24(3):390–400.
[72] Franke M, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature 2016;538(7624):265–9.
[73] Groschel S, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell 2014;157(2):369–81.
[74] Lupianez DG, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell 2015;161(5):1012–25.
[75] Despang A, et al. Functional dissection of the Sox9-Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet 2019;51(8):1263–71.
[76] Symmons O, et al. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. Dev Cell 2016;39(5):529–43.
[77] Dixon JR, et al. Chromatin architecture reorganization during stem cell differentiation. Nature 2015;518(7539):331–6.
[78] Ji X, et al. 3D chromosome regulatory landscape of human pluripotent cells. Cell Stem Cell 2016;18(2):262–75.
[79] Harmston N, et al. Topologically associating domains are ancient features that coincide with metazoan clusters of extreme noncoding conservation. Nat Commun 2017;8(1):441.

[80] Fabre PJ, Benke A, Manley S, Duboule D. Visualizing the HoxD gene cluster at the nanoscale level. Cold Spring Harb Symp Quant Biol 2015;80:9–16.
[81] Noordermeer D, et al. Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife 2014;3, e02557.
[82] Noordermeer D, et al. The dynamic architecture of Hox gene clusters. Science 2011;334(6053):222–5.
[83] Bonev B, et al. Multiscale 3D genome rewiring during mouse neural development. Cell 2017;171(3). 557–72.e24.
[84] Giorgetti L, et al. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. Cell 2014;157(4):950–63.
[85] Hnisz D, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science 2016;351(6280):1454–8.
[86] Narendra V, et al. CTCF-mediated topological boundaries during development foster appropriate gene regulation. Genes Dev 2016;30(24):2657–62.
[87] Narendra V, et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science 2015;347(6225):1017–21.
[88] Nora EP, et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell 2017;169(5). 930–44.e22.
[89] Rao SSP, et al. Cohesin loss eliminates all loop domains. Cell 2017;171:305–20.
[90] Schwarzer W, et al. Two independent modes of chromatin organization revealed by cohesin removal. Nature 2017;551:51–6.
[91] Wutz G, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J 2017;36(24):3573–99.
[92] Barbieri M, et al. Complexity of chromatin folding is captured by the strings and binders switch model. Proc Natl Acad Sci U S A 2012;109(40):16173–8.
[93] Jost D, et al. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. Nucleic Acids Res 2014;42(15):9553–61.
[94] Alipour E, Marko JF. Self-organization of domain structures by loop-extruding enzymes. Nucleic Acids Res 2012;40(22):11202–12.
[95] Fudenberg G, et al. Formation of chromosomal domains by loop extrusion. Cell Rep 2016;15(9):2038–49.
[96] Nasmyth K. Disseminating the genome: joining, resolving, and separation sister chromatids during mitosis and meiosis. Annu Rev Genet 2001;35:673–745.
[97] Vian L, et al. The energetics and physiological impact of cohesin extrusion. Cell 2018;175(1):292–4.
[98] Haarhuis JHI, et al. The cohesin release factor WAPL restricts chromatin loop extension. Cell 2017;169(4):693–707.e14.
[99] Nagano T, et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. Nature 2017;547(7661):61–7.
[100] Wang S, et al. Spatial organization of chromatin domains and compartments in single chromosomes. Science 2016;353(6299):598–602.
[101] Bantignies F, et al. Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 2011;144(2):214–26.
[102] Osborne CS, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. Nat Genet 2004;36(10):1065–71.
[103] Papantonis A, et al. TNFalpha signals through specialized factories where responsive coding and miRNA genes are transcribed. EMBO J 2012;31(23):4404–14.
[104] Grimaud C, et al. RNAi components are required for nuclear clustering of Polycomb group responsive elements. Cell 2006;124(5):957–71.
[105] Denholtz M, et al. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. Cell Stem Cell 2013;13(5):602–16.

[106] Fanucchi S, et al. Chromosomal contact permits transcription between coregulated genes. Cell 2013;155(3):606–20.
[107] Zhao ZW, et al. Spatial organization of RNA polymerase II inside a mammalian cell nucleus revealed by reflected light-sheet superresolution microscopy. Proc Natl Acad Sci U S A 2014;111(2):681–6.
[108] Kang J, et al. A dynamical model reveals gene co-localizations in nucleus. PLoS Comput Biol 2011;7(7), e1002094.
[109] Rajapakse I, et al. The emergence of lineage-specific chromosomal topologies from coordinate gene regulation. Proc Natl Acad Sci U S A 2009;106(16):6679–84.
[110] Phair RD, Misteli T. High mobility of proteins in the mammalian cell nucleus. Nature 2000;404(6778):604–9.
[111] Rowley MJ, et al. Evolutionarily conserved principles predict 3D chromatin organization. Mol Cell 2017;67(5). 837–52.e7.
[112] Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol 2006;4(5), e138.
[113] Rosa A, Everaers R. Structure and dynamics of interphase chromosomes. PLoS Comput Biol 2008;4(8), e1000153.
[114] Chambeyron S, Bickmore WA. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev 2004;18(10):1119–30.
[115] Lomvardas S, et al. Interchromosomal interactions and olfactory receptor choice. Cell 2006;126(2):403–13.
[116] Spilianakis CG, et al. Interchromosomal associations between alternatively expressed loci. Nature 2005;435(7042):637–45.
[117] Fuss SH, et al. Local and cis effects of the H element on expression of odorant receptor genes in mouse. Cell 2007;130(2):373–84.
[118] Osborne CS, et al. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. PLoS Biol 2007;5(8), e192.
[119] Rocha PP, et al. Close proximity to Igh is a contributing factor to AID-mediated translocation. Mol Cell 2012;47(6):873–85.
[120] Zhang Y, et al. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. Cell 2012;148(5):908–21.
[121] Raccaud M, Suter DM. Transcription factor retention on mitotic chromosomes: regulatory mechanisms and impact on cell fate decisions. FEBS Lett 2018;592:878–87.
[122] Naumova N, et al. Organization of the mitotic chromosome. Science 2013;342(6161):948–53.
[123] Deluz C, et al. A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev 2016;30(22):2538–50.
[124] Festuccia N, et al. Mitotic binding of Esrrb marks key regulatory regions of the pluripotency network. Nat Cell Biol 2016;18(11):1139–48.
[125] Dileep V, et al. Topologically associating domains and their long-range contacts are established during early G1 coincident with the establishment of the replication timing program. Genome Res 2015;25(8):1104–13.
[126] Hsiung CC, et al. A hyperactive transcriptional state marks genome reactivation at the mitosis-G1 transition. Genes Dev 2016;30(12):1423–39.
[127] Du Z, et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 2017;547(7662):232–5.
[128] Flyamer IM, et al. Single-nucleus hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. Nature 2017;544(7648):110–4.
[129] Ke Y, et al. 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell 2017;170(2). 367–381.e20.
[130] Hug CB, et al. Chromatin architecture emerges during zygotic genome activation independent of transcription. Cell 2017;169(2). 216–228.e19.

[131] Robert Finestra T, Gribnau J. X chromosome inactivation: silencing, topology and reactivation. Curr Opin Cell Biol 2017;46:54–61.
[132] Darrow EM, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. Proc Natl Acad Sci U S A 2016;113(31):E4504–12.
[133] Giorgetti L, et al. Structural organization of the inactive X chromosome in the mouse. Nature 2016;535(7613):575–9.

A static view of chromatin architecture

Over the past decades, many studies have assessed the spatial proximity and nuclear organization of specific genomic loci, using microscopic techniques, such as fluorescent *in situ* hybridization (FISH) (Giorgetti & Heard, 2016; McCord, Kaplan, & Giorgetti, 2020), or molecular biology techniques, such as variants of chromosome conformation capture (3C) (Dekker et al., 2002), adapted to high-throughput sequencing (Hi-C) (Lieberman-aiden et al., 2009). Collectively, these studies demonstrated a correlation between chromatin topology and underlying gene activity. 3C-based approaches provide genome-wide contact maps, usually performed on populations of millions of cells, generating snapshots of the population and thus complementary methods are needed to investigate heterogeneity or variability at the single-cell level.

The 3C-based methods involve formaldehyde crosslinking, followed by ligation of the overhangs generated by digestion with a restriction enzyme and determination of pair-wise contacts using either PCR or sequencing (Denker & De Laat, 2016; Wit & Laat, 2012). The number of detected ligation products is a measure of how often the two corresponding genomic sequences were in close physical proximity, in a cell population and at the time point of the fixation. Depending on the biological question being asked, chromosome conformation capture-based approaches can be divided into "one- to-one" (3C) (Dekker et al., 2002), "one-to-all" (circular chromosome conformation capture combined with high-throughput sequencing (4C-seq)) (Van De Werken et al., 2012), "many-to-many" (5C (Dostie et al., 2006), ChIA-PET (Fullwood et al., 2009), Capture-HiC (Hughes et al., 2014)) and "all-to-all" (high-throughput chromosome conformation capture (Hi-C) (Lieberman-aiden et al., 2009)) (**Figure 15**). After a reverse crosslinking step, different approaches can be used to detect chromatin interactions. For example, the 4C-seq protocol entails a second round of digestion and re-ligation is utilized, followed by inverse PCR with specific primers to detect all of the chromatin interactions involving a specific locus of interest (the "bait") (Simonis et al., 2006; Van De Werken et al., 2012). However, the crosslinking and ligation steps used in 3C-based approaches can potentially produce biased results. For that purpose, ligation-free GAM (genome architecture mapping) (Beagrie et al., 2017) and crosslinking- and ligation-free DamC (Redolfi

et al., 2019) techniques have been developed, both methods showing good agreement with 4C-seq and Hi-C data.

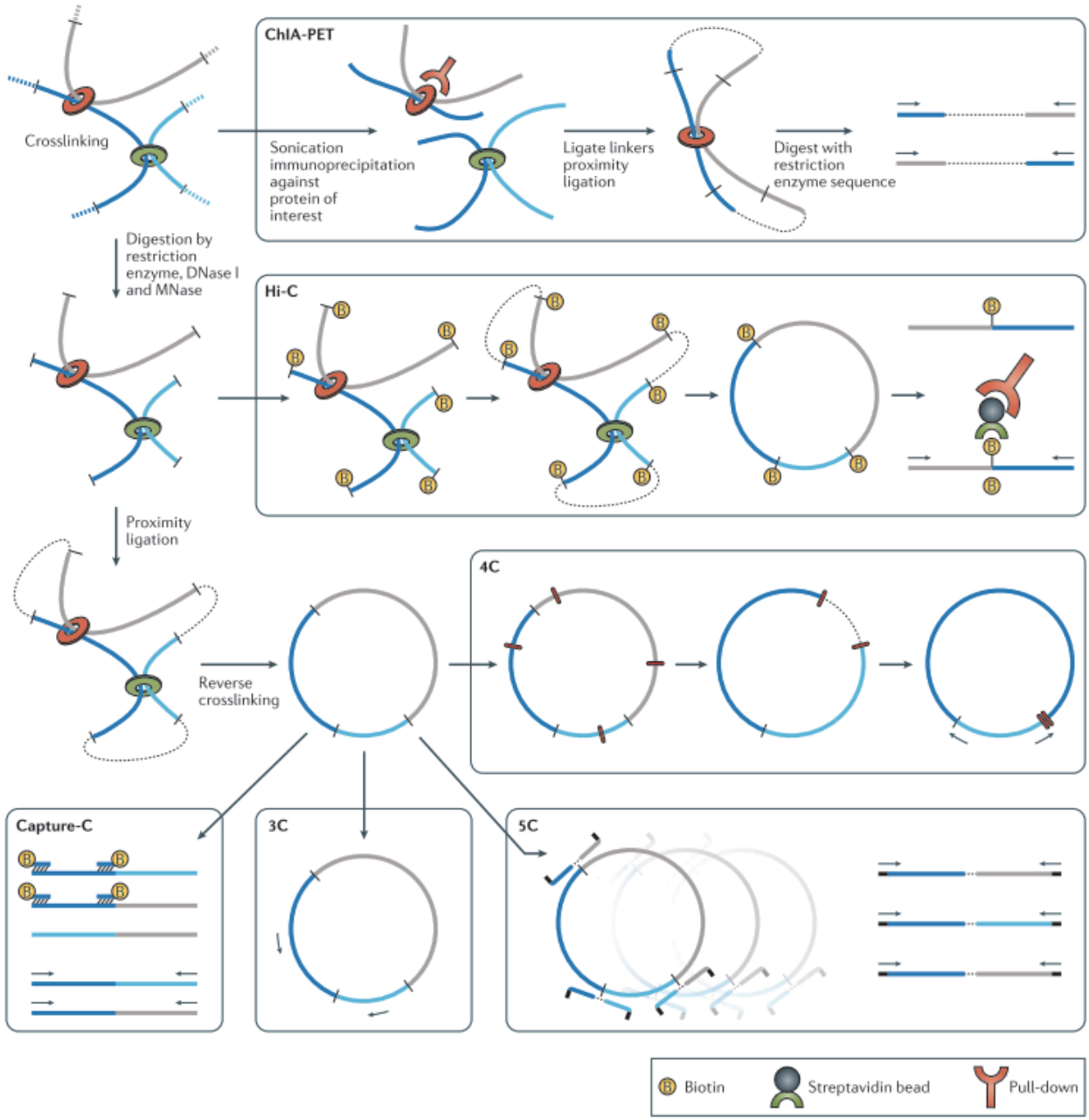


Figure 15: 3C-based methods to study chromatin architecture. Adapted from (Bonev & Cavalli, 2016).

Microscopy-based methods, such as DNA FISH, have largely confirmed 3C-based findings (Bintu et al., 2018; Cardozo Gizzi et al., 2019; Finn et al., 2019; Giorgetti et al., 2014; E. P. Nora et al., 2012). However, DNA FISH and 3C are conceptually different techniques that provide essentially different and complementary types of information (Giorgetti & Heard, 2016). DNA FISH provides a powerful means to directly determine the actual positions, distribution, 3D distances between genomic loci, as well as their position relative to TF/Pol II foci (when coupled with immunostaining) inside single cells, an inaccessible information in 3C-based experiments (Giorgetti & Heard, 2016). Nevertheless, conventional FISH is limited to probing few candidate loci, due to the restricted number of spectrally distinguishable fluorophores that can be simultaneously imaged by standard microscopes. Additionally, FISH probes are designed for each target region of interest and can thus detect only few genomic loci in parallel. These limitations constrain the genomic coverage achievable by DNA FISH. Although 3C-related techniques remain much superior in coverage, FISH has nonetheless uncovered many fundamental principles of genome organization, such as the preferential association between similarly regulated genes (Schoenfelder et al., 2010).

While these methods have provided major insights into 3D genome organization, recent advances using single-cell biochemistry (Cattoni et al., 2017; Flyamer et al., 2017; Nagano et al., 2013; Stevens et al., 2017) and super-resolution microscopy on fixed cells (Beliveau et al., 2015; Bintu et al., 2018; Cattoni et al., 2017; Finn et al., 2019; Mateo et al., 2019; Nir et al., 2018; Su, Zheng, Kinrot, Bintu, & Zhuang, 2020; Szabo et al., 2018; S. Wang et al., 2016), have demonstrated high variability of chromatin conformations between individual cells. Results from these single-cell methods confirmed compartmentalization (S. Wang et al., 2016) and the physical existence of TADs, initially characterized by population-average Hi-C, albeit with high spatial heterogeneity (Bintu et al., 2018; Mateo et al., 2019; Szabo et al., 2018; S. Wang et al., 2016), formed by a plethora of low-frequency interactions rather than one stable chromatin loop (Cattoni et al., 2017; Finn et al., 2019; Giorgetti et al., 2014) (**Figure 16**).

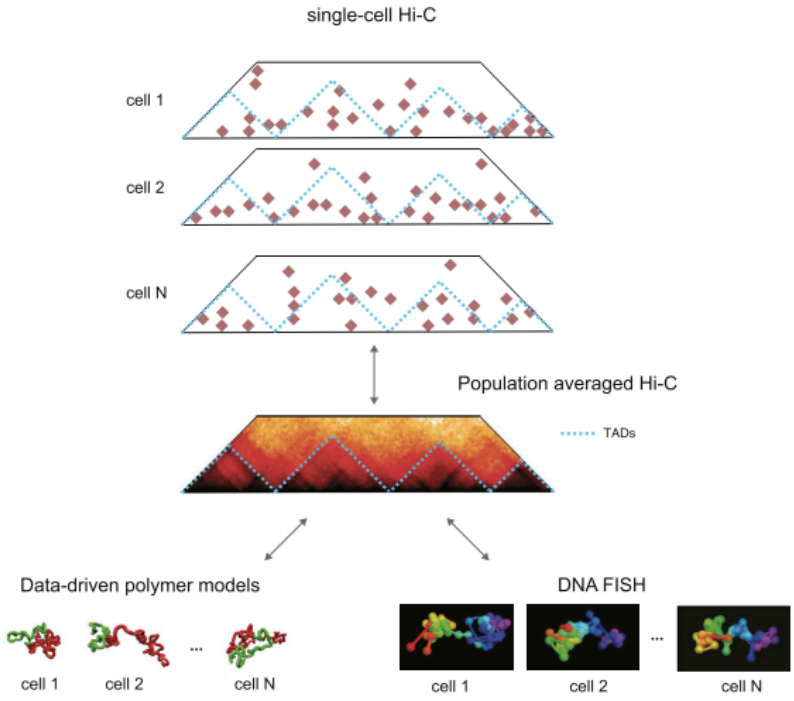


Figure 16 : Chromatin structure in single cells. Single-cell Hi-C, polymer simulations and super-resolution DNA FISH (Bintu et al., 2018) demonstrate that population-averaged chromosome conformations arise from highly variable conformations of the chromatin fiber that occur simultaneously in single cells. Adapted from (McCord et al., 2020).

Interestingly, the boundaries are highly variable across cells and depletion of cohesin, which ablates TADs at the population level (Rao et al., 2017), did not diminish TAD-like structures in individual cells (Bintu et al., 2018). The relatively low-frequency interactions in the population and the occurrence of diverse conformations of the chromatin fiber in individual cells suggest the coexistence of a wide spectrum of genome configurations across a cell population (Finn et al., 2019), in agreement with previous physical models which have demonstrated that interaction maps are the average of multiple diverse genome conformations (Giorgetti et al., 2014; Kalhor, Tjong, Jayathilaka, Alber, & Chen, 2011). Overall, Hi-C and imaging studies on fixed cells have uncovered a wealth of information on chromatin architecture, but the emerging models to explain links between topology and genome control all imply a certain dynamic nature of the underlying chromatin fiber to adopt transient and metastable conformations.

Chromatin dynamics

Visualizing DNA in living eukaryotic cells

Whereas chromosomal structures have been deeply characterized in the last years (Bonev et al., 2017; Hsieh et al., 2020; Krietenstein et al., 2020; Rao et al., 2014; Sexton et al., 2012) there remains a knowledge gap in how these structures alter over time, despite growing appreciation that chromatin dynamics likely play key roles in transcriptional regulation. Technical and analytical hurdles have, until relatively recently, limited assessment of locus-specific gene mobility. The first generation of locus-tagging methods entailed integration of multiple tandem copies of repetitive lacO sequences, bound by ectopic fluorescent-tagged lac repressor (W. F. Marshall et al., 1997; Robinett et al., 1996), but these very large constructs (~10 kb) are laborious to clone and may alter the chromatin state of the tagged locus (M. Dubarry, Loïodice, Chen, Thermes, & Taddei, 2011) (**Figure 17A**). Due to technical difficulties in targeting the lacO constructs to specific regions, the first chromatin labelling experiments only showed general principles of their dynamics in metazoan nuclei: constrained and anomalous movement within a restricted volume, and mobility generally reduced even further in heterochromatin, consistent with results on super-resolution studies of chromatin mobility by tagging histones (Nozaki et al., 2017). Subsequently, several editing-free systems such as fluorescent zinc finger proteins or TALEs (transcription activator like-effectors) (Ma, Reyes-Gutierrez, & Pederson, 2013; Miyanari, Ziegler-Birling, & Torres-Padilla, 2013) (**Figure 17B**) and CRISPR-dCas9 (catalytically dead Cas9) (B. Chen et al., 2013; Ma et al., 2015) (**Figure 17C**) constructs were designed to directly bind the endogenous sequences of interest, but it has been very challenging to obtain robust signals outside of repetitive heterochromatic sequences.

More recent approaches have proposed alternative ways to overcome the limitation of repetitive loci. Wang *et al*. (H. Wang et al., 2019) developed CRISPR LiveFISH, an approach that deploys Cy3-labeled sgRNAs assembled with dCas9-EGFP (enhanced GFP) *ex vivo* as fluorescent ribonucleoproteins (fRNPs), which are delivered to cells by electroporation and allow labelling of genomic sequences in living cells. Importantly, the signal-to-noise-ration (SNR) of gRNA was 4-fold higher than that of dCas9-EGFP and gRNA did not accumulate in

the nucleolus, unlike dCas9-EGFP. While this approach has a great potential for real-time visualization of DNA and RNA transcripts in living cells, through the combination of Cas9 and Cas13 (see below), it has only been shown to work at repetitive regions so far (**Figure 17C**).

Chimeric array of gRNA oligonucleotides (CARGO) coupled with dCas9, is another noninvasive strategy that overcomes the issue of SNR through the use of 12 gRNAs which recruit 12 dCas9-EGFP molecules to a genomic region spanning around 2 kb (Gu et al., 2018) (**Figure 17C**). Interestingly, CARGO-dCas9 allows imaging of non-repetitive loci contrary to classical CRISPR-dCas9 or CRISPR LiveFISH methods. An alternative approach for imaging non-repetitive regions in living cells is the CRISPR-Tag system (B. Chen, Zou, Xu, Liang, & Huang, 2018) in which a repeating DNA tag, with a minimal size of 250 bp, is inserted near the locus of interest. The integrated repeat sequence are binding sites for four sgRNAs (**Figure 17C**). Instead of the standard dCas9-EGFP, dCas9 is fused with 14 copies of GFP11 tags (dCas9-GFP_{14x}), a non-fluorescent fragment of GFP that upon complementation with GFP1-10 becomes a functional fluorescent protein (Kamiyama et al., 2016). CRISPR-Tag system represents the shortest available DNA tag, and thus is less likely to interfere with the structure and/or function of the target locus. Nevertheless, binding of Cas9 to the DNA implies DNA unwinding, which might affect the localization of histones (Sternberg, Redding, Jinek, Greene, & Doudna, 2014). Moreover, another disadvantage of the system is the fact that GFP1-10 has to be expressed exogenously.

The ANCHOR/ParB DNA labeling system is an alternative to the previous approaches. The system is based on the kinetochore-like nucleoprotein complexes, whose function is to ensure mitotic stability of bacterial chromosomes and large plasmids (N. Dubarry, Pasta, & Lane, 2006; Passot, Calderon, Fichant, Lane, & Pasta, 2012; Sanchez et al., 2015). ANCHOR entails introduction of a short (~1kb), non-repetitive DNA sequence (ANCH or *parS*) to which fluorescently-tagged OR (bacterial partition protein or ParB) bind site-specifically (Germier et al., 2017; Saad et al., 2014). Then, ORs self-oligomerize via N-terminal protein-protein interaction and recruit further OR dimers which bind non-specifically and relatively weakly to adjacent DNA and spread along chromatin to produce a robust signal (**Figure 17A**). These features, together with the small size of the binding site, allow the insertion of the ANCH sequence immediately adjacent to regulatory elements with minimal perturbation of endogenous chromatin (Saad et al., 2014).

Importantly, orthologous parS/ParB pairs allow for simultaneous double-label experiments (Germier et al., 2017; Oliveira et al., 2021; Saad et al., 2014).

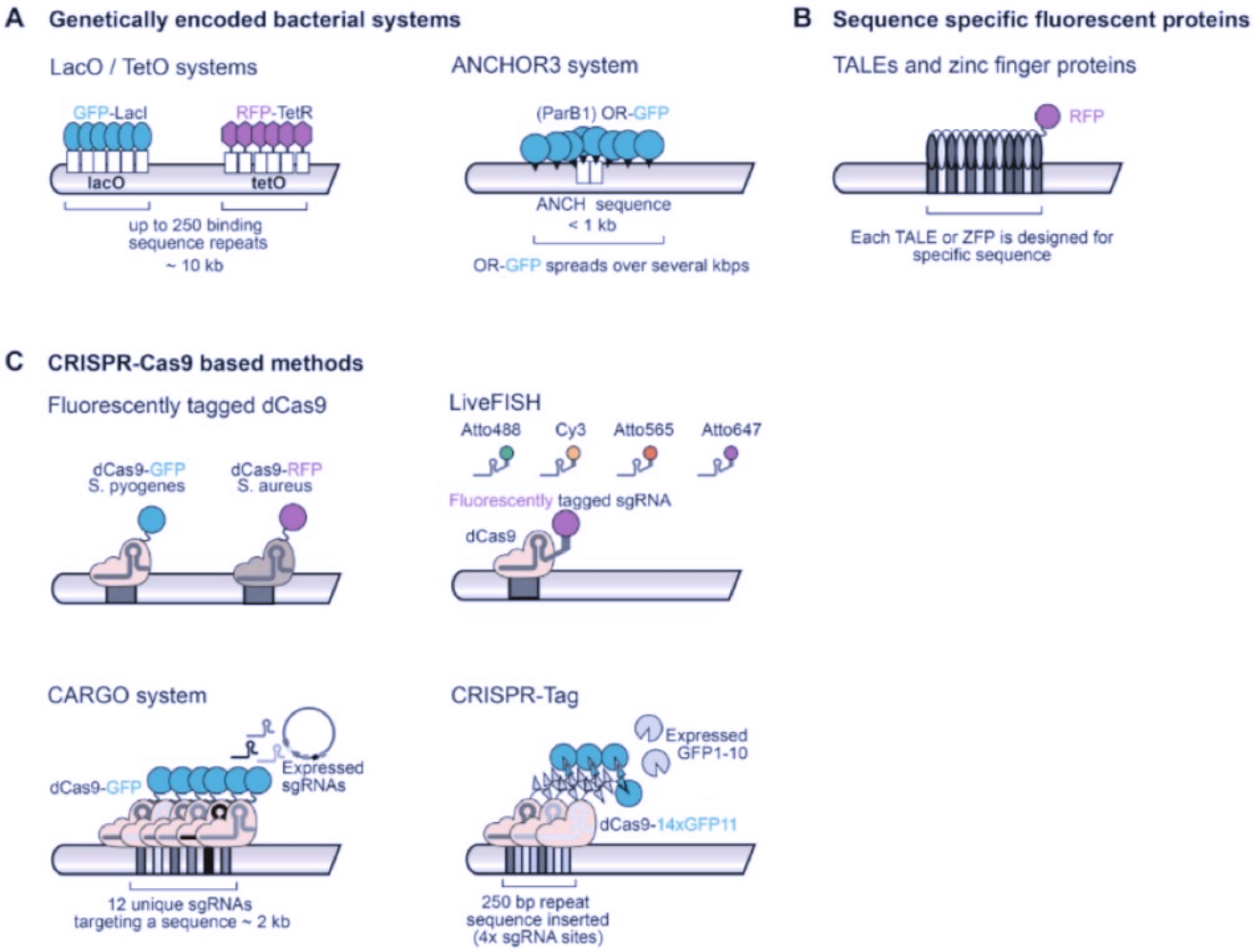


Figure 17: Systems to fluorescently tag genomic loci in living cells. (A) LacO/TetO systems require the insertion of large (~10 kb) repeat binding sequences into the genome in order to visualize the locus. The ANCHOR/ParB DNA system entails introduction of a small (~1kb), non-repetitive DNA sequence to which fluorescently-tagged OR bind and spread over the surrounding chromatin. (B) Fluorescently-labeled TALEs and zinc finger proteins (ZFs) are designed to bind to a specific endogenous locus of interest. (C) CRISPR-Cas9 based approaches to visualize genomic loci in living cells using the original fluorescently tagged dCas9. Adapted from (Shaban & Seeber, 2020).

Visualizing RNA in living eukaryotic cells

Recent advances in live-imaging methods have enabled the direct visualization of nascent transcripts at single-cell resolution. The MS2-MCP system has been largely used to visualize single RNAs in real time in living cells (Bertrand et al., 1998; Fusco et al., 2003; Germier et al., 2017; Golding et al., 2005; Larson et al., 2011; Yunger et al., 2010). Transcripts are engineered to encode a repetition of 24 19-bp RNA stem-loops that are recognized *in vivo* by a fluorescently-tagged bacterial coat protein MS2 (Bertrand et al., 1998; Fusco et al., 2003; Yunger et al., 2010) (**Figure 18A**). These RNA-protein interactions have been thoroughly studied, enabling the fine-tuning of the technique for different applications. However, a common concern is whether genetic insertion of MS2 aptamers into gene loci can affect RNA structure, expression or function. Other methods using fluorogenic RNA aptamers, such as Spinach (Paige, Wu, & Jaffrey, 2011) and Broccoli (Filonov, Moon, Svensen, & Jaffrey, 2014), provide background-free signals but have been mostly applied in bacteria. Molecular beacons, whose fluorescence increases after target binding, have also been used to visualize endogenous, untagged RNA transcripts (M. Chen et al., 2017). Nevertheless, their use is costly and requires their exogenous supply to the cells.

CRISPR-Cas13 is a newly identified RNA-programmable and RNA-binding RNase protein family and catalytically inactive Cas13s (dCas13s) fused to a fluorescent protein have been used for precise RNA targeting and imaging in living mammalian cells (Abudayyeh et al., 2017; Cox et al., 2017; Yang et al., 2019) (**Figure 18B**). dLwaCas13a has been initially used to label the abundant *ACTB* mRNA after stress (Abudayyeh et al., 2017), although a latter study proposed dPspCas13b and dPguCas13b to be the most efficient dCas13s that allow robust (comparable signals to MS2-MCP) and rapid real-time RNA tracking in living cells (Yang et al., 2019). In the same study, application of orthogonal dCas13s or combination of dCas13 and MS2-MCP allowed simultaneous dual-color label of RNAs, while combination of dCas13 with dCas9 enabled RNA-DNA visualization in living cells. While this tool has not been yet applied in a wide range of studies, its high efficiency, robustness and the absence of genetic manipulations offer great promise.

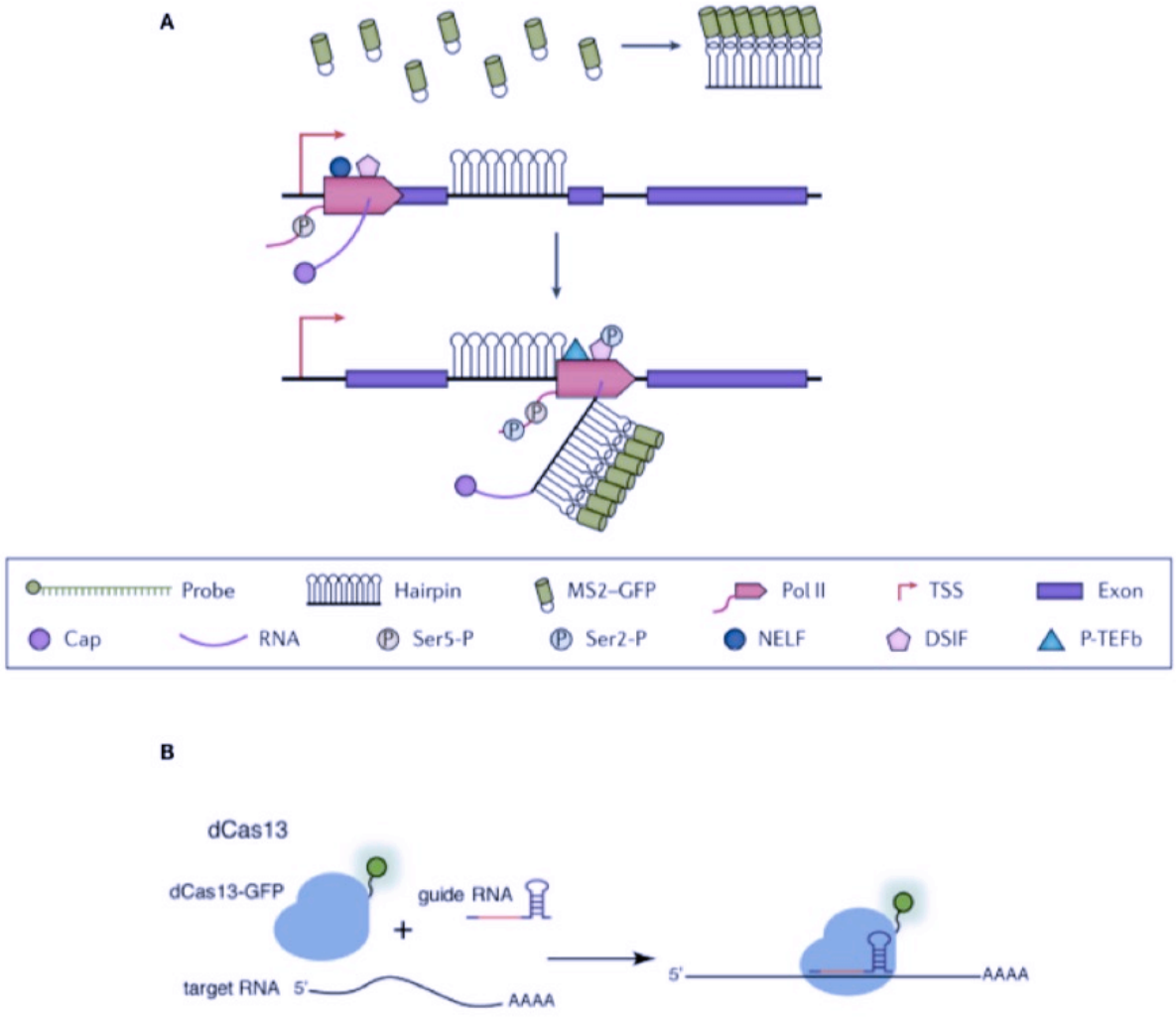


Figure 18: Systems to fluorescently visualize RNA in living cells. (A) GFP- tagged protein MS2 (MCP) binds an array of hairpin-forming sequences engineered into the RNA of interest. Adapted from (Wissink, Vihervaara, Tippens, & Lis, 2019). (B) Fluorescently-tagged dCas13 fused binds target RNAs in the presence of a gRNA. Adapted from (Pichon, Lagha, Mueller, & Bertrand, 2018).

Single-gene dynamics during transcription

Recent tagging of individual genes and enhancers has given seemingly conflicting views on the effect of transcription on chromatin dynamics. On the one hand, transcription coincided with reduced mobility and confinement to smaller nuclear volume (Germier et al., 2017; Ochiai, Sugawara, & Yamamoto, 2015). In particular, Germier *et al.* (Germier et al., 2017)

combined the ANCHOR3 DNA-labeling system with MS2-labeled mRNAs to study the motion of an estrogen-inducible CyclinD1 transgene in human cells before and after the appearance of MCP-EGFP-labeled mRNAs in the same cell. Analysis of the motion of the ANCHOR3-tagged locus before and after stimulation by E2 (17 β-estradiol) showed that while the motion of the transgene domain remained sub-diffusive in the absence of transcription, transcription initiation by RNA Pol II rapidly confined the mRNA-producing gene. Interestingly, confinement was maintained upon treatment of the cells with the transcriptional elongation inhibitor DRB (5,6-dichloro-1-b-D-ribofuranosyl- benzimidazole). In contrast, addition of the transcriptional initiation inhibitor, triptolide, to the transcriptionally stimulated cells released locus confinement. These results suggest that initiating but not elongating RNA Pol II induces confinement of chromatin, consistent with engagement of the gene within a restricted nuclear transcriptional hub. However, in other experiments, genes and enhancers activated upon differentiation were found to move faster, and were slowed on treatment with drugs inhibiting transcriptional initiation or elongation (Gu et al., 2018; Nozaki et al., 2017).

Gu *et al*. (Gu et al., 2018) applied the CARGO-dCas9 imaging system to monitor the dynamics of the *Fgf5* enhancer and promoter in mESCs and two days after inducing differentiation to epiblast-like cells (mEpiLCs), where the enhancer is activated and *Fgf5* is induced (Buecker et al., 2014). Their results demonstrated that both enhancer and promoter mobility increases in the mEpiLC state and since the *Fgf5* is activated during mESCs to mEpiLCs differentiation they proposed that transcriptional activation increases locus motion. In contrast to the results provided by Germier *et al*. (Germier et al., 2017), these authors reported that acute perturbation of RNA Pol II activity by DRB and flavopiridol, to target transcriptional elongation, or triptolide, to inhibit transcriptional initiation, significantly reduced the mobility of the locus. Collectively, the increased motion of the locus was interpreted as an increase in thermal energy via the active transcription process, which may aid gene search for regulatory sequences within the TAD, rather than the formation of stable enhancer-promoter loops. Interestingly, an earlier study has reported that chromatin dynamics tend to increase during early mESC differentiation (Masui et al., 2011). In order to distinguish whether these observations were linked to the activity status of the *cis*-regulatory elements or due to differences in global chromatin organization between mESCs and mEpiLCs, the authors (Gu et al., 2018) performed CARGO-dCas9 imaging at the *Tbx3* locus, where the gene becomes down-regulated during mESC to mEpiLC transition. Imaging of *Tbx3* promoter and distal super-enhancer revealed changes in

mobility that were opposite to those detected at the *Fgf5* locus. Nevertheless, it has to be noted that nascent mRNA was not measured in this study (Gu et al., 2018) , thus the locus is assumed as always being transcriptionally active at the mEpiLC state, yet without having evidence for mRNA production in the analyzed cells.

Notwithstanding, Neumann *et al*. (Neumann et al., 2012) measured the movement and the transcriptional output of tagged genomic loci in budding yeast, to which transcriptional activators were targeted. According to the study, local targeting of the VP16, but not of the Gal4 acidic domain, increased chromatin mobility and this increase in motion did not correlate with the activity of elongating RNA Pol II. The authors proposed that local chromatin remodeling and nucleosome eviction, depended on the INO80 remodeler, rather than transcriptional activity drive chromatin mobility in yeast. Together, the link between chromatin dynamics and transcription largely remains unclear, not least because so few loci have been assessed.

Global chromatin diffusion dynamics

As illustrated by single particle tracking (SPT), interphase chromatin undergoes diffusive motion within the nucleus, however this motion is constrained such that a given segment of chromatin is free to move within a small subregion of the nucleus (W. F. Marshall et al., 1997). The constrained diffusion of the chromatin fiber is functionally important, since it enables a locus to explore for potential interaction partners: for example, for an enhancer to encounter its target promoter or for multiple genes to cluster into transcriptional hubs (discussed earlier in the introduction). Generally speaking, the simplest model to describe the diffusion of microscopic systems is Brownian motion, whereby movements are caused by random collisions of any particle of interest with neighboring smaller particles within the system. However, as a bulky and extremely long polymer interacting with the crowded nuclear environment, chromatin frequently displays sub-diffusive behavior, with more limited movement than for classical Brownian motion (Bancaud et al., 2009; Tortora, Salari, & Jost, 2020). Therefore, the mean squared displacement (MSD) of chromatin is expected to follow this relationship with time: $\langle (r(t) - r(0))^2 \rangle = 2nDt^{\alpha}$, with r(t) the n-dimensional vector representing the position of particles in time and $\langle . \rangle$ the time-average over one trajectory or the ensemble-average over many trajectories.

Two parameters thus describe the diffusion properties of chromatin: the apparent diffusion coefficient D, indicating the speed of motion of a particle in a given amount of time, and the anomalous diffusion co-efficient α, defined in the range 0< α <2, which points to the type of diffusion undertaken by the particle of interest α<1 (sub-diffusive behavior, with smaller values indicating greater constraint); α=1 (pure Brownian diffusion); α>1 (a more directed type of motion) (**Figure 19**).

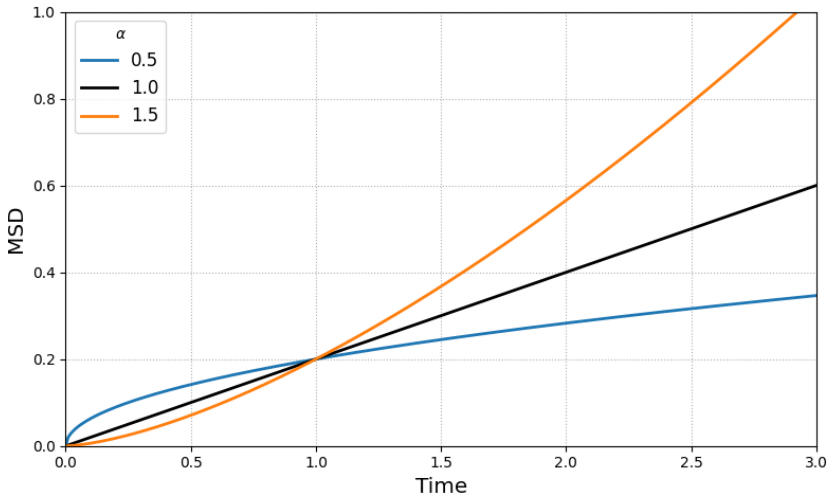


Figure 19: Comparison between different values of α and the expected MSD behavior.

Nonetheless, imaging chromatin within the crowded nuclear environment might be challenging, thus novel approaches, such as Dense Flow reconstruction and Correlation (DFCC) (Shaban, Barth, & Bystricky, 2018) and High Diffusion Mapping (Hi-D) (Shaban, Barth, Recoules, & Bystricky, 2020), have been developed to supplement SPT for studying bulk chromatin motion, although these are not adapted for the study of specifically tagged loci. Very recently, a collaboration between my group and the lab of Nacho Molina resulted in the development of a tool for the precise measurement of diffusive properties of such tagged chromatin loci, finding that local dynamics can significantly vary with genomic context (Oliveira et al., 2021). However, the dynamics of promoters and enhancers relative to transcriptional status was not stud

Aims of Study

Population-average "C" methods and DNA FISH support a model of chromatin loops bringing individual enhancers into contact with target genes, but this is challenged by recent findings showing enhancers at large distances from activated genes. Furthermore, it has become clear that static views of chromatin architecture are insufficient to fully understand dynamic processes such as transcription, or even maintenance of the architectures themselves. Despite growing anecdotal evidence that chromatin mobility varies with gene locus and transcriptional activity, very little is known about the "rules" governing chromatin dynamics, nor how they may affect gene regulation.

In this thesis I aimed to visualize enhancer-promoter communication in real time, coupled to labelling of nascent RNA, and to characterize chromatin dynamics at these loci. For this reason, I have adapted the ANCHOR/ParB technology, and have optimized the system for CRISPR/Cas9 knock-in of double labels into mouse embryonic stem cells. As a model for my experiments I used the locus around the pluripotency gene Sox2 and tagged its endogenous promoter and distal (>100 kb downstream) enhancer (SCR). The real-time behavior of these components were dissected in response to a battery of different perturbations, such as deletions of critical regulatory elements within the SCR, pharmacological inhibition of transcription and cell differentiation, in order to see whether and how promoter-enhancer distances and/or diffusive properties linked to transcriptional output.

Results

ANCHOR system allows visualization of specific loci without perturbing genome function

To assess enhancer-promoter proximity in real-time, we engineered ANCHOR (*parS*) tags (Germier et al., 2017; Saad et al., 2014) at the promoter and distal enhancer (SCR (Sox2 control region);(Zhou et al., 2014)) of *Sox2* in mouse embryonic stem cells. To ensure incorporation of the two tags on the same allele, we used the *M. musculus*/*M. castaneus* hybrid F1 ESC line (Mlynarczyk-Evans et al., 2006) and performed allele-specific knock-in of ANCH1 and ANCH3 tags with CRISPR/Cas9 at sites where SNPs result in the PAM only being present on the *musculus* allele. Due to the small size (<1 kb) of the ANCH sequences, we were able to incorporate a tag fully within the SCR, between the elements (SRR; *Sox2* regulatory region) containing the hallmarks of enhancer activity (Zhou et al., 2014) (**Fig 20a**; **Fig S1**). PCR screens, sequencing and visualization of the labels by microscopy all confirmed the allele-specific and unique incorporation of the ANCH tags into their expected sites (**Fig S2**). As a control, we also generated an ESC F1 line with *musculus* allele-specific incorporation of ANCH tags at sites shifted ~60 kb, labeling presumably non-regulatory regions with maintained genomic separation (**Fig 20a, b**; **Figs S1,S2a, b**). Virtual 4C plots derived from the high-resolution Hi-C map of ESCs (Bonev et al., 2017) confirm the *Sox2*-SCR interaction predicted from previous 3C and 4C studies (de Wit et al., 2013; Zhou et al., 2014), and show that the control tagged regions are not expected to form interactions (**Fig S2a**). Introducing ectopic sequences so close to regulatory regions could potentially interfere with their normal functioning, particularly if the sequences are tightly bound by proteins (M. Dubarry et al., 2011), although ANCHOR was previously shown to be minimally invasive in yeast (Saad et al., 2014). We confirmed that the ANCHOR system has no detectable effects on promoter and/or enhancer functioning in ESCs by various assays. Firstly, qRT-PCR showed that expression of pluripotency marker genes *Nanog* and *Pou5f1* (Oct4), as well as *Sox2* itself, is unaffected by incorporation of the ANCH sequences, even when the binding OR proteins are also introduced (**Fig 20b**). Allele-specific qRT-PCR also demonstrated that *Sox2* expression from either *musculus* (ANCH-labeled) or *castaneus* (unmodified) alleles are unaltered (**Fig**

Figure 20

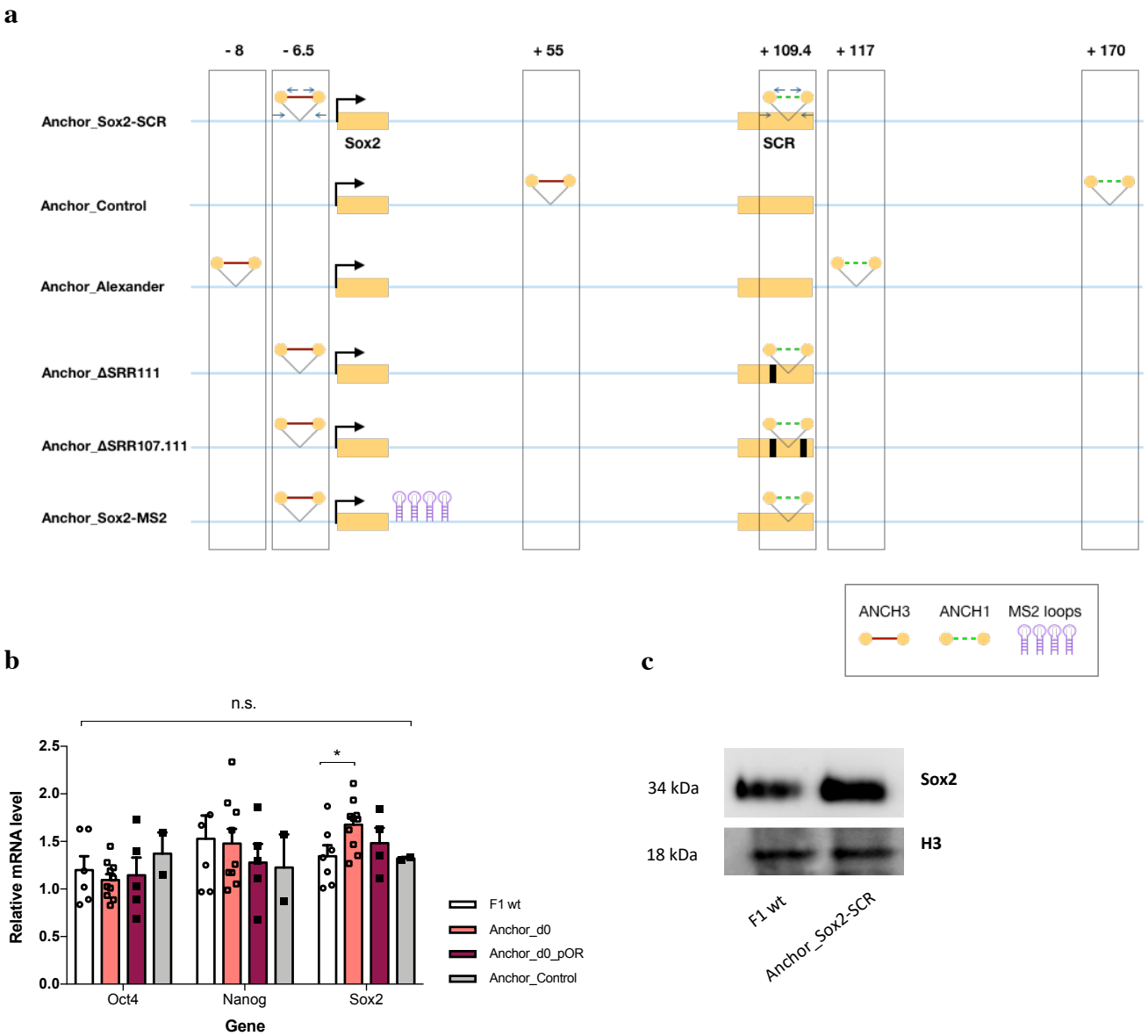


Figure 20: ANCHOR is non-invasive in mouse ESCs. a) Schematic of the different ESC lines generated in this work, showing the positions of ANCH1 (green) and ANCH3 (red) tags, relative to the *Sox2* promoter (the distances from the transcription start site, in kb, are given above the tag), and the SCR. Black rectangles denote the positions of SCR sub-deletions, and purple stem loops indicate the inclusion of MS2 repeats at the *Sox2* 3'UTR. **b**) qRT-PCR results, normalized to SDHA housekeeping gene, for F1, Anchor_Sox2-SCR (with or without transfection of OR proteins) and Anchor_Control ESCs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **c**) Western blot for Sox2 in F1 and Anchor_Sox2-SCR ESCs, using histone H3 as a loading control.

Figure 21

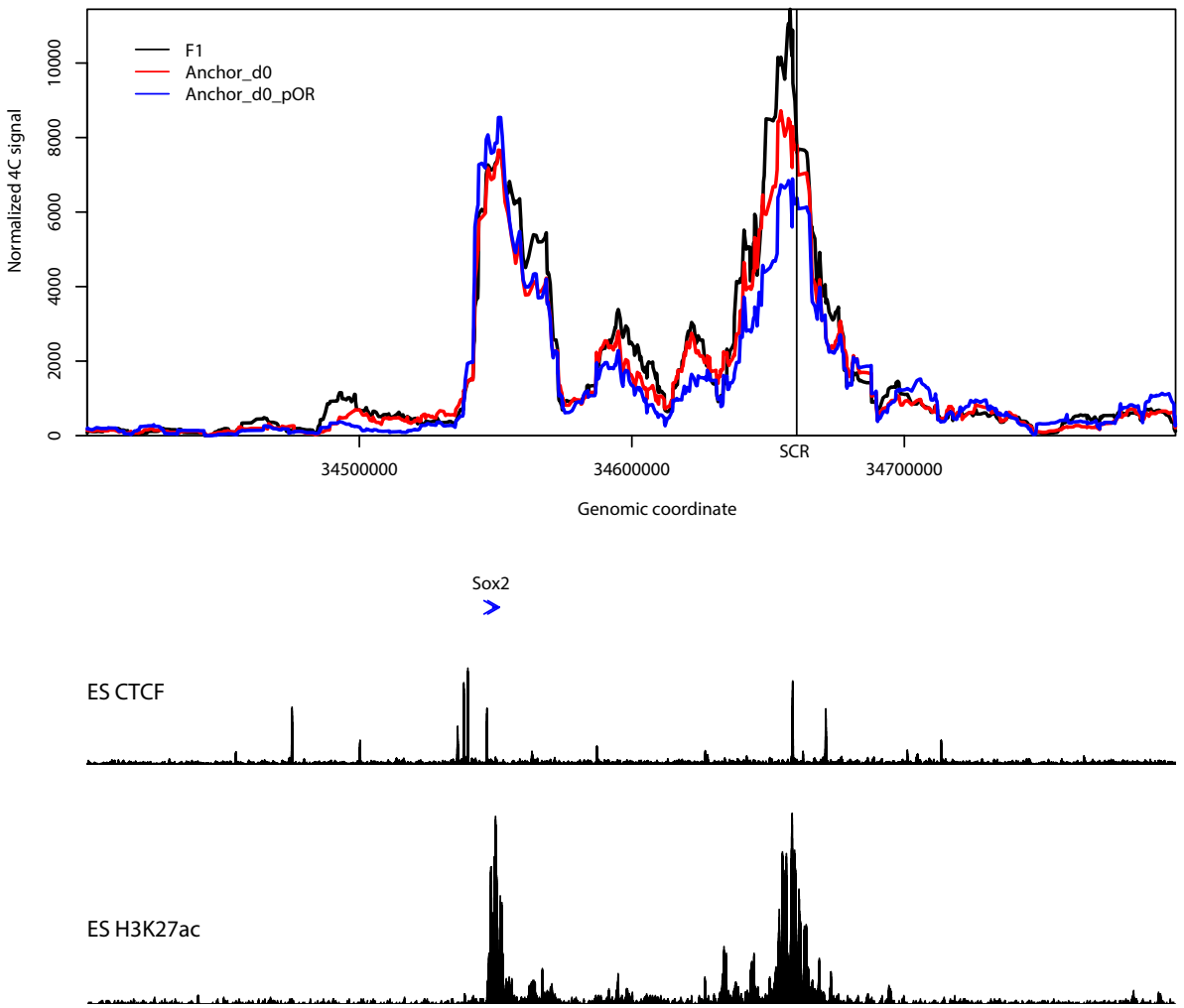


Figure 21: ANCHOR does not alter chromatin topology. Allele (*musculus*)-specific 4C results, with the SCR as bait, for F1 and Anchor_Sox2-SCR (with or without transfection of OR proteins) ESCs. Position of Sox2 gene is shown underneath in blue, and ChIP-seq tracks for H3K27ac and CTCF (GEO: GSE49847) are also shown below.

S2c). Secondly, western blot showed that Sox2 protein levels were similarly unaffected by the presence of ANCH sequence (**Fig 20c**). Finally, allele (*musculus*)-specific and canonical 4C showed that apparent *Sox2*-SCR interaction levels are unaffected by promoter/enhancer tagging (**Fig 21**; **Fig S3**). Thus the ANCHOR system does not alter labeled gene expression or local chromatin architecture, either in *cis* or *trans*.

Enhancer-promoter proximity is accompanied by greater constraint of chromatin regulatory elements

To assess enhancer-promoter interaction dynamics in living cells, we performed spinning disc confocal microscopy on ANCH-labelled ESCs, obtaining images every 500 ms for ~2-3 minutes. We observed a consistent proximity between enhancer and promoter, but with fluctuations in their exact spatial arrangement and direct juxtaposition (**Fig 22a**). We did not identify large-scale spatial chromatin rearrangements within the timeframe of individual movies (median interquartile range for distances is 95 nm; maximum interquartile range for distances within one movie is 153 nm; **Fig S4a**). With our current setup, large nuclear contractions indicative of phototoxicity became apparent with longer imaging times, so we were unable to reliably track enhancer-promoter interactions over longer periods. However, comparisons of distance distributions between movies, while showing extensive cell-to-cell heterogeneity, revealed that the majority of cells maintain close proximity of promoter and enhancer (**Fig S4a**; see also **Appendix 1**). The average *Sox2*-SCR distance was significantly shorter than the distance between the control regions (median 146 nm for *Sox2*-SCR, 195 nm for control; **Fig 22b**), in agreement with the interaction differences suggested by Hi-C. However, we note a high heterogeneity of inter-label distances between cells, as previously reported (Finn et al., 2019), and that the control labels are also rather close, as may be expected for a genomic separation of ~115 kb.

A previous study reported much larger distances between *Sox2* and its downstream enhancer (mean 339 nm, compared to mean 165 nm that we observed) when they were labelled with large repeats of tetO and cuO sequences, suggesting that promoter-enhancer proximity was not required for transcriptional activation (Alexander et al., 2019). To investigate further, we generated a new F1 ESC line with ANCH labels integrated at the exact same positions as

Figure 22

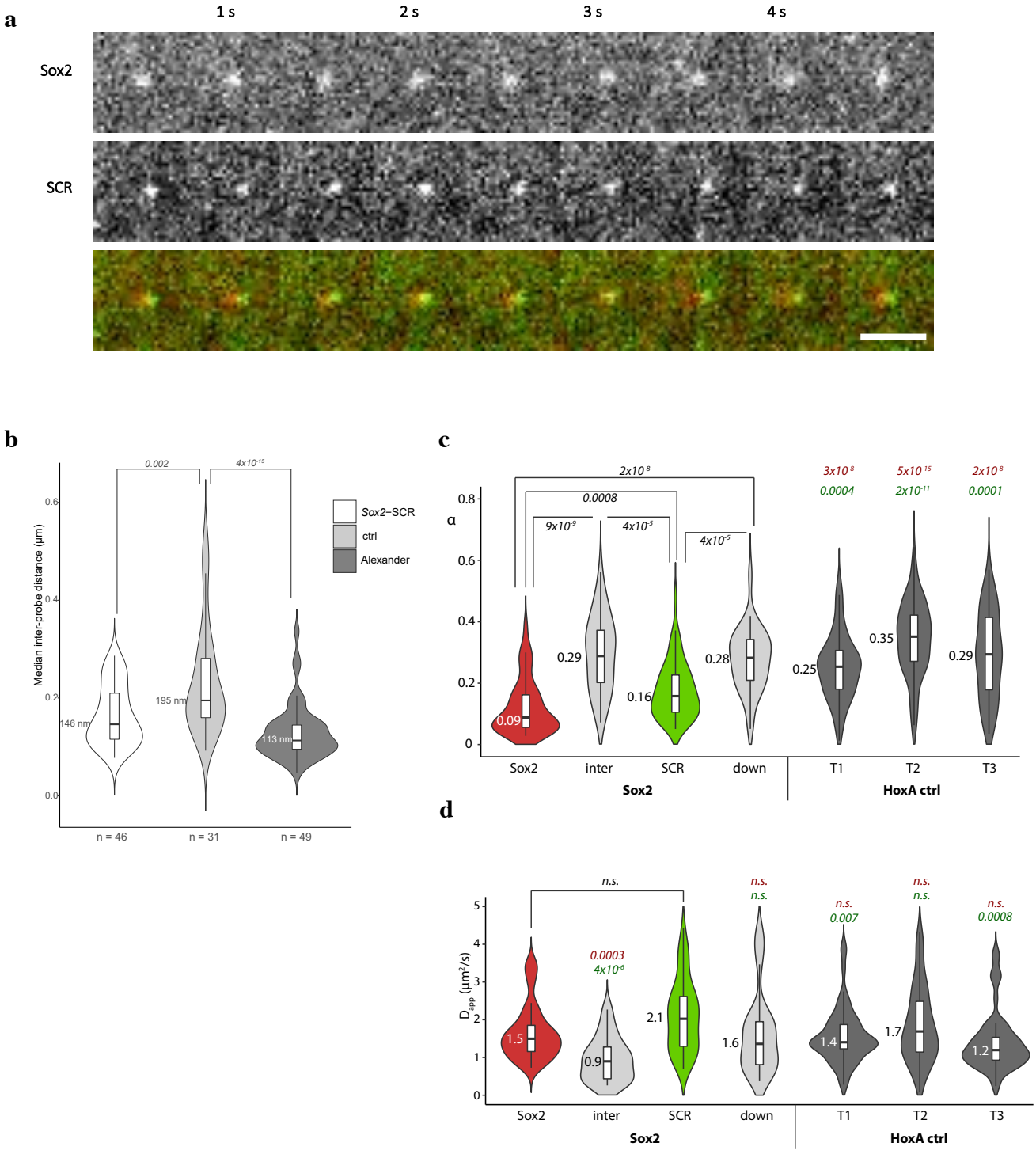

a
1 s
2 s
3 s
4 s
Sox2
SCR
b
Median inter-probe distance (µm)
0.002
4x10-15
Sox2–SCR
ctrl
Alexander
146 nm
195 nm
113 nm
n = 46
n = 31
n = 49
c
2x10-8
0.0008
9x10-9
4x10-5
4x10-5
3x10-8
0.0004
5x10-15
2x10-11
2x10-8
0.0001
0.09
0.29
0.16
0.28
0.25
0.35
0.29
Sox2
inter
SCR
down
T1
T2
T3
Sox2
HoxA ctrl
d
Dapp (µm²/s)
n.s.
n.s.
n.s.
n.s.
0.007
n.s.
n.s.
0.0008
0.0003
4x10-6
1.5
0.9
2.1
1.6
1.4
1.7
1.2
Sox2
inter
SCR
down
T1
T2
T3
Sox2
HoxA ctrl

Figure 22: Close proximity and constrained motion of *Sox2* promoter and enhancer. a) Representative time-lapse images of an Anchor_Sox2-SCR cell, simultaneously visualizing ANCH3/*Sox2* and ANCH1/SCR. Merged images (*Sox2* in red, SCR in green) shown below. Scale bar: 2 μm. **b**) Violin plots for median inter-probe distance distributions for Anchor_Sox2-SCR (white), Anchor_Control (light grey) and Anchor_Alexander (grey) cells. Median values and numbers of analyzed movies are given on the figure. Numbers in italics denote the p-value for pairwise comparisons (Wilcoxon rank sum test). **c**) Violin plots for anomalous diffusion coefficient distributions for *Sox2* (shades of red), SCR (shades of green), and control loci within the Sox2 locus (light grey) or the HoxA locus (grey). Median values are given on the figure. Numbers in italics denote the q-value for pairwise comparisons within the Sox2 locus; for the HoxA loci, numbers in red indicate the q-value for pairwise comparison with *Sox2* and numbers in green indicate the q-value for pairwise comparison with SCR (Wilcoxon rank sum test, with Benjamini-Hochberg multiple testing correction). **d**) Violin plots for apparent diffusion speed distributions for *Sox2* (shades of red), SCR (shades of green), and control loci within the Sox2 locus (light grey) or the HoxA locus (grey), exactly as for **c** (*n.s.* denotes non-significant; $q > 0.05$).

the tetO/cuO sites (*Sox2* site 3 kb further upstream of the promoter than our insertion site; SCR site outside and a further 9 kb downstream; **Fig 20a; Figs S1-2**). These sites are similarly expected to participate in chromatin interactions based on Hi-C data (**Fig S2a**), and total and allele-specific *Sox2* expression is also unaltered in this cell line (**Fig S2c**). In line with our ANCHOR results and contrary to what was previously reported, these labels are also frequently proximal, with a median distance (113 nm) actually slightly smaller than in the original ANCHOR line, and significantly smaller than the control label distance (**Fig 22b**). We propose therefore that the difference in reported inter-label distances in our results and those of the previous study are due to technical differences, perhaps as a result of using the large repetitive tet/cu operators. In particular, the previous study also reported a rather large distance (mean ~250 nm) between the two operators when their genomic separation was reduced to ~14 kb by a large deletion in the intervening region (Alexander et al., 2019), around double the average distances we observe in the wild-type state.

As well as assessing how physical distances fluctuate with time, the relatively high temporal resolution of our movies also allows us to measure the local diffusive properties of the chromatin at the different tagged loci. We previously reported that apparent diffusion speed (D_{app}) and the anomalous diffusion parameter (α) can significantly vary between different chromatin loci (Oliveira et al., 2021), although the underlying principles determining these differences were unclear. We applied GP-FBM (Gaussian Processes applied to Fractional Brownian Motion; Oliveira et al., 2021) to measure D_{app} and α, and found that both the *Sox2* promoter and enhancer are significantly more constrained than control regions within the locus (**Fig 22c**). Further, the *Sox2* promoter is significantly more constrained than the SCR. The measurements of α are virtually identical between the two Sox2 ANCH integration sites and between the two SCR ANCH integration sites (**Fig S4b**). In a previous study, we measured the local chromatin diffusive properties of three loci near the HoxA locus, finding one locus (termed T1), which is <15 kb from a putative enhancer, to be significantly more constrained than the other two (T2 and T3) in ESCs (Oliveira et al., 2021). The *Sox2* promoter and enhancer are significantly more constrained than all of the HoxA regions, and the control regions within the Sox2 locus have similar α to the other inactive regions around the HoxA locus (T2 and T3) (**Fig 22c**). Apparent diffusion speeds also differed across loci (**Fig 22d**), albeit with less obvious links to underlying chromatin state. Within the Sox2 locus, the SCR had the highest diffusive speed (median 2.1 $\mu m^2/s$) and the control region between the gene and enhancer had the lowest (median 0.9 $\mu m^2/s$); diffusive speeds were very similar between *Sox2* and the

downstream control region, as well as to the control loci within the HoxA locus, in between these two extremes. Notably, whereas the intervening control region diffuses significantly more slowly than all other tested loci, the faster measured diffusion speed of the SCR is not statistically significant for all other compared loci. Further, whereas the *Sox2* diffusion speeds were essentially identical for the two alternative ANCH labels, there was a very large discrepancy between measurements for the two SCR labels: the label precisely within the SCR had an overall higher diffusion speed than the other loci, whereas the label 9 kb downstream had a much slower diffusion rate (median 0.7 $\mu m^2/s$), statistically similar to the intervening control region (**Fig S4c**). Overall, active chromatin elements, such as promoters and enhancers, appear to have more constrained chromatin motion than nearby, inactive regions within the same locus or topologically associated domain, and their relatively shorter separation distances reflect chromosome conformation capture data. Apparent diffusion speeds can also vary with genomic locus, but there is no obvious direct link to underlying transcriptional or regulatory activity.

Sox2 inactivation during differentiation coincides with reduced chromatin constraints

In our previous study of the HoxA locus, local chromatin diffusive parameters were altered when differentiation was induced with retinoic acid (Oliveira et al., 2021). We similarly assessed the effects of *in vitro* differentiation on enhancer-promoter proximity and chromatin dynamics at the Sox2 locus. As expected, removal of LIF (leukemia inhibitory factor) and 2i (a combination of two kinase inhibitors which maintain ESCs in a ground pluripotency state; (Ying et al., 2008)) from the medium, as well as supplementing with retinoic acid, causes rapid silencing of pluripotency factor expression, including *Sox2*, in ANCH-labelled ESCs, with steady transcriptional shutdown almost completed by the third day (**Fig 23a**). Allele-specific 4C shows a concomitant reduction in interaction between the ANCH-tagged *Sox2* promoter and SCR, although this is only apparent by day three of differentiation (**Fig 23b**). We thus performed live imaging experiments during the first three days of *in vitro* differentiation, where *Sox2* silencing has been achieved, and cells have not completely aggregated to prevent their imaging. Cells and nuclei at day 1 of the protocol had abnormal morphology, presumably due to insufficient recovery after transfection of the OR vectors, so this timepoint was not analyzed.

Figure 23

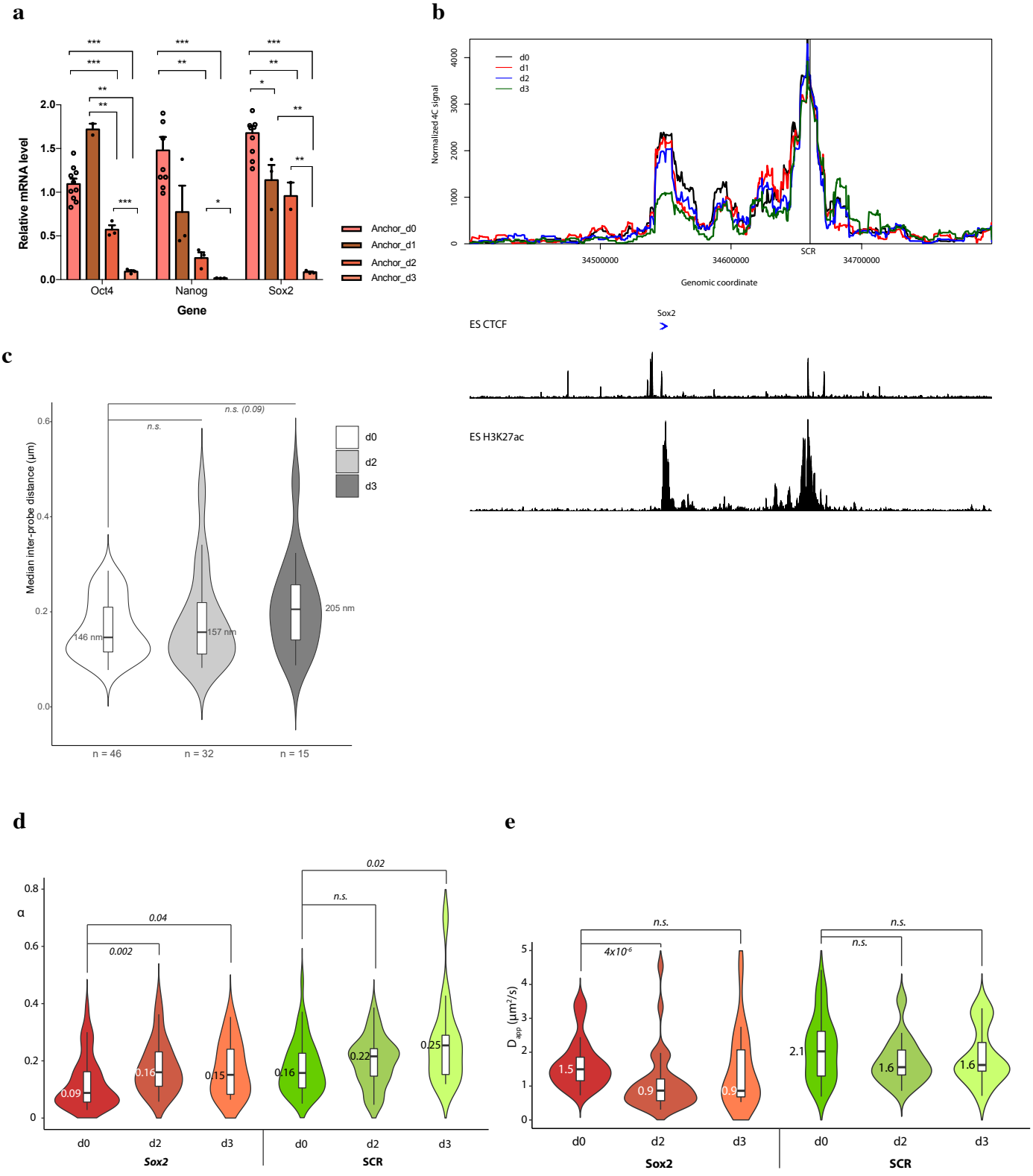

a
Relative mRNA level
Oct4
Nanog
Sox2
Gene
Anchor_d0
Anchor_d1
Anchor_d2
Anchor_d3
b
Normalized 4C signal
d0
d1
d2
d3
SCR
34500000
34600000
34700000
Genomic coordinate
ES CTCF
Sox2
ES H3K27ac
c
Median inter-probe distance (µm)
n.s. (0.09)
n.s.
146 nm
157 nm
205 nm
n = 46
n = 32
n = 15
d0
d2
d3
d
α
0.04
0.002
0.02
n.s.
0.09
0.16
0.15
0.16
0.22
0.25
Sox2
SCR
e
D app (µm²/s)
n.s.
4x10⁻⁵
n.s.
n.s.
1.5
0.9
0.9
2.1
1.6
1.6
Sox2
SCR

Figure 23: Local regulatory element dynamic changes on ESC differentiation. a) qRT-PCR results, normalized to SDHA housekeeping gene, for Anchor_Sox2-SCR ESCs during days 0-3 of *in vitro* differentiation. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **b**) Allele (*musculus*)-specific 4C results, with the SCR as bait, for Anchor_Sox2-SCR ESCs during days 0-3 of *in vitro* differentiation, exactly as for **Fig 21**. **c**) Violin plots for median inter-probe distance distributions, exactly as for **Fig 22b**, for ANCHOR_Sox2-SCR ESCs during days 0-3 of *in vitro* differentiation. **d,e**) Violin plots for **d**) anomalous diffusion coefficient and **e**) apparent diffusion speed distributions of *Sox2* and SCR in ANCHOR_Sox2-SCR ESCs during days 0-3 of *in vitro* differentiation, exactly as for **Figs 22c,d**, except that p-values without Benjamini-Hochberg correction are shown.

In agreement with the 4C results, the average promoter-enhancer distances are unaltered at day 2, where *Sox2* expression is still more than 50% of ESC levels (median 157 nm), but are increased by day 3 (median 205 nm) (**Fig 23c**). The difference is currently not statistically significant ($p = 0.09$; Wilcoxon rank sum test) because only 15 movies have been obtained for day three; more data acquisition is ongoing to obtain a sufficient sample size. We also observe changes in *Sox2* chromatin dynamics on differentiation: both the gene promoter and enhancer become less constrained (**Figs 23d**), in line with an apparent greater constraint at active elements (**Fig 22c**; Germier et al., 2017). Notably, the relaxation of the promoter had already occurred by day 2, before any obvious loss of enhancer proximity, as assessed by ANCHOR or 4C. Interestingly, diffusion speed was reduced significantly at the *Sox2* promoter, again a day before detectable enhancer proximity differences, with only a weak, non-significant reduction observed for the SCR (**Fig 23e**). Overall, these results suggest that promoter-enhancer proximity correlates somewhat with transcriptional output, but that local chromatin dynamics precede these changes. Promoters and enhancers appear to be more constrained when active, whereas the *Sox2* promoter but not the SCR has increased diffusive speed on activation. These experiments will be imminently performed on the cell line with labelled control probes, to dissect effects of gene inactivation from potential consequences of cell differentiation on chromatin in general.

Promoter diffusive speed is directly modulated by transcriptional initiation

To directly assess the effect of transcription on chromatin dynamics and promoter-enhancer interactions, we performed live imaging experiments on ANCH-labelled ESCs after treatment with inhibitor drugs: triptolide, which covalently binds to XPB subunit of TFIIH and mediates proteasomal degradation of promoter-bound RNA polymerase II before initiation, and flavopiridol, which inhibits CDK9 subunit of P-TEFb and hence phosphorylation of serine-2 of the RNA polymerase II C-terminal domain and subsequent transcriptional elongation (Bensaude, 2011). We incubated the cells with either drug (or DMSO control) under conditions specifically causing large reduction in mRNA levels of short-lived species, such as *Myc* or *Sox2*, without affecting levels of longer-lived mRNA species such as *Actb* (**Fig S5a, b; Table S1**), thus ensuring as far as possible treatments that acutely and efficiently block transcription with minimal pleiotropic effects. As may be expected, inhibitor treatments of the cells labelled

Figure 24

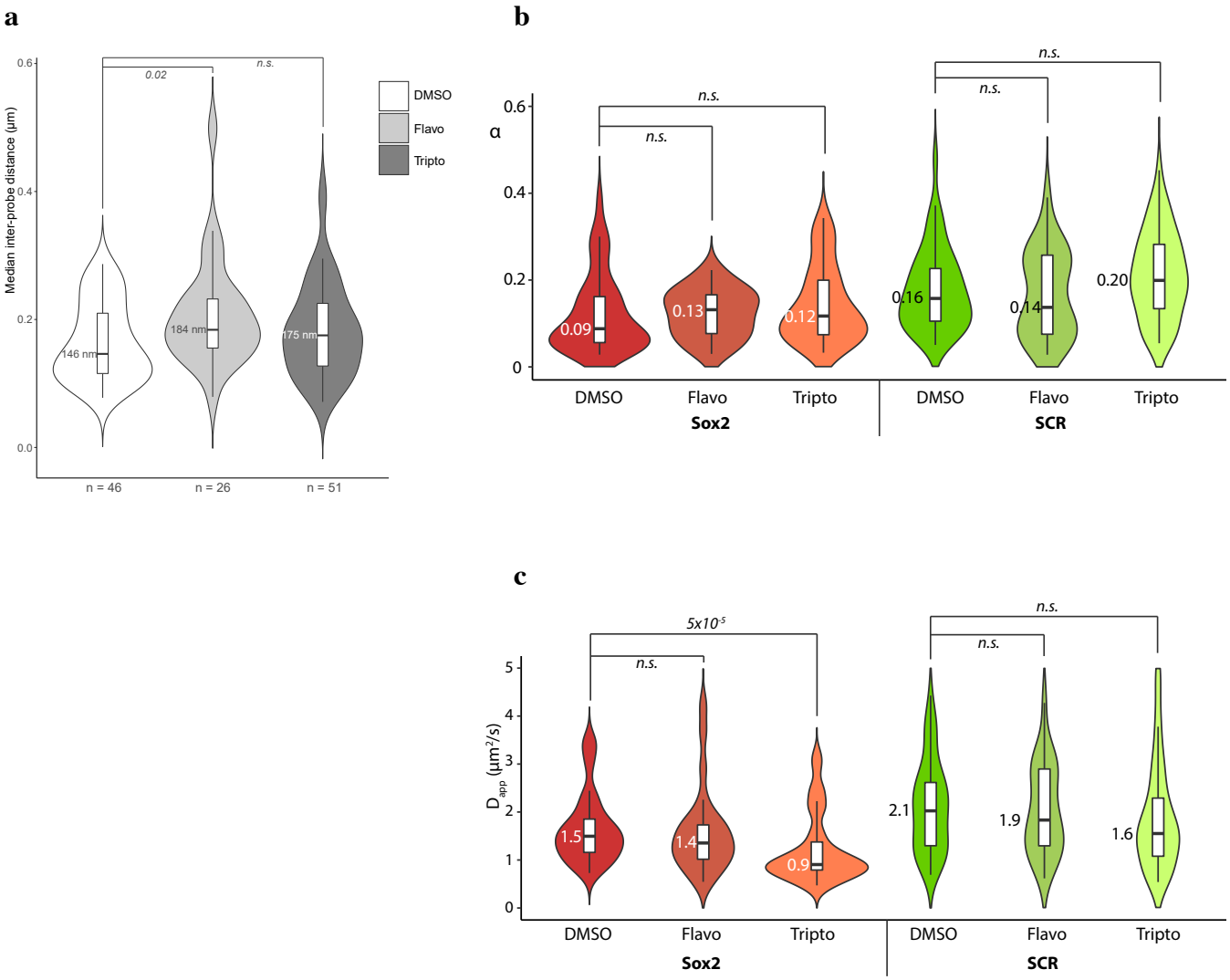


Figure 24: Inhibition of transcriptional initiation slows apparent promoter diffusion speed. a) Violin plots for median inter-probe distance distributions, exactly as for **Fig 22b**, for Anchor_Sox2-SCR ESCs after treatment with different drugs. **b,c**) Violin plots for **b**) anomalous diffusion coefficient and **c**) apparent diffusion speed distributions of *Sox2* and SCR in Anchor_Sox2-SCR ESCs after treatment with different drugs, exactly as for **Figs 23d,e**.

at control, non-transcribed regions, had no effect on average distance between the labels nor any measured diffusive parameter on either locus (**Fig S5c-e**). For the *Sox2*/SCR-labelled cells, inhibitors caused a weak increase in average distance, which is only statistically significant for flavopiridol treatment (p = 0.02; Wilcoxon rank sum test; **Fig 24a**), and a very weak, non-significant reduction in chromatin constraint on triptolide treatment (**Fig 24b**). On the other hand, triptolide but not flavopiridol treatment causes a strong reduction in promoter diffusion speed, with minimal effects of either drug on the SCR (**Fig 24c**). This finding, that inhibition of transcriptional initiation specifically reduces diffusive speed at promoters, partly agrees with a previous report that inhibition of either initiation or elongation reduced diffusive speed at both promoters and enhancers (Gu et al., 2018). Overall, most locus- and differentiation stage-dependent differences we observed in chromatin dynamics and promoter-enhancer proximity appear to only be partially explained by acute transcriptional processes. Conversely, RNA polymerase II engagement and/or promoter melting at transcriptional initiation appears to directly increase apparent diffusive speed at gene promoters.

Enhancer mutations link gene expression to local diffusive properties but not promoter proximity

Despite carefully controlled conditions, transcriptional inhibitor drugs may still have non-specific or pleiotropic effects. To more precisely assess the specific links between *Sox2* transcription and local chromatin architecture and dynamics, we performed live imaging after making *musculus*-specific deletions of SRRs within the SCR with CRISPR/Cas9 (**Fig 20a; Fig S1**) (Moorthy & Mitchell, 2016a). Deletion of SRR111 alone results in a weak (~1.4-fold) reduction in allele-specific *Sox2* expression, whereas a compound deletion of SRR111 and SRR107, maintaining the Intervening, ANCH1-labelled sequence, causes a strong (>4-fold) loss of *musculus Sox2* expression (**Fig 25a**). Expression of *Sox2* from the unmodified *castaneus* allele, or of Oct4 and Nanog, is unaffected (**Fig 25a**; **Fig S6**). Allele-specific 4C using a bait region near the SCR which is not deleted in any of the lines did not reveal any *Sox2*-SCR interaction differences between wild-type and Δ111 alleles, and only a weak reduction in interaction in the Δ107,111 allele (**Fig 25b**), in agreement with findings in analogous, non-ANCH-labelled cells (Shchuka et al., in prep). Neither deletion affected the average *Sox2*-SCR distance, largely in line with the 4C results (**Fig 25c**). Interestingly, although the deletions

Figure 25

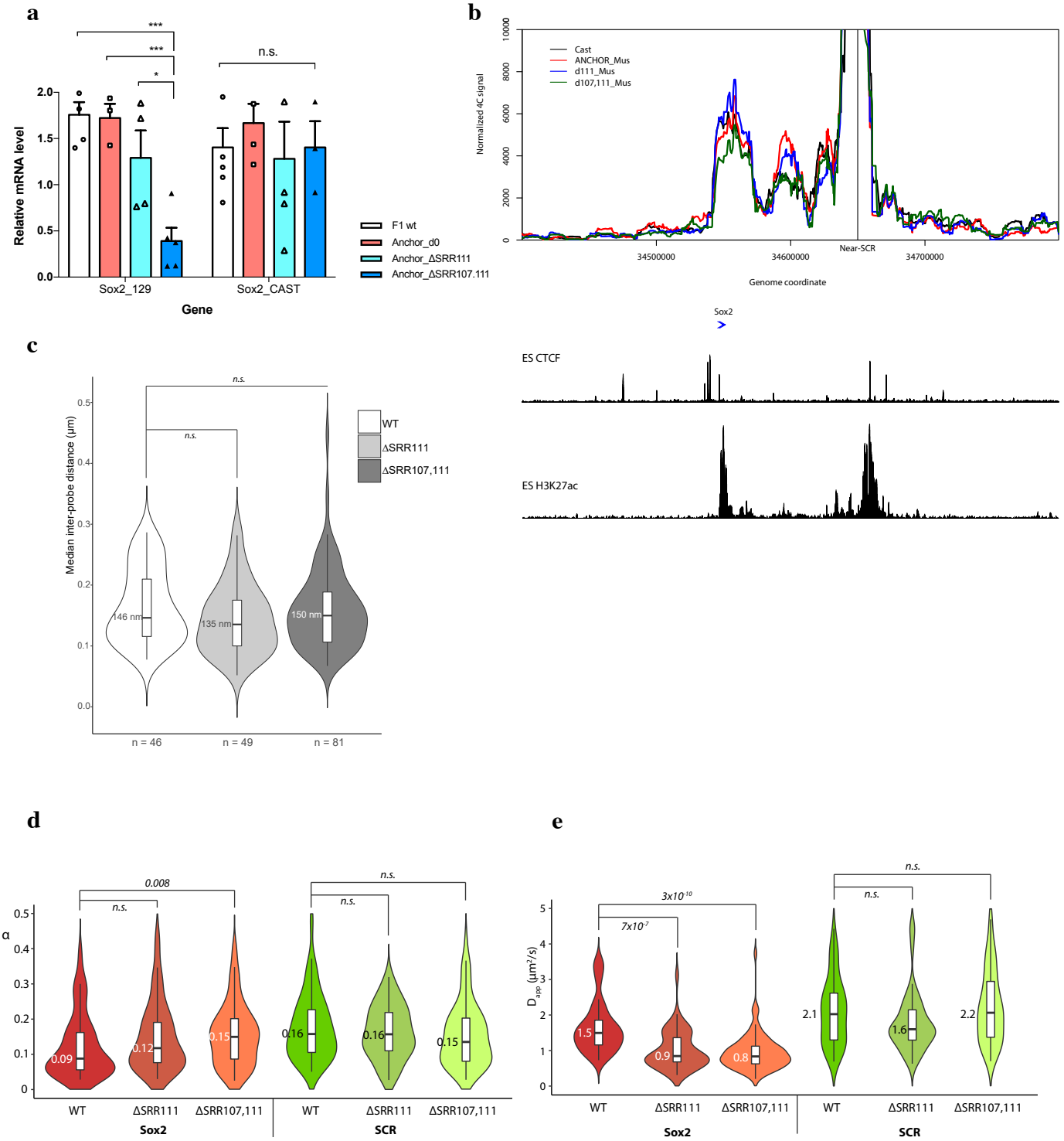

a
***
***
*
n.s.
Relative mRNA level
2.0
1.5
1.0
0.5
0.0
Sox2_129
Sox2_CAST
Gene
F1 wt
Anchor_d0
Anchor_ΔSRR111
Anchor_ΔSRR107.111
b
Cast
ANCHOR_Mus
d111_Mus
d107,111_Mus
Normalized 4C signal
10000
8000
6000
4000
2000
0
34500000
34600000
34700000
Near-SCR
Genome coordinate
Sox2
ES CTCF
ES H3K27ac
c
n.s.
n.s.
Median inter-probe distance (µm)
0.5
0.4
0.3
0.2
0.1
0.0
146 nm
135 nm
150 nm
n = 46
n = 49
n = 81
WT
ΔSRR111
ΔSRR107,111
d
0.008
n.s.
n.s.
n.s.
α
0.5
0.4
0.3
0.2
0.1
0
0.09
0.12
0.15
0.16
0.16
0.15
WT
ΔSRR111
ΔSRR107,111
WT
ΔSRR111
ΔSRR107,111
Sox2
SCR
e
3x10-10
7x10-7
n.s.
n.s.
Dapp (µm2/s)
5
4
3
2
1
0
1.5
0.9
0.8
2.1
1.6
2.2
WT
ΔSRR111
ΔSRR107,111
WT
ΔSRR111
ΔSRR107,111
Sox2
SCR

Figure 25: Promoter-specific local dynamics changes on enhancer disruption. a) Allele-specific qRT-PCR for Sox2 expression, normalized to SDHA housekeeping gene, in F1, Anchor_Sox2-SCR, Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **b**) Allele-specific 4C results, with a region flanking near the deleted SCR region as bait, for Anchor_Sox2-SCR, Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs, with the same annotations as in **Fig 21**. The *castaneus* results from the unmodified allele serve as an internal control. **c**) Violin plots for median inter-probe distance distributions, exactly as for **Fig 22b**, for Anchor_Sox2-SCR, Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs. **d,e**) Violin plots for **d**) anomalous diffusion coefficient and **e**) apparent diffusion speed distributions of *Sox2* and SCR in Anchor_Sox2-SCR, Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs, exactly as for **Figs 23d,e**.

remove binding sites for sequence-specific transcription factors and nucleation points for active histone modifications at the SCR, which might be expected to affect local chromatin fibre stiffness, neither deletion significantly affected local chromatin dynamics at the enhancer (**Figs 25d,e**). Conversely, the *Sox2* promoter has a significant slowing of diffusion speed in both deletion lines and reduced constraint, specifically in the compound deletion of SRR107 and SRR111. These results reinforce the link between local promoter diffusion speed, constraint of motion and transcription, but also show an apparent uncoupling with diffusive parameters at the spatially proximal enhancer, and that enhancers and promoters appear to remain close even when transcription is reduced.

Transcribing and non-transcribing alleles have similar chromatin dynamics

The results so far suggest that gene activity somehow has a direct but rather complex effect on local chromatin architecture and dynamics. To directly assess if and how transcribing loci differ from non-transcribing genes, we integrated 24 copies of the MS2 stem loop repeat into the 3'UTR of the *musculus Sox2* allele in ANCH-labelled F1 ESCs, allowing allele-specific transcripts to be visualized by binding of mScarletI-labelled MCP (see Methods; (Bertrand et al., 1998)). Triple-label live imaging thus allowed *Sox2* and SCR dynamics and proximity to be simultaneously tracked with gene transcription (**Fig 26**). A potential limitation of the MS2-MCP system is that the stem loops may disrupt expression via effects on transcription or downstream processes, such as mRNA stability or translation. This is a special concern for the *Sox2* gene, since it lacks introns and MS2 repeats can only be inserted into the 3'UTR, linked to regulation of polyadenylation, mRNA stability and translation efficiency. Indeed, a similar MS2-MCP setup in the mouse Sox2 locus has already been reported to heavily reduce Sox2 protein levels (Alexander et al., 2019). We assayed Sox2 expression in our own F1 ESC line (Anchor_Sox2-MS2, containing double-ANCHOR and MS2 integrations, as well as stable expression of both ANCHOR OR proteins and MCP (**Fig 26a**)) by different approaches. First, single-molecule RNA FISH was used to quantify the numbers of *Sox2* and MS2 mRNA species in individual cells, finding that wild-type and Anchor_Sox2-MS2 cells contained equivalent numbers of *Sox2* molecules, and that Anchor_Sox2-MS2 cells specifically contained ~50% the amount of MS2 transcripts (**Fig 26b**), implying equivalent transcription from both alleles,

Figure 26

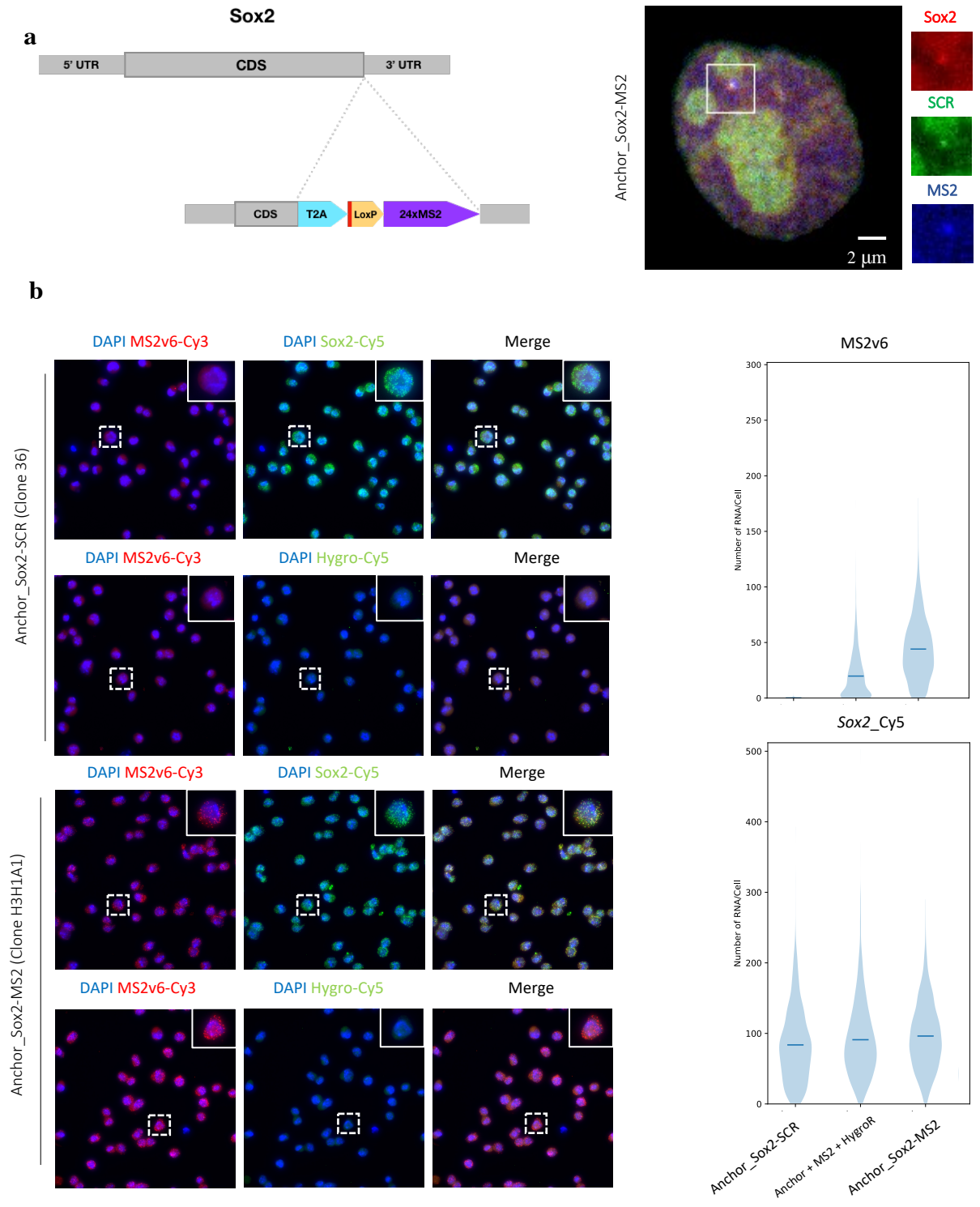

a
Sox2
5' UTR
CDS
3' UTR
CDS
T2A
LoxP
24xMS2
Anchor_Sox2-MS2
Sox2
SCR
MS2
2 µm
b
Anchor_Sox2-SCR (Clone 36)
Anchor_Sox2-MS2 (Clone H3H1A1)
DAPI MS2v6-Cy3
DAPI Sox2-Cy5
DAPI Hygro-Cy5
Merge
MS2v6
Sox2_Cy5
Number of RNA/Cell
Anchor_Sox2-SCR
Anchor + MS2 + HygroR
Anchor_Sox2-MS2
c
Sox2
SCR
MS2
d
Silent
Burst
Time (s)

Figure 26: Simultaneous visualization of promoter-enhancer dynamics and transcriptional bursting. a) Schematic of MS2 system incorporated at the 3' end of the *musculus* allele of *Sox2*, with a representative image of a triple-labeled cell, showing ANCH3/*Sox2* (red), ANCH1/SCR (green) and MCP/Sox2-MS2 (blue). Scale bar: 2 μm. **b**) Representative smRNA FISH maximum projection images (left) and violin plots for per-cell molecule quantitation (right) within Anchor_Sox2-SCR and Anchor_Sox2-MS2 ESCs for labeling of MS2 (red) with either Sox2 or excised hygromycin-resistance marker (green); DAPI staining of nuclei in blue. **c**) Representative time-lapse images of an Anchor_Sox2-MS2 cell, simultaneously visualizing ANCH3/*Sox2*, ANCH1/SCR and MCP/MS2. Merged images (*Sox2* in red, SCR in green, MS2 in blue) shown below. Scale bar: 2 μm. **d**) MS2 "barcode" for imaged Anchor_Sox2-MS2 ESCs. Each row represents a movie, and each column represents an acquired frame, with black shading indicating presence of a detectable MCP spot at the ANCHOR-labeled focus, and white indicating absence of MS2 detection. Zoomed in barcode for a single movie on the right shows easy distinction of refractory periods with extended absence of MS2 signal from sporadic loss of signal due to movement out of focal plane.

regardless of the MS2 tag. However, qRT-PCR and western blot results show an appreciable loss of steady-state *Sox2* mRNA (both at the tagged *musculus* and unlabeled *castaneus* alleles) and protein levels (**Figs S7a, b**). The maintained levels of nascent transcripts determined by single-molecule RNA FISH suggests that expression effects are predominantly post-transcriptional, reducing mRNA stability and/or translation efficiency, but since Sox2 protein regulates its own expression (Chew et al., 2005), and untagged mRNA levels are slightly reduced (**Fig S7a**), a slight inhibition of transcription compared to wild-type levels is likely. Notably, the previously published Sox2-MS2 system included a selectable marker gene (Alexander et al., 2019), which could further perturb transcription and mRNA stability. Our strategy for MS2 insertion also includes a hygromycin-resistance selection marker, but this is flanked by loxP sites and excised before imaging experiments (see Methods). We note that the line maintaining the selection marker has equivalent steady-state mRNA levels as measured by qRT-PCR, but ~50% nascent transcription as measured by single-molecule FISH and even more greatly reduced protein levels (**Figs S7a-c**).

We performed triple-label live imaging experiments, obtaining images every 3 s for ~10 minutes (**Fig 26c**). This setup provided sufficient temporal resolution for GP-FBM to robustly measure diffusive parameters, while spanning a long enough period for transcriptional bursts, which have a typical duration of a few minutes (Tunnacliffe & Chubb, 2020), to be detected. The imaging setup required for sufficiently rapid ANCHOR spot detection is incompatible with three-dimensional tracking, so MCP foci indicative of genes undergoing transcriptional bursting were frequently lost for sporadic intervals on movement out of the imaged focal plane. However, on visual inspection of frames with/without MCP detection, it was relatively easy to resolve bursting and non-transcribing events (**Fig 26d**). In total, ~25% of all frames had a detectable MCP signal, and of 54 imaged cells, 28 (52%) had defined bursting events, with a median duration of ~4.5 minutes. Since a large number of the bursts had already started at the onset of image acquisition, or had not yet finished when acquisition was stopped, this duration is likely an underestimate, but is within the expected range reported for other transcriptional bursting studies (Tunnacliffe & Chubb, 2020). Although greater numbers of movies are required to improve statistical confidence, this measured bursting frequency is much higher than reported in a previous study (~2/3 cells do not demonstrate bursting within a 30 min window; the vast majority of the rest have <20% frames with MCP detection; Alexander et al., 2019), further suggesting that the labelled system within Anchor_Sox2-MS2 better represents active Sox2 loci.

We extracted subsets of movies comprising the called bursting events (19 of sufficient duration for accurate measurement of diffusive parameters) and prolonged periods of absence of transcription (24; minimum duration 4.9 minutes), and performed GP-FBM to compare promoter-enhancer proximity and local chromatin diffusive properties between transcribing and non-transcribing loci. Although insufficient movies have so far been acquired to obtain statistical significance, *Sox2* and SCR are slightly closer when the gene is actively transcribing (**Fig 27a**). Indeed, when comparing transcribing and non-transcribing distributions to the total distribution from the double-label experiments, the transcribing subset of Sox2 loci have significantly closer *Sox2*-SCR distances, suggesting that at least for this locus, physical promoter-enhancer juxtaposition is linked to transcriptional activation. For the triple-labelled cell lines, both the *Sox2* promoter and the SCR had less constrained movement and slower diffusive speed than for the non-MS2-tagged cell line (**Figs S7d-e**). Since the MS2 tag causes reduced expression, this finding is in line with the other perturbation experiments showing that gene inactivation causes an overall reduction in D_{app} and increase of α (e.g. compare with the effects of SRR deletions on *Sox2* dynamics; **Figs 25d,e**). However perhaps surprisingly, there were no differences in constraint (**Fig 27b**) or apparent diffusive speed (**Fig 27c**) between transcribing and non-transcribing loci. Tracking of a greater number of cells will be required to lend greater support to these preliminary findings. *Sox2* contains high levels of paused RNA polymerase in mouse ESCs (Williams et al., 2015) (**Fig S1**). It therefore appears that paused and actively transcribing genes reside in a sufficiently similar nuclear microenvironment to have essentially the same local chromatin diffusive properties, further supported by unchanged dynamics on treatment with transcriptional elongation inhibitor, flavopiridol (**Figs 24b,c**).

Overall, chromatin diffusive properties is locus-specific and appears to be linked to underlying activity. Discussed in more detail in the following section, regulatory regions tend to have more constrained motion, and promoters have greater apparent diffusive speeds when active, without necessarily being coupled to changes in promoter-enhancer spatial separation.

Figure 27

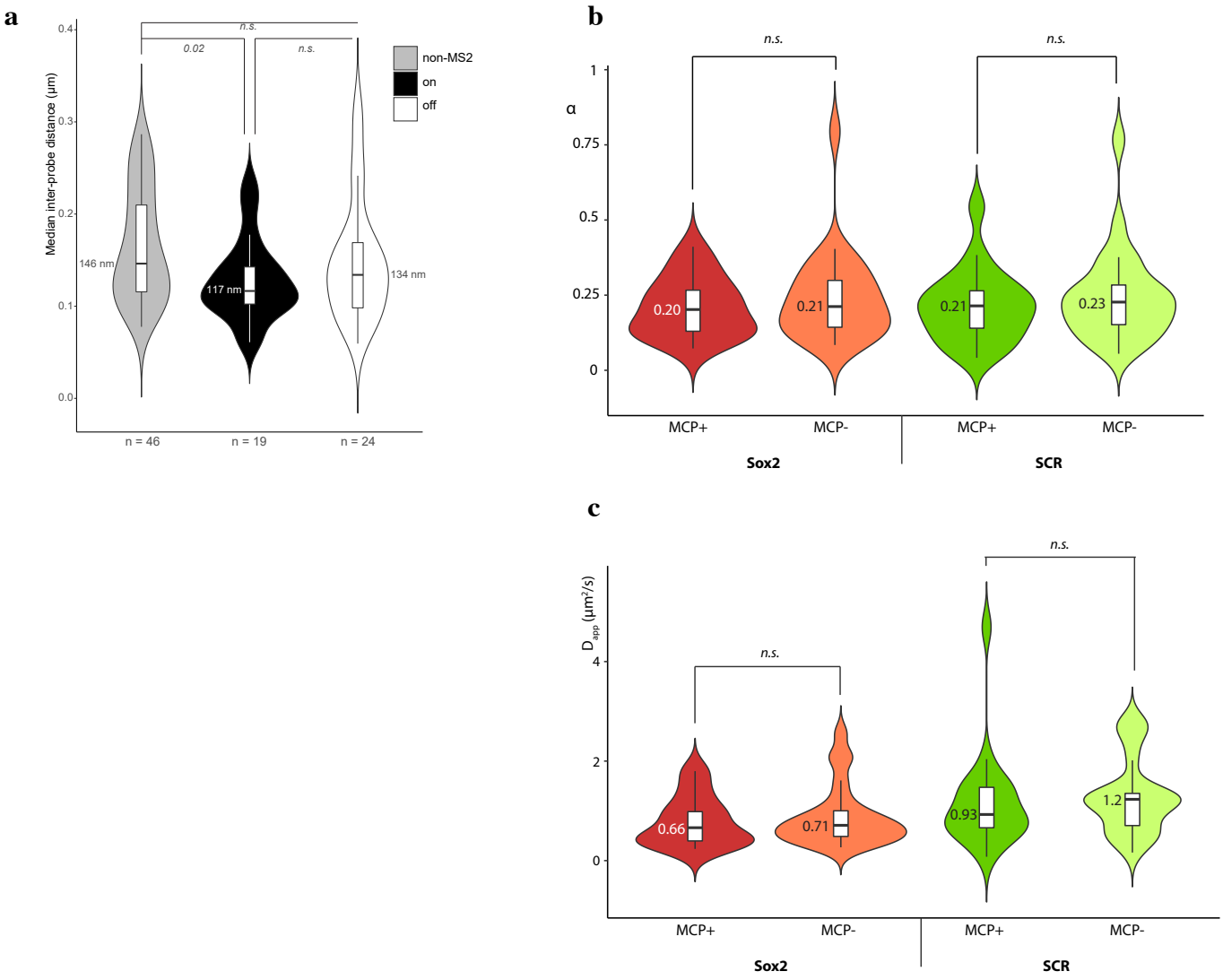
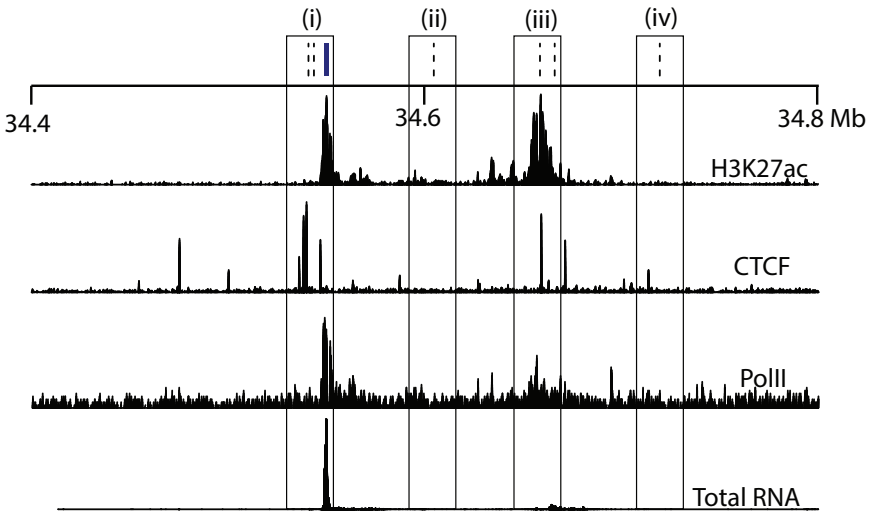


Figure 27: No apparent difference in local chromatin dynamics between bursting and non-bursting alleles. a) Violin plots for median inter-probe distance distributions, exactly as for **Fig 22b**, for Anchor_Sox2-SCR ESCs (grey) and Anchor_Sox2-MS2 ESCs, stratified according to the presence (black) or absence (white) of an MCP focus. **b,c**) Violin plots for **b**) anomalous diffusion coefficient and **c**) apparent diffusion speed distributions of *Sox2* and SCR in MCP-positive or -negative Anchor-MS2 ESCs, exactly as for **Figs 23d,e**.

Supplementary Figure S1



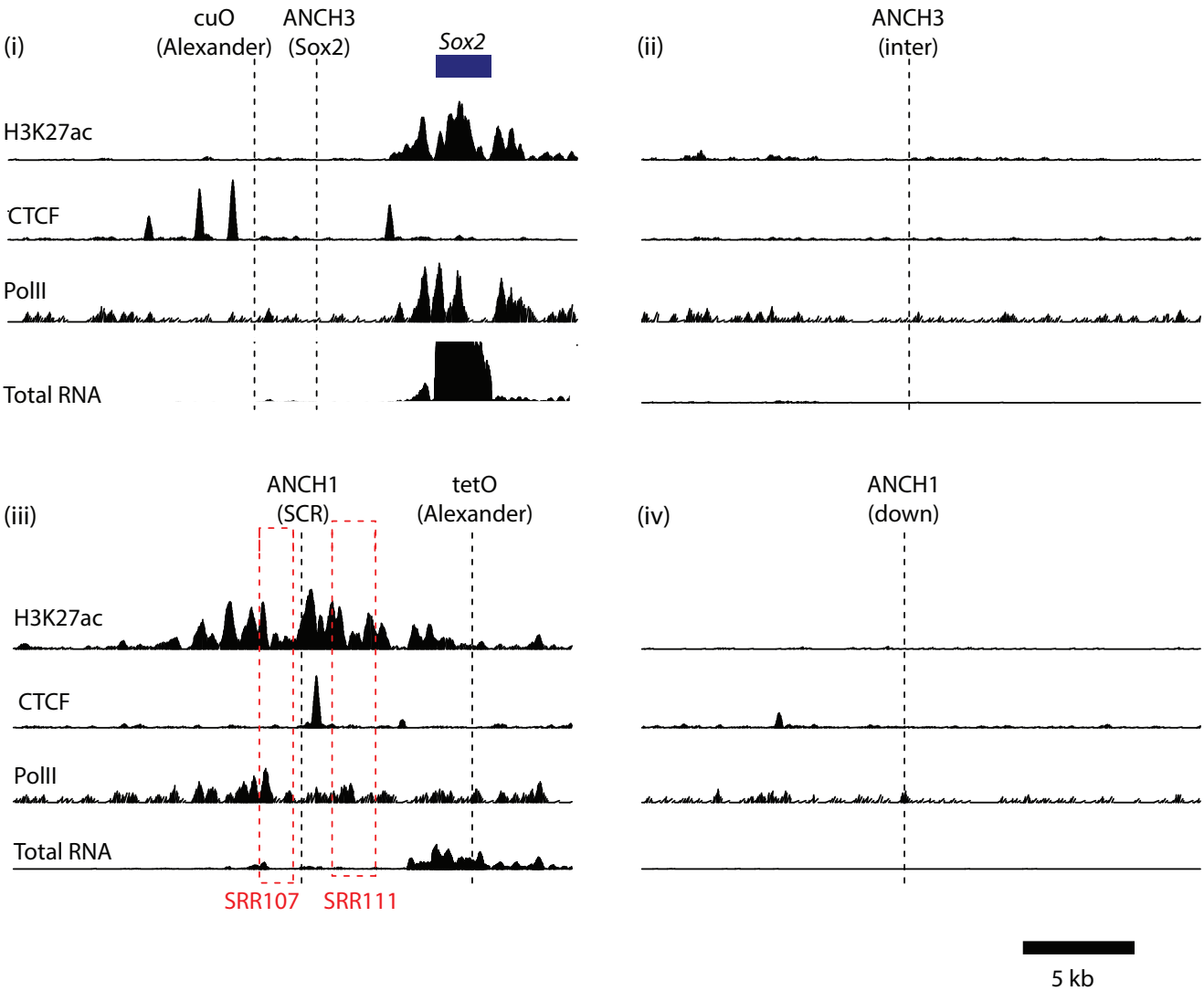
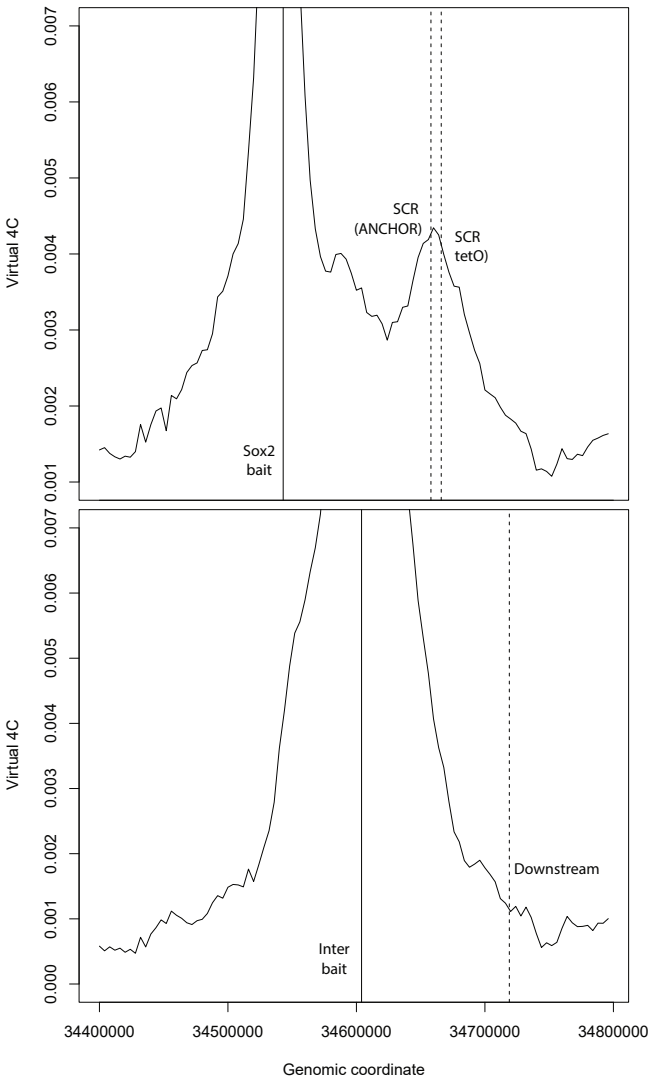


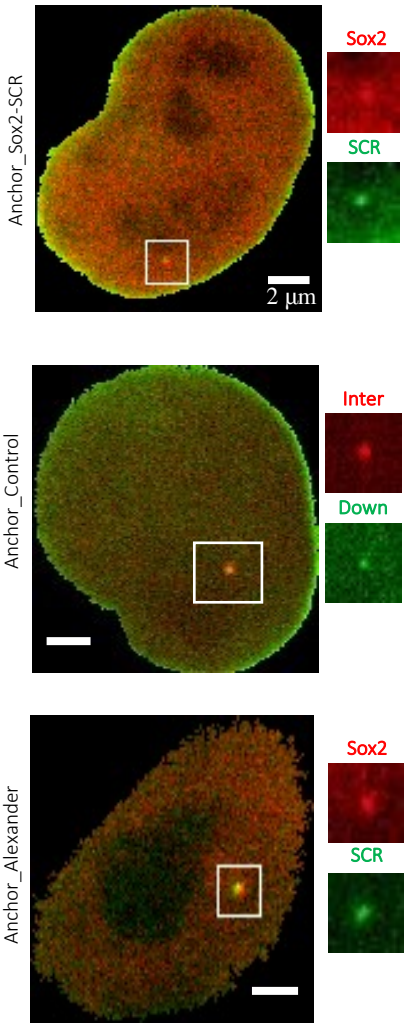
Figure S1: Detailed positions of ANCHOR labels and deletions. ChIP-seq tracks for H3K27ac, CTCF, RNA polymerase II and total RNA (GEO: GSE49847) are shown around the Sox2 locus alongside the positions of the different ANCHOR labels (black dashed lines) and *Sox2* gene (blue). (i-iv) denote zooms at 25 kb regions, encompassing (i) *Sox2*, (ii) the intervening control region, (iii) the SCR, and (iv) the downstream control region. Red dashed lines delimit the SRRs that are deleted on the *musculus* allele in Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs.

Supplementary Figure S2

a



b



c

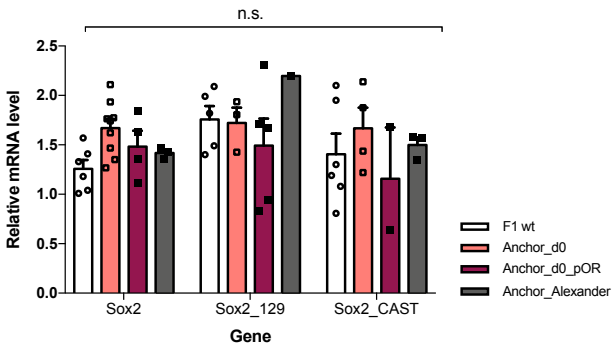


Figure S2: Validation of ANCHOR setup. a) Virtual 4C plots derived from ESC Hi-C data (Bonev et al., 2017), using the *Sox2* promoter (top) or intervening control region location (bottom) as bait (filled black line). The location of the "interacting" region where ANCHOR labels are incorporated are denoted with dashed black lines. **b**) Representative double-label images for Anchor_Sox2-SCR, Anchor_Control and Anchor_Alexander ESCs, showing unique ANCH3 (red) and ANCH1 (green) spots at proximal sites in the nucleus, indicating that both tags have been inserted in the same allele. Scale bar: 2 μm. **c**) Allele-specific qRT-PCR for Sox2 expression, normalized to SDHA housekeeping gene, for F1, Anchor_Sox2-SCR (with or without transfection of OR proteins) and Anchor_Control ESCs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Supplementary Figure S3

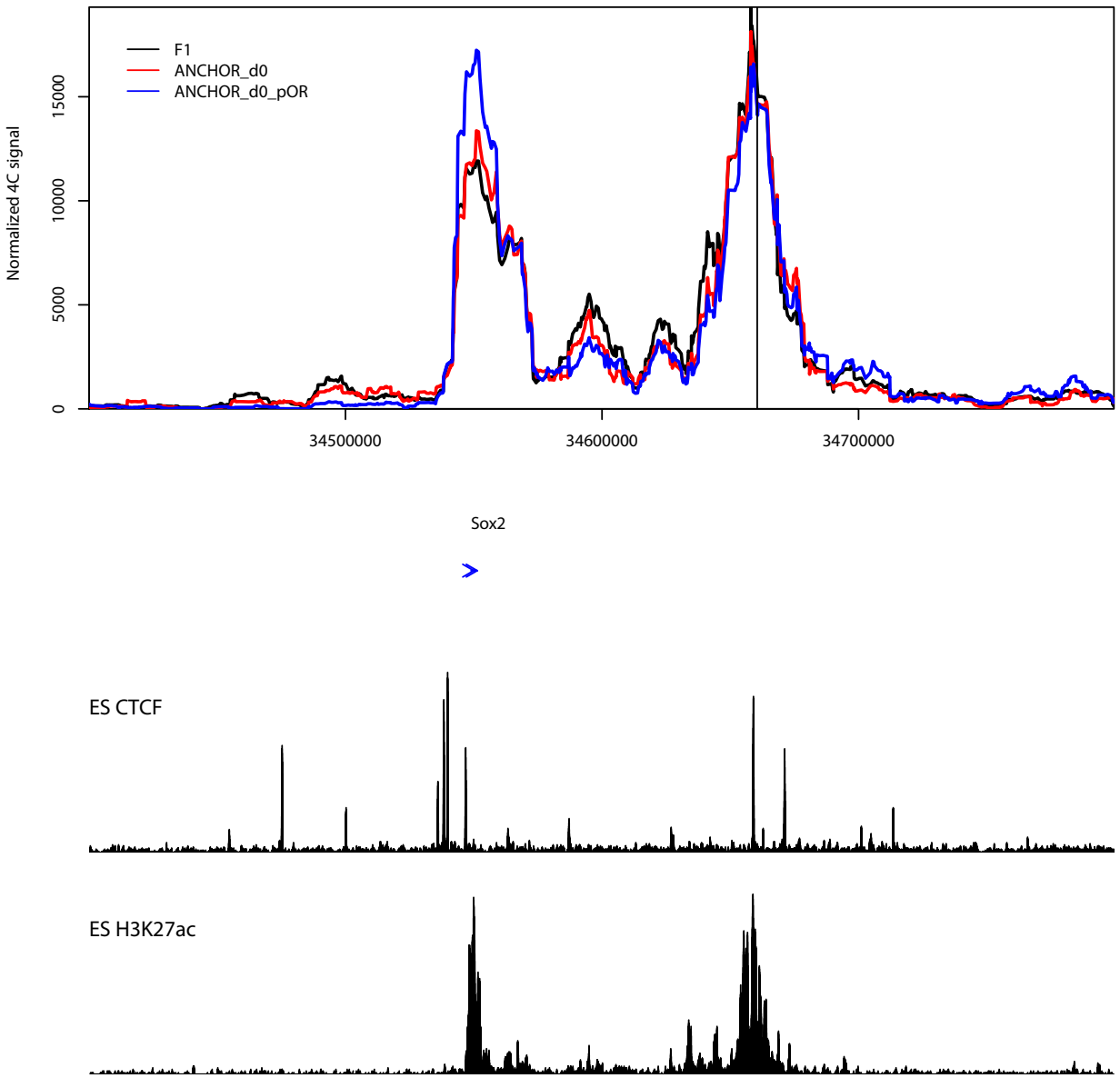
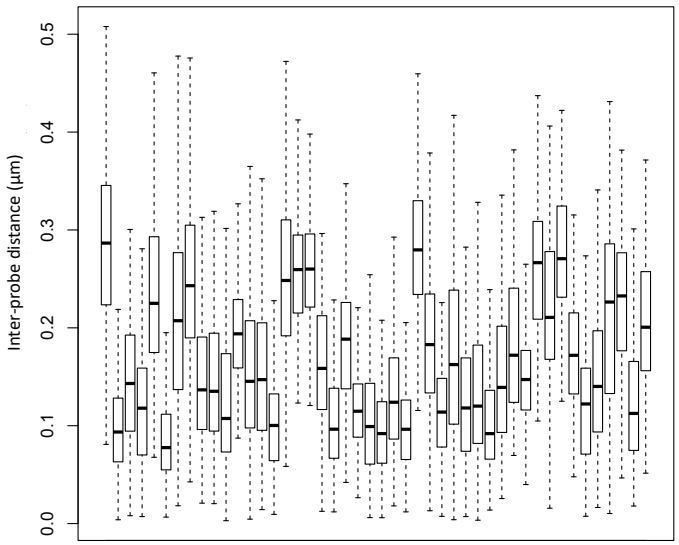


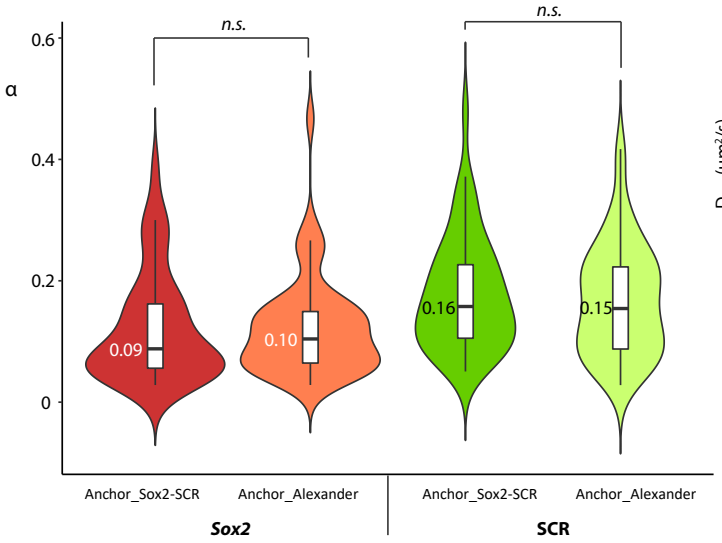
Figure S3: ANCHOR does not alter chromatin topology. Non-allele-specific 4C results, with the SCR as bait, for F1 and Anchor_Sox2-SCR (with or without transfection of OR proteins) ESCs. Position of Sox2 gene is shown underneath in blue, and ChIP-seq tracks for H3K27ac and CTCF (GEO: GSE49847) are also shown below.

Supplementary Figure S4

a



b



c

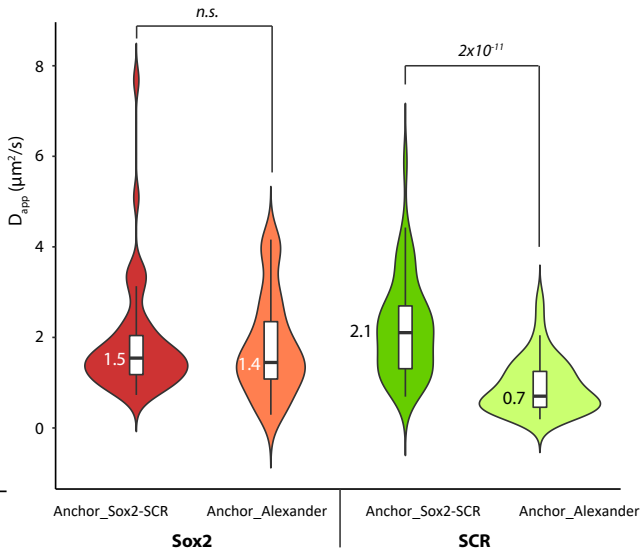
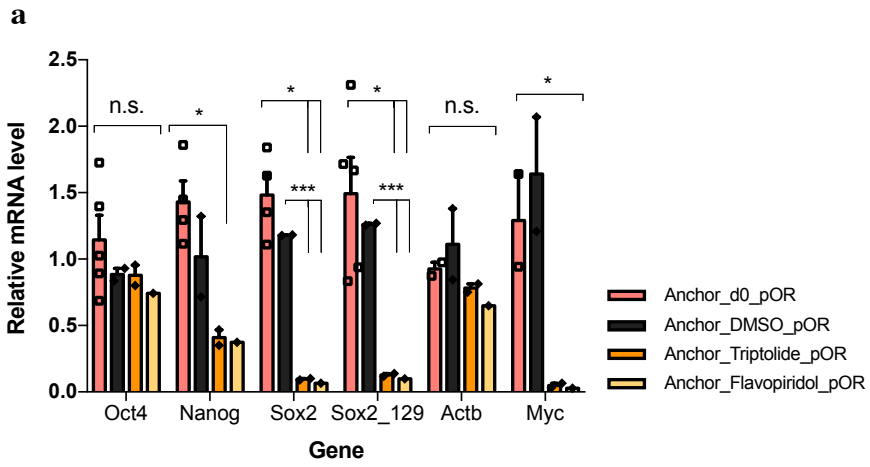
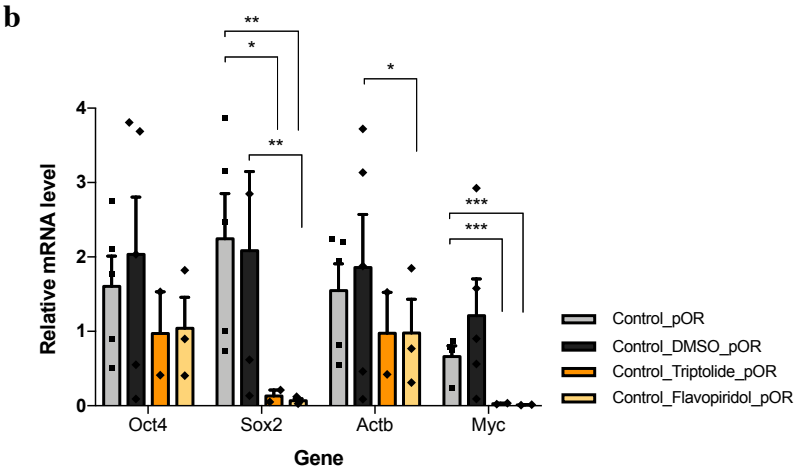
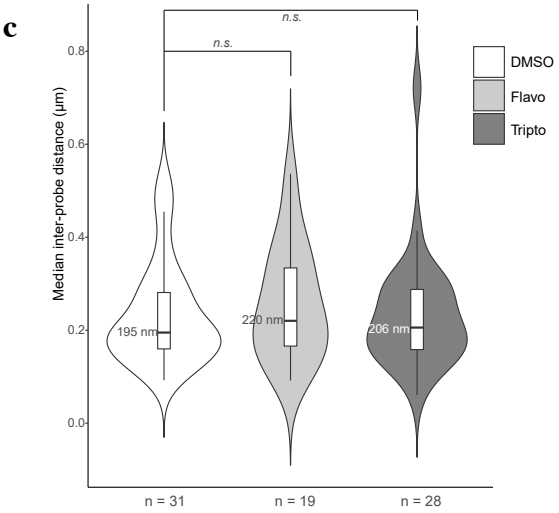


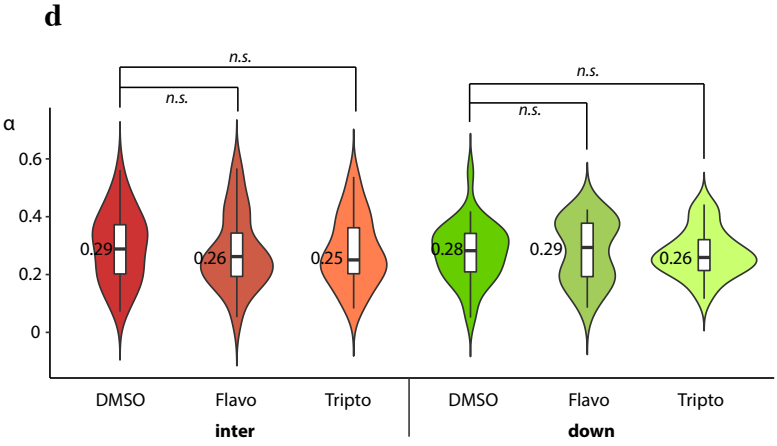
Figure S4: Comparison of alternative Sox2/SCR label locations. a) Boxplots for each individual Anchor_Sox2-SCR movie, showing distributions of inter-probe distances within each cell over the time of imaging. **b,c**) Violin plots for **b**) anomalous diffusion coefficient and **c**) apparent diffusion speed distributions of *Sox2* and SCR in Anchor_Sox2-SCR and Anchor_Alexander ESCs, exactly as for **Figs 23d,e**.

Supplementary Figure S5









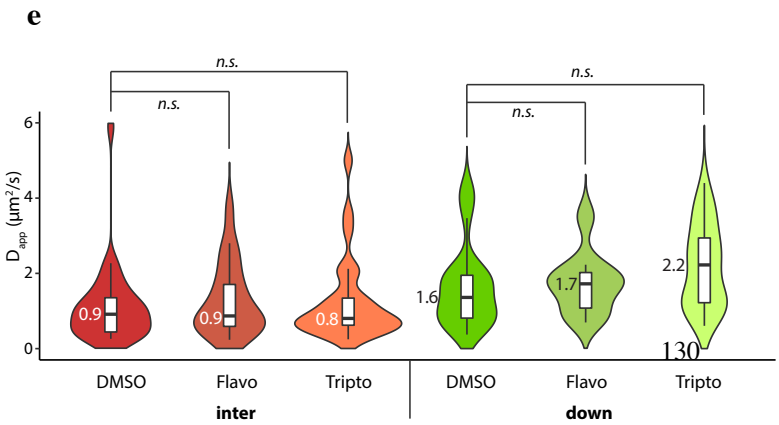


Figure S5: Transcriptional inhibition does not alter dynamics of control regions. a) qRT-PCR results, normalized to SDHA housekeeping gene, for Anchor_Sox2-SCR ESCs after treatment with different drugs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **b**) Exactly as for **a**, but in Anchor_Control cells. **c**) Violin plots for median inter-probe distance distributions, exactly as for **Fig 22b**, for Anchor_Control ESCs after treatment with different drugs. **d,e**) Violin plots for **d**) anomalous diffusion coefficient and **e**) apparent diffusion speed distributions of control loci in Anchor_Control ESCs after treatment with different drugs, exactly as for **Figs 23d,e**.

Supplementary Figure S6

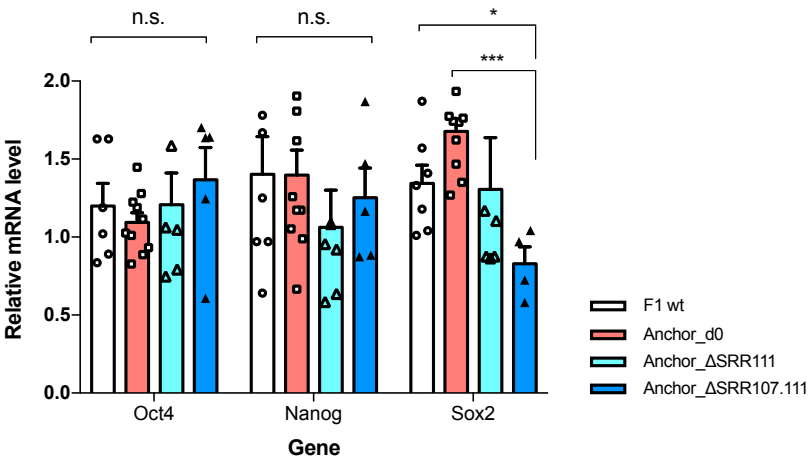


Figure S6: Heterozygous SCR mutations do not affect pluripotency. qRT-PCR results, normalized to SDHA housekeeping gene, for F1, Anchor_Sox2-SCR, Anchor_ΔSRR111 and Anchor_ΔSRR107,111 ESCs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Supplementary Figure S7

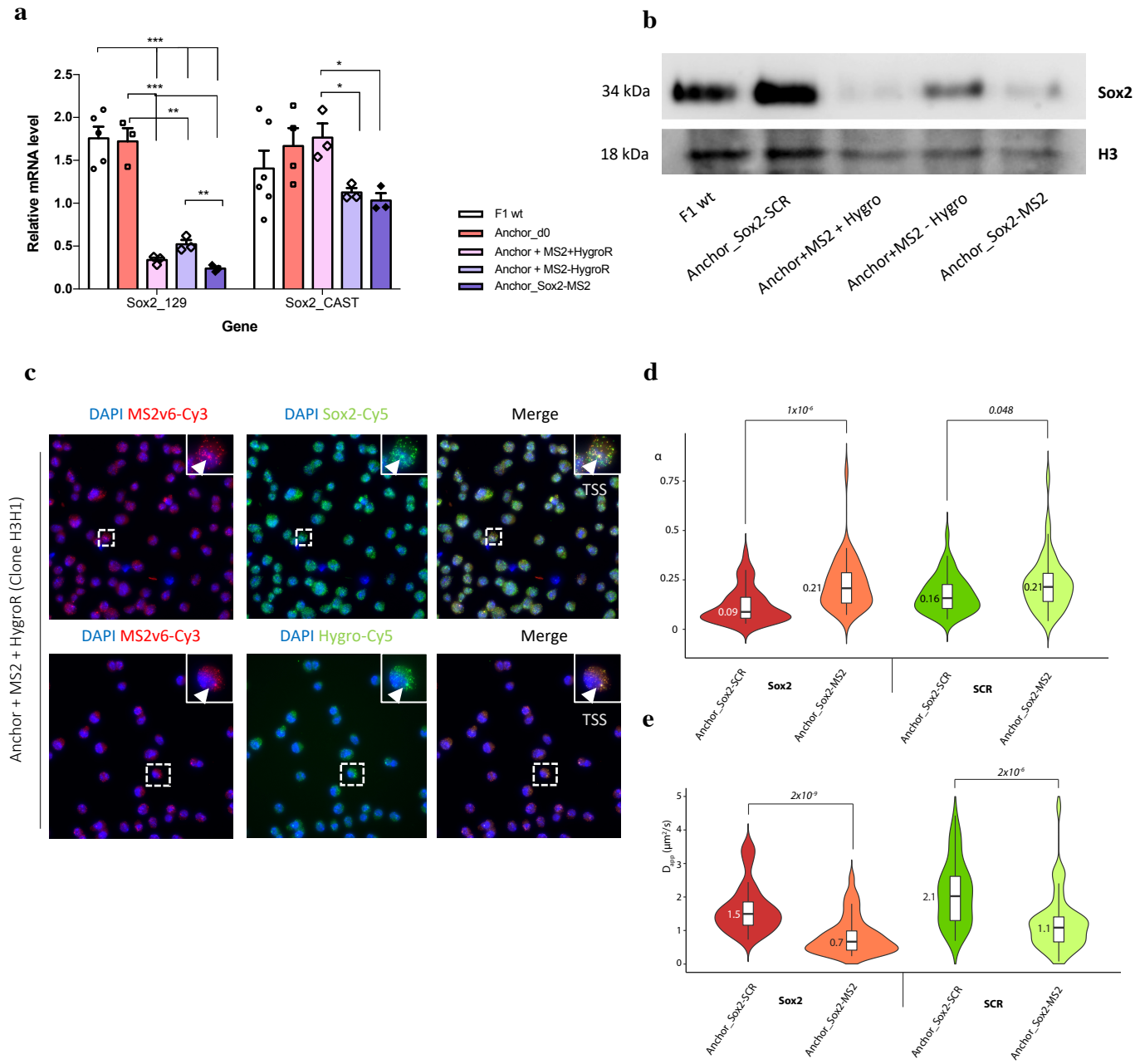


Figure S7: Transcriptional and post-transcriptional perturbation of Sox2 expression on incorporation of MS2 tag. a) Allele-specific qRT-PCR for Sox2 expression, normalized to SDHA housekeeping gene, for F1, Anchor_Sox2-SCR, and Anchor_Sox2-MS2 (before and after excision of resistance marker, and after stable integration of OR protein) ESCs. Error bars represent SEM. Significant differences in gene expression were determined by t-test: $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **b**) Western blot for Sox2 in F1, Anchor_Sox2-SCR, and Anchor_Sox2-MS2 (before and after excision of resistance marker, and after stable integration of OR protein) ESCs, using histone H3 as a loading control. **c**) Representative smRNA FISH maximum projection images within Anchor_Sox2-MS2 (before excision of selectable marker) ESCs for labeling of MS2 (red) with Sox2 (green); DAPI staining of nuclei in blue. **d,e**) Violin plots for **d**) anomalous diffusion coefficient and **e**) apparent diffusion speed distributions of control loci in Anchor_Sox2-SCR and Anchor-MS2 ESCs (without stratification according to bursting properties), exactly as for **Figs 23d,e**.

Table S1: Half-lives of mRNA species for which qRT-PCR was performed after treatment with transcriptional inhibitor drugs. Values taken from (Sharova et al., 2009).

mRNA of genes	Half life, h
Oct4	7.4
Nanog	5.2
Sox2	1.4
Actb	24
Myc	1.2

Appendix 1

Batch effects and variability within Anchor_Sox2-SCR line

The Anchor_Sox2-SCR line was the first to be generated, and has been used for experiments over a longer time period than other lines (Anchor_Control or Anchor-Alexander) or its derivatives (Anchor_ΔSRR111, Anchor_ΔSRR107,111 and Anchor_Sox2-MS2), undergoing more passages. When the inter-probe distance distributions of all of the accumulated double-label movies for Anchor_Sox2-SCR are processed together, the cells strikingly group into two populations, a population where the SCR and *Sox2* promoter are very close (median distance ~110 nm), and a group where they are noticeably farther apart (median distance ~225 nm). In the violin plots this creates a clear bimodal distribution, with not much density around the population median. Although it may have been interesting to speculate that these different distances could represent two functional states of the locus, such as transcriptionally active or competent versus silent/refractory, this pattern is not observed for any other cell line or experimental condition. These instead form a unimodal distribution, with density around the median, either approximately centralized around the median or with longer tails for larger separations (compare with Anchor_Control in **Fig A1**, or all other distributions in **Figs 22b, 23c, 24a, 25c, 27a and S5C**), suggesting a more technical issue with using the total Anchor_Sox2-SCR results. These differences are not apparently due to batch effects from performing the experiments on different days (**Fig A1**), nor do distances correlate with measurements of D_{app} or α, which themselves form unimodal distributions (**Fig A2**). The high variability and unusual distance distribution heavily reduces the statistical significance of any comparisons made between the "untreated" Anchor_Sox2-SCR and other cell lines or experimental conditions, even when average differences are apparent in the plots.

As the control condition for treatment with drugs inhibiting transcription, this same cell line is treated with DMSO solvent, which does not affect *Sox2* or other pluripotency gene expression (**Figs S5a,b**). Perhaps due to the experiments being performed in fewer batches over a shorter period of time, these results are much less variable and have a unimodal distance distribution (**Fig A3**). The diffusive parameters of the *Sox2* promoter, but not SCR or control loci, are significantly different between the total untreated cell dataset and DMSO-treated cells; in fact the DMSO-treated *Sox2* results are nearly identical to that obtained in the alternative

Sox2 probe site used in the Anchor_Alexander cell line. Although the reasons for the discrepancies in the untreated Anchor_Sox2-SCR results are unclear, there is sufficient evidence to suggest that the specific batches of results with DMSO treatment on the same cell line are more reliable representations of physiological Sox2 loci. The DMSO-treated Anchor_Sox2-SCR and DMSO-treated Anchor_Control datasets were thus used throughout the analyses in favor of the untreated ones. Further batches of experiments will be required to attain more robust measurements from untreated Anchor_Sox2-SCR cells, perhaps with stable integration of OR proteins instead of transient transfections, which may generate batch-to-batch variation.

Figure A1

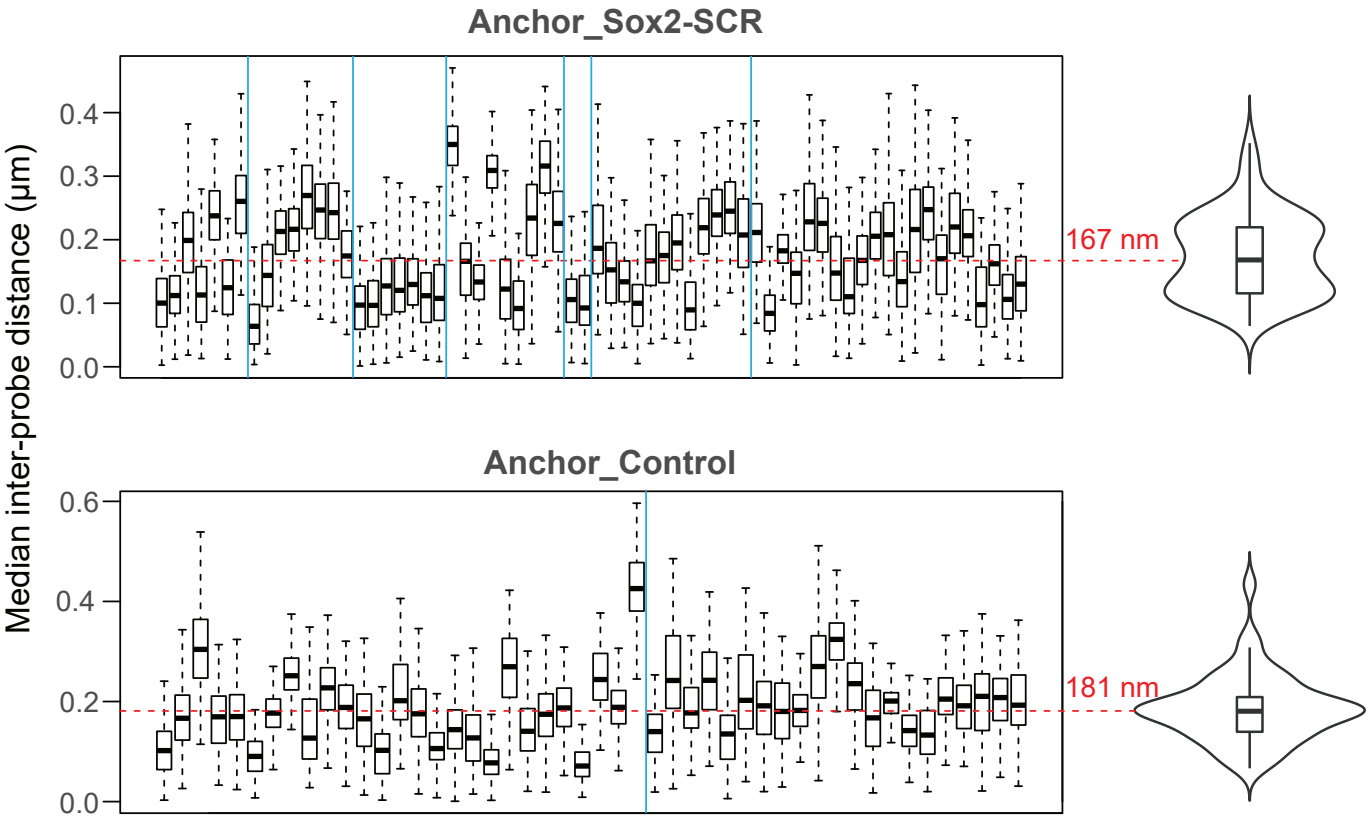


Figure A1: Two populations of promoter-enhancer distances in Anchor_Sox2-SCR ESCs. Boxplots for each individual Anchor_Sox2-SCR movie, and each individual Anchor_Control movie (untreated), showing distributions of inter-probe distances within each cell over the time of imaging. Total population median is shown by dashed red line, and violin plot for overall distribution is shown to the right. Individual movie boxplots are displayed in chronological order of their acquisition; blue lines delimit series of movies acquired on the same day.

Figure A2

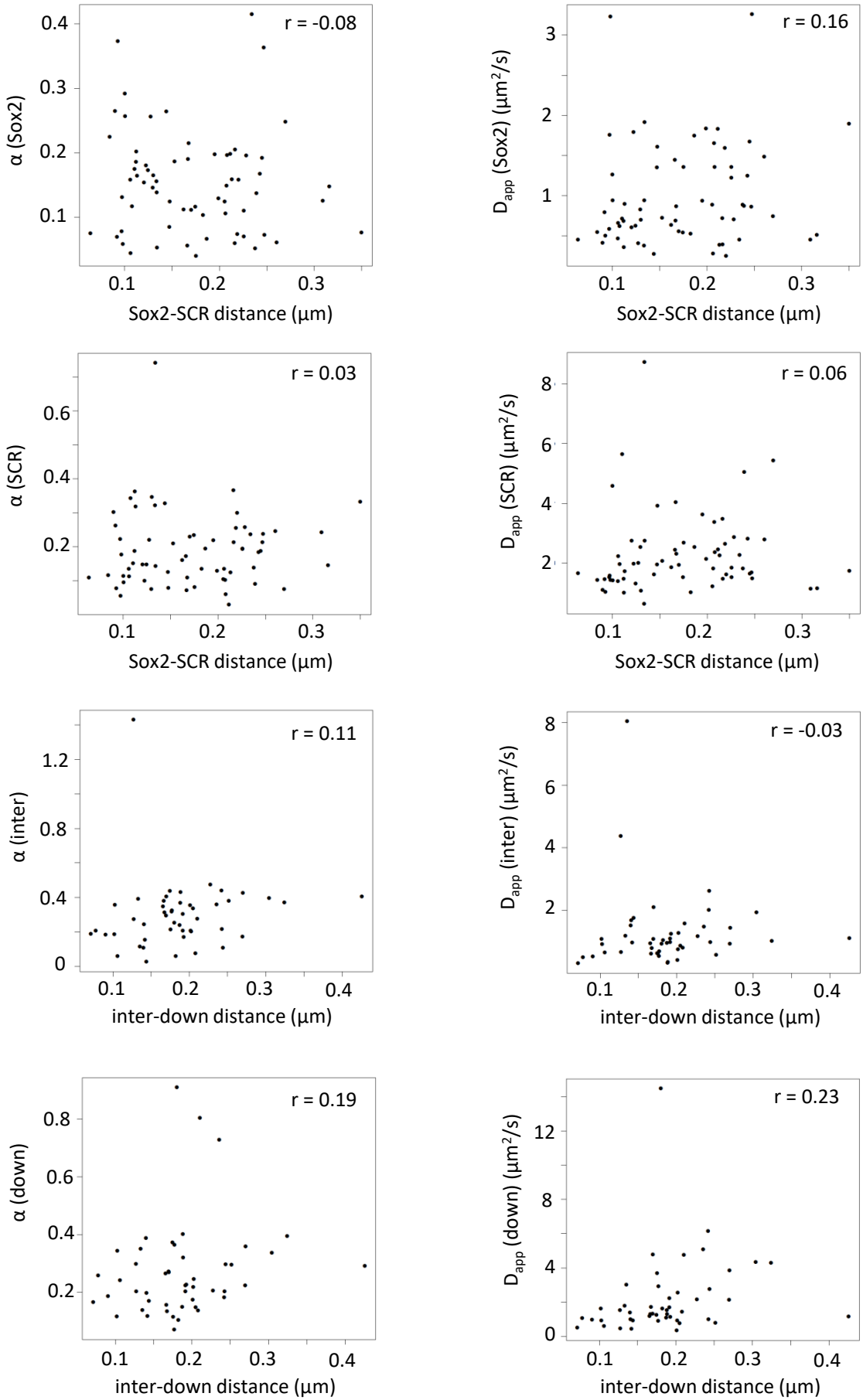
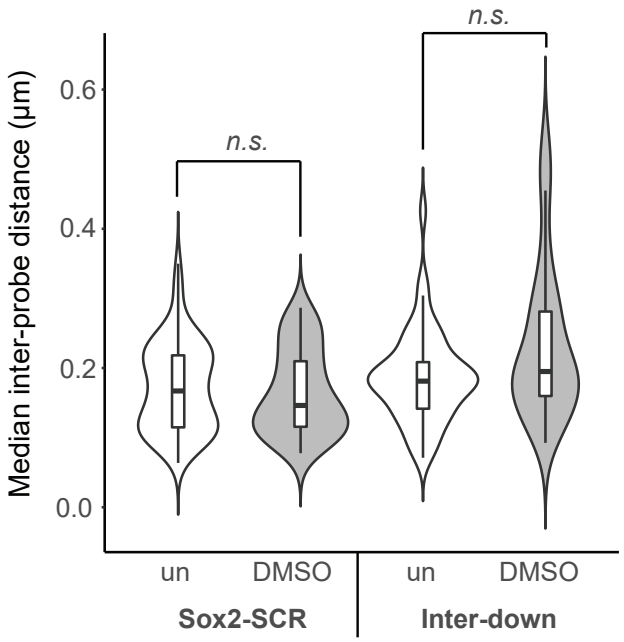
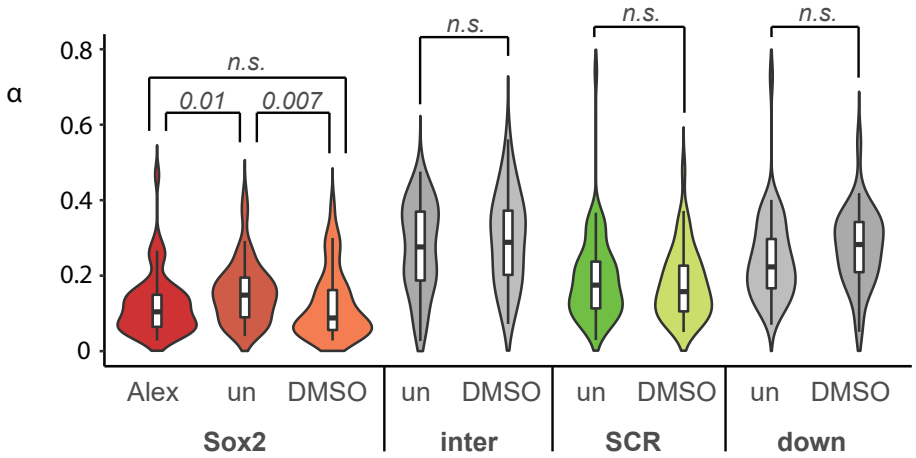

r = -0.08
α (Sox2)
Sox2-SCR distance (μm)
r = 0.16
D_{app} (Sox2) (μm²/s)
Sox2-SCR distance (μm)
r = 0.03
α (SCR)
Sox2-SCR distance (μm)
r = 0.06
D_{app} (SCR) (μm²/s)
Sox2-SCR distance (μm)
r = 0.11
α (inter)
inter-down distance (μm)
r = -0.03
D_{app} (inter) (μm²/s)
inter-down distance (μm)
r = 0.19
α (down)
inter-down distance (μm)
r = 0.23
D_{app} (down) (μm²/s)
inter-down distance (μm)

Figure A2: Proximity does not correlate with local diffusive parameters in individual movies. Scatterplots showing lack of correlation between inter-probe distance and either anomalous diffusion coefficient or apparent diffusive speed, at either *Sox2* or SCR in Anchor_Sox2_SCR ESCs, or at either control region in Anchor_Control ESCs.

Figure A3


n.s.
n.s.
Median inter-probe distance (µm)
0.6
0.4
0.2
0.0
un
DMSO
un
DMSO
Sox2-SCR
Inter-down


α
n.s.
n.s.
n.s.
n.s.
0.01
0.007
0.8
0.6
0.4
0.2
0
Alex
un
DMSO
un
DMSO
un
DMSO
un
DMSO
Sox2
inter
SCR
down

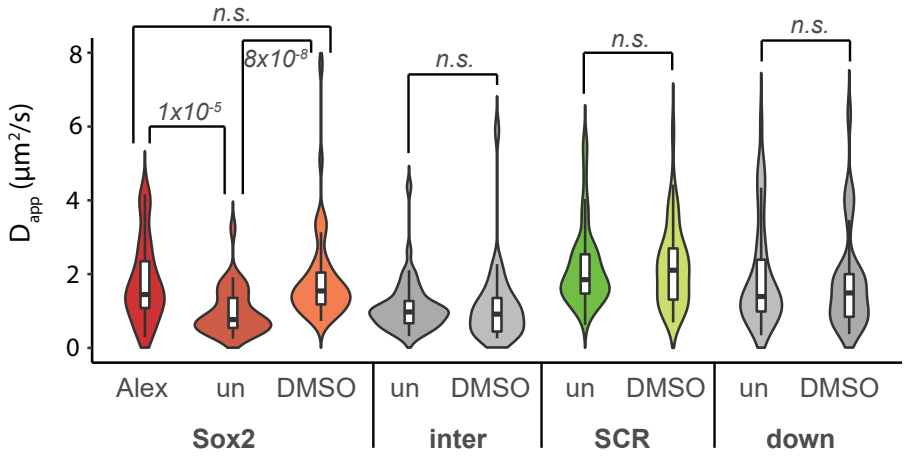

n.s.
n.s.
n.s.
n.s.
1x10-5
8x10-8
D app (µm2/s)
8
6
4
2
0
Alex
un
DMSO
un
DMSO
un
DMSO
un
DMSO
Sox2
inter
SCR
down

Figure A3: Comparison of DMSO-treated and untreated Anchor_Sox2-SCR proximities and local dynamics. Violin plots for median inter-probe distance (top), anomalous diffusion coefficient of *Sox2*, SCR and control regions (middle), and apparent diffusive speed of *Sox2,* SCR and control regions (bottom), in untreated or DMSO (drug control)-treated Anchor_Sox2-SCR or Anchor_Control ESCs. The diffusive properties of the *Sox2* promoter are also shown for the Sox2_Alexander line for comparison.

Discussion

In this study, with a combination of genome editing and multi-color live-cell imaging, I was able to extract some very interesting features about the separation distances and the local chromatin dynamics of regulatory elements, an aspect that has been overlooked in most previous studies of genome functions. Overall, I found that the *Sox2* gene and its cognate ES enhancer are frequently in close spatial proximity, but that the exact distance does not necessarily correlate with transcriptional firing. Instead, transcriptional activity is generally associated with alterations in local chromatin dynamics, whereby regulatory regions are generally more constrained than "neutral" sequences, and apparent diffusive speed is increased at promoters when genes are activated. Alterations in chromatin dynamics correlate with gene activity (summarized in **Table 2**), but loci participating in active transcriptional bursts do not have notably different chromatin dynamics to poised loci.

Table 2: Recapitulative table of measured *Sox2*-SCR distances and diffusive parameters upon perturbations of *Sox2* expression.

	***Sox2* expression**	***Sox2*-SCR distance**	**α for *Sox2***	**α for SCR**	**D for *Sox2***	**D for SCR**
Differentiation (day 2)	Reduced	Unchanged	Increased	Unchanged	Reduced	Unchanged
Differentiation (day 3)	Almost completely lost	Increased	Increased	Increased	Reduced	Reduced
Triptolide	Almost completely lost	Unchanged	Weak, non-significant increase	Unchanged	Reduced	Unchanged
Flavopiridol	Almost completely lost	Increased	Unchanged	Unchanged	Unchanged	Unchanged
ΔSRR111	Reduced (~1.4-fold)	Unchanged	Unchanged	Unchanged	Reduced	Unchanged
ΔSRR107.111	Reduced (>4-fold)	Unchanged	Increased	Unchanged	Reduced	Unchanged

Revisiting enhancer looping models

As has been discussed earlier (see Introduction), a large body of research suggests that enhancer-promoter looping is required for gene expression (Bartman et al., 2016; H. Chen et al., 2018b; Deng et al., 2012, 2014; Palstra et al., 2003). However, recent imaging experiments demonstrated, at least for some specific loci, enhancers at large distances from activated target genes (Alexander et al., 2019; Benabdallah et al., 2019; Lim et al., 2018), thus decoupling enhancer-promoter proximity from gene expression. In this study, I show that *Sox2*-SCR distances are significantly closer (median=146 nm) than control regions of equivalent genomic separation (median=195 nm), in agreement with Hi-C data (Bonev et al., 2017). Nonetheless, both *Sox2*-SCR and control pairs were found in close proximity, as might be expected for genomic separations of ~115 kb. Recent high-throughput imaging on fixed cells revealed a remarkable degree of cell-to-cell heterogeneity and variability in spatial genome organization (Bintu et al., 2018; Cattoni et al., 2017; Finn et al., 2019; Giorgetti et al., 2014). Although my current experimental setup does not allow simultaneous visualization of all four regions, it is highly likely that there would be individual cells where the control regions are closer than *Sox2*-SCR. Future work is needed to solidify this hypothesis. A very ambitious experiment is to perform triple-label ANCHOR in order to determine whether, and how often, *Sox2*-SCR are really closer than *Sox2*-Inter control and SCR-Inter control pairs.

Besides characterizing chromatin topology at the pluripotent state, where *Sox2* is expressed, this study also revealed insights into enhancer-promoter separation distances at early stages of ESC *in vitro* differentiation, where *Sox2* is silenced. In agreement with the 4C results, *Sox2* and SCR average distances were found significantly increased (median=205 nm) three days after retinoic acid (RA)-induced differentiation of mESCs, yet the probes were still in relatively close spatial proximity, and much closer compared to the average distances separating *Sox2* and SCR in mESCs proposed by a previous study (Alexander et al., 2019). To directly examine if and how chromatin topology relates with gene activation, I tagged the native *Sox2* mRNA of the *musculus* allele of ANCHOR-labelled mESCs with MS2 stem loops (Bertrand et al., 1998), providing a direct readout of transcription from the endogenous *Sox2* locus. Triple-label live imaging allowed *Sox2* and SCR proximity to be simultaneously measured with gene transcription. Although imaging of a greater number of cells is required to lend greater support to these preliminary results, I find here that *Sox2* and SCR are slightly

closer during active transcription and that the actively transcribing subset of *Sox2* loci has significantly closer Sox2-SCR distances compared to the total distribution from the double-label experiments. Together, these findings fit with a model in which physical enhancer-promoter proximity correlates somehow with transcriptional activation. Nevertheless, when I assessed the link between *Sox2* transcription and local chromatin architecture by other means, in particular after inhibition of transcriptional initiation upon treatment with triptolide, which represses *Sox2* expression to almost undetectable levels, and after deletions of critical sites within the SCR, which strongly reduced Sox2 expression (>4-fold), I did not observe any effect on *Sox2*-SCR proximity. Overall, these results suggest that the *Sox2* gene and its cognate ESC enhancer, SCR, are frequently in close proximity, but that the exact separation distance does not necessarily correlate with transcription. In light of these and others' results (Benabdallah et al., 2019; Lim et al., 2018), the simplistic looping model needs to be revised. Direct physical juxtaposition of promoters and enhancers does not seem to be required (nor sufficient) for transcriptional regulation to be conferred; instead, at least for *Sox2*, crosstalk across the elements appears to be possible when they are within a certain, yet-to-be-determined, radius.

Sox2 chromatin dynamics are consistent with transcriptional hubs and/or condensates

Recent research has pointed to the existence of transcriptional hubs that create local microenvironments, wherein enhancers and promoters can share common clusters of Pol II and regulatory factors (Lim & Levine, 2021). Moreover, recent imaging of regulatory factors and *cis*-elements at pluripotency genes, including *Sox2*, documented ~100- to 200-nm-sized clusters of enhancer-associated TFs, created by frequent proximity of distal enhancers to target genes (J. Li et al., 2019; Jieru Li et al., 2020). Notably, all the separation distances observed in this study, including the ones between the control probes, are within the ~100- to 200-nm range described by Li *et al*. (J. Li et al., 2019; Jieru Li et al., 2020). Taking into consideration my findings on enhancer-promoter separation distance and chromatin diffusion parameters, it is tempting to speculate that such a transcriptional hub could be embracing *Sox2* and SCR, thus constraining their mobility relative to control regions. Since the intervening sequence is also less constrained, it may loop out of the transcriptional hub and thus be freer to explore the nucleoplasm. My results are thus globally consistent with a model whereby promoters and

enhancers reside, without necessarily directly juxtaposing, in the same transcriptional hub, which appears to constrain local chromatin movement. However, my results do not support a simple model whereby hubs exclusively contain actively transcribing loci. Chromatin constraints are largely unchanged on treatment with drugs inhibiting transcription and, more tellingly, there are no obvious differences in the dynamics of transcribing or non-transcribing alleles within wild-type ESCs. The links, if any, between transcriptional hubs and apparent diffusive speeds are even more obscure.

The challenge now is to identify whether *Sox2* and SCR co-localize within the proposed transcriptional hub. To understand the dynamics and spatial organization of the effector proteins within the suggested hub, I will attempt to use super-resolution microscopy to visualize and track RNA Pol II and key factors that have been previously found in clusters, such as Mediator, Sox2 and Brd4 (Cho et al., 2018; Jieru Li et al., 2020; Sabari et al., 2018). A very ambitious goal, whose success is uncertain but would be extremely rewarding, is to couple ANCHOR labelling of *Sox2* with single-particle tracking of labeled transcription factors and directly observe the major assumption of a transcriptional hub model: that TF exchange is more rapid when a gene and an enhancer are in the same microenvironment. The three- and four-fluorophore experiments will be very challenging, and even more difficult to analyze. I anticipate that this technical difficulty will be facilitated by future technical advances. However, technical issues aside, it is also likely that the much faster dynamics of freely diffusing transcription factors cannot be meaningfully modeled with the much slower changes to the bulky chromatin fiber. Future work should also explore whether regulatory factors and Pol II are recruited into clusters to form transcription hubs via the formation of phase-separated condensates (Sabari et al., 2018) or through binding to cognate DNA and chromatin binding sites (Jieru Li et al., 2020). Treatment of cells with 1,6-hexanediol, which disrupts condensates (Yi Lin et al., 2016), followed by live-imaging experiments in order to measure separation distances between *Sox2* and the SCR and calculate local chromatin dynamics might be informative, although indirect effects of this solvent change could confound interpretations. Nevertheless, 1,6-hexanediol has been very recently shown to weaken enhancer-promoter interactions and TAD insulation (Ulianov et al., 2021), thus alternative experiments might be needed to explore this interesting avenue of research. Ultimately, even if for technical reasons this experiment might not be feasible to perform, combination of triple-label ANCHOR for *Sox2*, internal control probe and SCR with RNA Pol II and/or TF immunolabeling, would

reveal the dynamics and spatial organization of the tagged loci within the proposed transcriptional hub. Applying a live antibody-based imaging approach, such as VANIMA (Conic et al., 2018), and super-resolution microscopy may be required to obtain meaningful resolution of transcriptional hubs.

PolII loading and a more indirect effect of transcription on chromatin dynamics?

At first glance, my results appear to contain a large contradiction. The double-label results after various means of perturbing *Sox2* expression (differentiation, treatment with transcriptional inhibitors, deletion of enhancer regions) all point to a direct link between transcription and chromatin dynamics, with perturbed expression causing a generally reduced constraint at both enhancer and promoter, and a more specific reduction of apparent diffusive speed at the promoter. However, when simultaneously labeling nascent *Sox2* transcription, there are no apparent differences in average diffusive properties between transcribing and non-transcribing alleles, either at the promoter or the enhancer. One possible explanation may be that in wild-type ESCs, nearly all alleles of the highly-expressed Sox2 locus are "poised" for transcriptional bursting within the same permissive nuclear microenvironment, regardless of whether or not they are synthesizing mRNA at the moment of image acquisition. Chromatin dynamics are likely to be determined by the local environment, so may thus not be expected to differ. In support for this, the *Sox2* promoter/gene start contains high levels of paused Pol II in ESCs, a known characteristic of poised, rapidly-inducible genes (Williams et al., 2015). Perturbations which convert poised Sox2 loci to a more refractory state may do so by altering the nuclear microenvironment, with subsequent effects on chromatin dynamics. For example, differentiation causes bulk changes in Pol II loading at the promoter, as well as loss of TF binding at both promoter and the SCR; these changes may directly reduce local chromatin mobility constraints, or indirectly via dissociation from transcriptional hubs. Triple-label experiments performed over much longer time periods may also uncover if any wild-type loci transition between the proposed "refractory" and "poised" states, determined via longer-term alterations in local diffusive parameters. Interestingly, chromatin dynamics were largely unaltered on treatment with the transcription elongation inhibitor, flavopiridol, but promoter dynamics were sensitive to treatment with the transcription initiation inhibitor, triptolide, suggesting that Pol II recruitment and/or promoter melting, but not processive elongation, are

the determinant steps for local chromatin mobility. ChIP-seq experiments show that much less Pol II is bound to the SCR than the *Sox2* promoter, potentially explaining why dynamics of the promoter are more sensitive to perturbation experiments. Support for this hypothesis could be provided by more in-depth study of binding of total and engaged Pol II, with ChIP-seq and GRO-seq (gene run on-sequencing; Williams et al., 2015) studies, respectively, on the different perturbations of *Sox2* expression.

Comparison to previous findings

During my thesis another laboratory published its work, in which they interrogated the spatial organization and activity of *Sox2* and SCR in living F1 mESCs (Alexander et al., 2019). By utilizing cuO and tetO arrays to label the promoter and enhancer regions at the 129 allele, the authors demonstrated that *Sox2* and SCR are separated by much greater distances (mean ~340 nm), compared to the ones reported in my study, in the nucleus of mESCs. I directly assessed this discrepancy by introducing ANCHOR into the exact same genomic locations of this study, finding that these probes were also frequently proximal. I propose therefore that the large separation distances reported by the authors (Alexander et al., 2019) are due to technical differences, possibly caused by the large repetitive tet/cu operators. Notably, the authors included a control line, where a 111 kb fragment has been deleted between the cuO and tetO pairs, leaving a 14 kb tether between them (Alexander et al., 2019). Curiously, the mean separation distance observed was ~250 nm and never below 100 nm, as has been previously reported for DNA FISH probe pairs separated by few tens (~30-60) of kilobases (Giorgetti et al., 2014). I assume that the large separation distances reported by Alexander *et al*. (Alexander et al., 2019) are caused by the integration of multiple copies of repetitive operator sequences and the stable binding of the fluorescent-tagged repressors.

Furthermore, in this study, by using a *Sox2*-MS2 system, we measured much higher bursting frequencies compared to what was previously reported (Alexander et al., 2019). These differences might be caused by the resistance selection marker that was maintained in the locus in the other study (excised from my Anchor_*Sox2*-MS2 cells), and could further perturb *Sox2* transcription and mRNA stability. Indeed, I observed that the line maintaining the selectable marker gene, exhibited reduced nascent mRNA and protein levels, as measured by smFISH and western blot. These results confer an additional value to my ANCHOR and MS2-MCP

labelling systems. However despite the large discrepancy in distance measurements, both studies put the classical enhancer-promoter looping model into question.

Whereas I observed greater separation of *Sox2* and SCR during very early stages of *in vitro* ESC differentiation, Alexander *et al*. reported greater proximity between *Sox2*-SCR after full differentiation to neuronal precursors. Additionally, increased separation distances were observed between *Shh* and *Shh* brain enhancers (SBEs) during neuronal differentiation, where *Shh* is active (Benabdallah et al., 2019). While the latter could be attributed to the different genomic context, *Sox2*-SCR opposing results could arise from the different stages of differentiation monitored in each study. Alexander *et al*. proposed that the entire *Sox2* region adopts a more compact conformation upon ESC differentiation. Longer differentiation experiments of my Anchor_Sox2-SCR mESCs are required to confirm whether the *Sox2* locus is configured into a compact structure, and if and how this affects promoter-enhancer proximity and local dynamics.

Local diffusive parameters as "measurements" of regulatory activity?

To my knowledge, this work is the first live-imaging study that describes separation distances and the diffusion properties of an enhancer-promoter pair, simultaneously with gene transcription, and in response to inhibitors of transcription. Previous studies also addressed the link between chromatin dynamics and gene expression (Germier et al., 2017; Gu et al., 2018), attaining seemingly conflicting results. My results both support and contradict these previous results to some extent. Firstly, as described above for my own results, the anomalous diffusion parameter (α) appears linked to "activity", with greater constraint at active regions, in line with a recent report, where transcription was shown to coincide with confinement of a mRNA-producing gene to smaller nuclear volumes (Germier et al., 2017).

Specifically, I also found that the *Sox2* promoter, but not the SCR, diffuses more rapidly (higher D_{app}) on activation. This finding partially agrees with one recent report: both a gene and an enhancer activated upon differentiation were found to move faster (Gu et al., 2018), interpreted in this study as an increase in thermal energy via the active transcription process. An alternative explanation (see also above) is that the first steps of the transcriptional cycle, and in particular chromatin remodeling and nucleosome removal as PIC is recruited and/or the promoter melting after the assembly of the PIC, actually result in less dense chromatin which

eventually can move quicker. Reduced mobility of *Sox2*, as shown in this study upon treatment with triptolide, a known inhibitor of the helicase activity of TFIIH (Bensaude, 2011), is consistent with this hypothesis. On the other hand, opposite to Gu *et al*., I did not observe any link between the diffusive speed of the enhancer and gene activation. Publicly available RNA-seq data from mESCs (ENCODE ENCSR000CWC; (Dunham et al., 2012)) suggest that the SCR is not producing any eRNA. Although we have no access to analogous data from EpiLCs, where the assessed *Fgf5* enhancer is active (Gu et al., 2018), its transcriptional activity could explain why it behaves like the active *Fgf5* and *Sox2* promoters, and not the SCR. Moreover, contrary to Gu *et al*., I concluded that inhibition of transcriptional elongation has no effect on chromatin mobility of either regulatory element. It remains to be seen whether this can be explained by technical differences or gene-specific contexts.

Despite the clear "slowing" of promoter diffusion on gene inactivation, it should be noted that the apparent diffusive coefficient of the *Sox2* promoter is significantly slower than both the SCR and control regions, which are not (or very poorly) transcribed. Thus when comparing genomic loci, transcriptional activity alone is insufficient to "predict" local diffusion speed. Interestingly, another member of the lab, Guilherme Oliveira, has recently built a polymer model introducing long-range interaction forces based on available Hi-C maps (Bonev et al., 2017), and has used it to accurately predict locus-specific D_{app} and α values around the HoxA locus, finding good agreement with the experimental results (Oliveira et al., 2021). Chromatin architecture therefore also appears to be an important feature in determining local chromatin dynamics. This model failed to predict the experimentally-derived values at the *Sox2* locus; a more complex polymer model to also include gene activity may also be required to further comprehend these findings.

Utility of ANCHOR for probing chromatin dynamics

In this study, I adapted the minimally perturbative ANCHOR/ParB DNA labelling approach, originally developed in yeast (Saad et al., 2014), and optimized the system for *musculus* allele-specific CRISPR/Cas9 knock-in of double labels. Besides the laborious engineering of the double knock-in cell lines, allele-specificity wouldn't be so feasible with dCas9 approaches, due to the lack of enough SNPs around the regulatory elements. The small size (<1 kb) of the

ANCH sequence together with the non-specific and relatively weak DNA binding of OR proteins, allowed me to position the ANCH3 sequence (Germier et al., 2017) within the immediate vicinity of *Sox2* (6.5 kb centromeric to TSS) and the ANCH1 (Saad et al., 2014) inside the SCR (~109 kb telomeric to TSS), in contrast with a previous study that had to integrate the large cuO repeats downstream of the SCR (Alexander et al., 2019). Importantly, in order to maintain the native context of the endogenous genomic locus, I did not include any ectopic insulator sequences as has been used in a recent study examining enhancer-promoter associations in Drosophila (H. Chen et al., 2018a). As has been previously reported in yeast (Saad et al., 2014), the ANCHOR/ParB system does not interfere with nucleosome formation, nor create fragile sites. Similarly (and beyond), I demonstrated that insertion of ANCH sequences and their binding with OR proteins show no change in overall *Sox2* expression, neither affect markers of pluripotency despite genetic modification of the locus. Furthermore, by performing allele-specific and canonical 4C from the SCR bait I showed that *Sox2*-SCR interactions are maintained in wild-type (F1) and ANCHOR (Anchor_Sox2-SCR) cells, regardless of whether OR proteins are expressed. Collectively, these results highlight the non-invasive nature of ANCHOR compared to large and repetitive lac or tet operators, which can silence surrounding chromatin and create fragile sites (M. Dubarry et al., 2011; Jacome & Fernandez-Capetillo, 2011) or dCas9 binding which can repress transcription of a target gene by interfering with RNA Pol II binding, transcription elongation, or TF binding (Qi et al., 2013). As it has been discussed previously, the correct insertion of the ANCH sequences for each locus in the *musculus* (129) allele of F1 ESCs was verified using PCR with several pairs of primers that span the unique homology arms. Nevertheless, it would be interesting to combine ANCHOR with DNA FISH, using probes against *Sox2* and SCR, or ImmunoFISH, for simultaneous visualization of OR proteins bound to DNA, as another means to show how accurately the ANCH sequences represent *Sox2* and the SCR.

Still, although the ANCHOR approach was shown to be effective and non-invasive, it is currently limited by concerns over the relatively low brightness of the fluorophores labeling OR proteins. In this study, the duration of the movies has been adjusted (~2-3 min) to avoid photobleaching of the fluorophores and cell damage, and images have been restricted to 2D planes by confocal microscopy. To address this problem, I tried the HaloTag labeling technology (Los et al., 2008), that allows specific fluorescent labeling of fusion (Halo-OR) proteins, but unfortunately without obtaining a much higher fluorescent signal. In the future, it

will be critical to improve these technical limitations by using brighter and more versatile Halo-compatible tags and by decreasing photobleaching and phototoxicity, possibly by applying lattice light-sheet microscopy (B. C. Chen et al., 2014).

Perspectives

Although the *Sox2* locus has proven a powerful tool to study enhancer control, not least because the SCR is one of the few enhancers described to be absolutely essential for target gene expression in ESCs (Zhou et al., 2014), one experimental drawback is that the gene is already active in ESCs, and one can only follow the loss of expression during perturbation experiments or on differentiation. Therefore, I would like to perform the same ANCHOR setup for a promoter-enhancer pair that is silent in ESCs but activated on differentiation to neuronal precursors (NPC). Labelling of *Sox2* and the distal NPC-specific *Sox2* enhancer, recently identified in 5C and Hi-C maps (Bonev et al., 2017; Phillips-Cremins et al., 2013), could be an excellent candidate pair. Before this candidate pair is to be used for imaging experiments, I will first need to delete the NPC-specific candidate enhancer with CRISPR/Cas9 and confirm that it is required for *Sox2* expression. If this condition is met, I will be in a position to engineer the ANCHOR and MS2 labels, in order to follow what happens during gain of activity. However, in order to gain a better understanding of the interplay between chromatin mobility, transcriptional hub formation and function, it is necessary to study more enhancer-promoter pairs and see how general the *Sox2* properties are. In the future, it will be crucial to further probe the links between local chromatin mobility, Pol II clustering, TF dynamics and transcription readout. Combination of multicolor live-cell super-resolution imaging and single-molecule detection, at nanometer resolution, will be a major feat for further zooming into single genes and analyzing how transcription and chromatin are regulated.

My findings, together with others (Germier et al., 2017; Gu et al., 2018), highlight the importance of assessing chromatin dynamics at transcription sites as a means of understanding the nuclear microenvironment and regulatory principles. Notwithstanding, D_{app} and α are quite general diffusive parameters. For example, reduced α can be explained by multiple mechanisms, such as confinement to a reduced volume or “attraction” to specific regions, that cannot be distinguished. I look forward to advances in physical theory allowing more precise

inferred diffusive parameters to be defined, which could give better insights and allow even more accurate analysis of local chromatin dynamics. Coupled with more sophisticated imaging tools, and better methods for manipulating the nucleus (e.g. acute ablation of key regulatory proteins with the auxin-degradation system (N. Q. Liu et al., 2021; P. Nora et al., 2017; Rao et al., 2017), or manipulation of liquid-liquid phase separation properties (Y. Shin et al., 2018), comprehensive understanding of the four-dimensional aspects of transcriptional regulation is becoming more and more feasible.

Materials & Methods

ESC culture

The mouse F1 (M.*Musculus*^{129} x M.*Castaneus*) embryonic stem cells (ESC) were kindly provided by Pr. Jennifer Mitchell. Cells were cultured on 0.1% gelatin-coated plates in ES medium (DMEM 4.5 g/l glucose) containing 2mM Glutamax-I, 15% Foetal Calf Serum (FCS), 0.1 mM non-essential amino acids, 0.1 mM β - mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF; produced in house), 40 µg/ml gentamycin, 3 µM CHIR99021 (GSK3β inhibitor, Axon Medchem) and 1 µM PD0325901 (MEK inhibitor, Axon Medchem) in 5% CO_2 at 37 °C, in a humidified incubator. Media was changed daily and cells were passaged every 2 days (70-80% confluency).

Generation of stable cell lines containing ANCH1 and ANCH3 sequences

The presence of a single nucleotide polymorphism (SNP) in the PAM sequence of each sgRNA target on the M. Castaneus (CAST) allele caused preferential insertion of the ANCH sequences on the M.Musculus^{129} (129) allele (Zhou et al., 2014). The three ("Anchor_Sox2-SCR", "Anchor_Control" and "Anchor_Alexander", see **Table 3**) ANCHOR transgenic lines were generated by allele-specific (129) CRISPR/*Cas9*-mediated knock-in experiments in the following way. First, flanking homology arms (HA) (see **Table 3**) were amplified from a bacterial artificial chromosome (BAC RP23-274P9) accommodating the Sox2 locus using the Q5 High-Fidelity DNA Polymerase (New England Biolabs) and introduced into a vector containing either ANCH1 (Saad et al., 2014) or ANCH3 sequence (Germier et al., 2017) (both vectors were kindly provided by Pr. Kerstin Bystricky) using the Gibson Assembly Master Mix kit (New England Biolabs). Both donor vectors (1 µg) were co-transfected with 3 µg of a vector containing *Cas9*-GFP, a puromycin resistance marker (Puro^R) and a scaffold to encode the two sgRNAs specific to the two insertion sites (generated by the IGBMC Molecular Biology platform) in 1 million F1 ESCs using Lipofectamine 2000 Transfection Reagent (Invitrogen), according to the supplier's recommendation. Transfected cells were allowed to recover for two days, cultured for 24h with 3 µg/ml puromycin and then for 48h with 1 µg/ml puromycin. After

antibiotic selection, GFP-positive cells were sorted using a FACS Aria Fusion and seeded into 96-well plates. Once the single colonies had grown to 70-80% confluency they were split into four identical 96-well plates. The first plate was used for DNA extraction. Cells were washed with 1xPBS and then lysed for 1h at 56˚C with 30 µl lysis buffer (0.45% Tween, 0.45% Triton-X100, 2.5mM MgCl2, 50mM KCl, 10mM Tris-HCl pH8.3, 100 µg/ml proteinase K). 2 µl of genomic DNA was used for subsequent PCR screening using Expand Long Template PCR System (Roche). The second and third plates were frozen at -80C after resuspending the clones in 90% FCS/10% DMSO freezing media, while the fourth one was kept in culture for further experiments.

Determination of double-positive clones

Heterozygous clones with the correct sequences were screened by PCR in the following manner. A first (external) PCR was performed to amplify the whole candidate integration sites (5'HA-ANCH-3'HA) using Expand Long Template PCR System (Roche), a forward primer hybridizing upstream of the 5' HA and a reverse primer downstream of the 3' HA. Two additional (internal) PCRs were performed on the candidate clones in order to verify the insertion of each ANCH sequence using one external primer (forward or reverse) and one internal (reverse or forward) hybridizing to the ANCHOR. Insertions were further confirmed by sequencing. Primer sequences are given in **Table 4**.

RT-qPCR

Total RNA was extracted (Macherey-Nagel) from the final candidate clones, using at least one million cells, and reverse transcription for cDNA generation was performed using random hexamer primers (Thermo Fisher) and SuperScript IV (Invitrogen). RT-qPCR was performed on 10 ng cDNA in technical triplicates using QuantitTect SYBR Green PCR kit (Qiagen) on a LC480 Light Cycler (Roche). Amplification was normalized to SDHA (Succinate dehydrogenase complex flavoprotein subunit A). Primer sequences are given in **Table 4**. Microscopy experiments (see below) validated the heterozygous incorporation of the ANCH

sequences (detection of one single spot per ANCH sequence per cell) within the same allele (two spots were always in close proximity).

Generation of CRISPR/Cas9-mediated monoallelic SCR deletions

Monoallelic deletion (129 allele) of the 107 and 111 elements of the SCR, screening and expression analysis were facilitated by the presence of SNPs between the two alleles of the hybrid F1 (M.*Musculus*^{129} x M.*Castaneus*) ESCs and were prepared as previously described (Moorthy & Mitchell, 2016b). The vectors containing the sgRNAs, *Cas9*-GFP and *Cas9D10A*-GFP, as well as the allele-specific primers for genotyping and gene expression analysis were kindly provided by Pr. Jennifer Mitchell. In brief, four sgRNAs and Cas9D10A-GFP were used to delete SRR111 enhancer region from Anchor_Sox2-SCR ESCs. The Cas9D10A mutant cuts only one strand, so the 4 guides result in staggered cuts in the two strands on either side of the target region. Once the Anchor_ΔSRR111$^{129\text{-}/CAST+}$ ESCs were obtained (Clone 3, see **Table 3**), we used two sgRNAs and Cas9-GFP to delete the SRR107 region and make the Anchor_ΔSRR107.111$^{129\text{-}/CAST+}$ ESCs (Clone 30, see **Table 3**). Primer and sgRNA sequences are given in **Table 4**.

Generation of mESCs with 24 × MS2 loops at the Sox2 3′ UTR and stable expression of MCP-mScarletI, OR1-EGFP and OR3-IRFP

The targeting vector containing the 24xMS2 cassette (5'HA-T2A-LoxP-HSV_TK-T2A-Hygro^{R}-LoxP-24XMS2-3'HA) (**Figure 28**), as well as the vector containing *Cas9* and a scaffold to encode the sgRNA targeting the region upstream of the endogenous *Sox2* stop codon were designed and assembled by Dr. Tineke Lenstra. For introducing 24XMS2 loops in the *Sox2* locus, 1 million Anchor_Sox2-SCR ESCs were transfected with 2 µg sgRNA-Cas9 vector and 2 µg targeting Sox2-MS2 vector using Lipofectamine 2000 Transfection Reagent (Invitrogen). Cells were given three days to recover and were then subjected to 200 μg/ml hygromycin selection for nine days. Once a stable cell population was obtained we performed single cell sorting and seeded cells into 96-well plates. Once individual colonies were formed, clones were screened by PCR and heterozygous clones with a correctly targeted 129 allele were

further controlled by sequencing to confirm the correct insertion of the 24xMS2 array. In order to facilitate the PCR screening, primers #438_Sox2_FOR (5'-CGCCCAGTAGACTGCACAT) and #439_Sox2_REV (5'CCCTCCCAATTCCCTTGTAT) were inserted in the targeting vector. In this way, two bands were expected after heterozygous insertion of the loops (197 and 82 bp length), while only one band (155 bp) was present in the wt allele. Additional PCRs with allele-specific primers spanning the whole engineered region and single-molecule FISH (smFISH) (see below) were performed in order to confirm the insertion of the 24xMS2 loops in 129 allele. Primer sequences are given in **Table 4**.

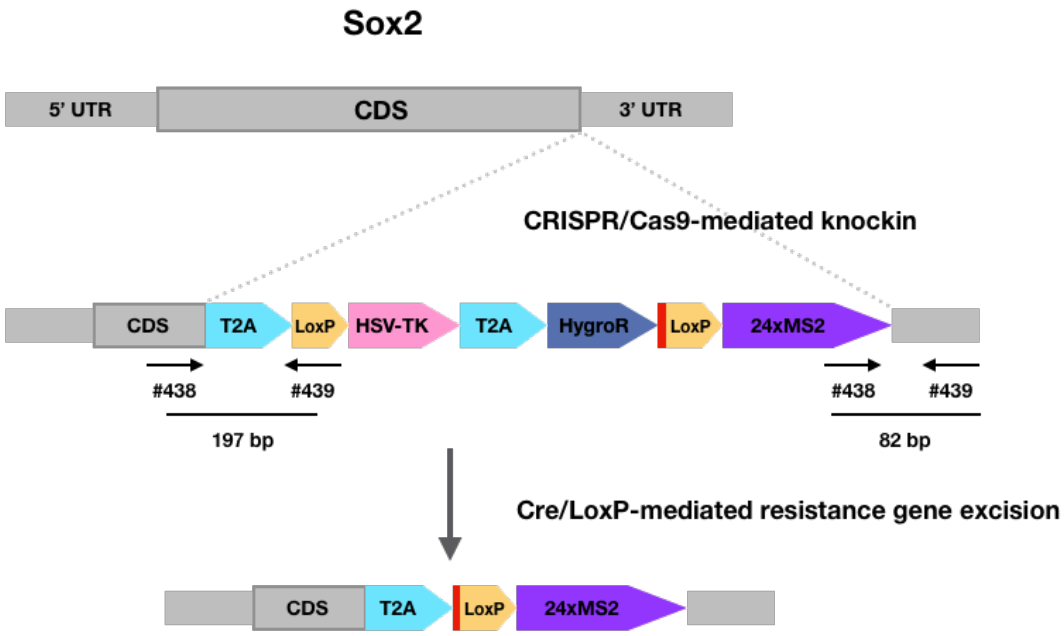


Figure 28: Schematic illustration of the reporter gene used to measure transcribed Sox2 in living mESCs.

To excise the LoxP-HSV_TK-T2A-Hygro^R resistance cassette, 1 million cells of a correctly targeted clone (H3H1, see **Table 3**) were transfected with 4 µg Cre-GFP vector using Lipofectamine 2000 Transfection Reagent (Invitrogen). After incubation for three days, cells were subjected to selection with 6 µM gancyclovir for ten days. Cells containing thymidine kinase (labeled as HSV-TK, **Figure 28**) did not survive after addition of gancyclovir and only the cells with successful LoxP removal survived. Once a stable cell population was obtained we performed single cell sorting and let individual clones grow for approximately ten days. To confirm the removal of the resistance marker (Hygro), smFISH was performed (see below) and

PCR over the insertion was used to select cells based on fragment length. A single clone (H3H1A1, see **Table 3**) with fully excised LoxP-HSV_TK-T2A-Hygro^{R} resistance cassette was used for all further experiments.

To generate cells stably expressing OR1, OR3 and MCP (MS2-coat protein) fluorescent fusion proteins, 1 million H3H1A1 cells, containing a 24 × MS2 cassette integrated in the 3′ UTR of 129 allele, were transfected with 1 µg ePiggyBac transposase expression plasmid (System Biosciences) and 1 µg epB-MCP-mScarletI, epB-EF1α-NLS-OR1-GFP, epB-EF1α-NLS-OR3-IRFP plasmids (OR1-EGFP and OR3-IRFP producing vectors available from NeoVirTech; were modified from the original source by changing the fluorescent protein at the C-terminus, introducing a Kozak sequence before the translation start site, introducing an N-terminal nuclear localization signal (NLS), and replacing the CMV promoter with EF-1α. For my thesis, both vectors were kindly provided by Pr. Kerstin Bystricky) using Lipofectamine 2000 (Invitrogen). Nine days after transfection, fluorescent cells were sorted in bins (low, medium and high) based on the expression of the fluorescent proteins. Imaging analysis was done for all bins to determine the optimal expression of the three fluorescent fusion proteins. The condition that we selected for all further experiments was: MCP-mScarletI medium/high, OR1-GFP low and OR3-IRFP low ("Anchor_Sox2-MS2" cell line, see **Table 3**).

Preparation of cells for live-imaging

150,000 cells were plated two days prior to imaging onto laminin-511-coated (BioLamina) 35mm glass bottom petri dishes (MatTek Corporation) and transfected with 2 µg of OR1-EGFP and OR3-IRFP producing plasmids (vectors available from NeoVirTech. For my thesis, both vectors were kindly provided by Pr. Kerstin Bystricky, and modified as described above except for the inclusion of NLS). Transfection was carried out using Lipofectamine 2000 Transfection Reagent (Invitrogen), according to the manufacturer's recommendation. The medium was changed just before imaging.

To examine the impact of the status of Pol II on chromatin motion in living cells, we used chemical inhibitors to block either transcription initiation (Triptolide) (Vispé et al., 2009) or elongation (Flavopiridol) (Chao & Price, 2001). ESC fresh medium was supplemented with

500 nM Triptolide (Sigma-Aldrich) or 1 μM Flavopiridol (Sigma-Aldrich) and cells were incubated for 3 or 4 hours respectively prior to image acquisition. Cells treated with DMSO (1:2000 dilution) for 4 hours were used as a control for each live-imaging experiment.

In order to obtain NPC-like cells (NPC, Neuronal Progenitor Cells) we modified a previously described protocol (Bibel, Richter, Lacroix, & Barde, 2007). On day 0, ESCs were passaged onto laminin-511-coated (BioLamina) 35mm glass bottom petri dishes (MatTek Corporation) and switched to a DMEM medium (4.5 g/l glucose) supplemented with 2mM Glutamax-I, 10% FCS, 0.1 mM non-essential amino acids, 0.1 mM β – mercaptoethanol and 40 μg/ml gentamycin, without LIF or CHIR99021 and PD0325901 inhibitors (2i). The following day, the medium was changed just before imaging ("day1" condition) and retinoic acid was added for the following two days ("day 2" and "day3") of culture.

Live-cell imaging of mESCs

Imaging experiments were performed on an inverted Nikon Eclipse Ti microscope equipped with a PFS (perfect focus system), a Yokogawa CSU-X1 confocal spinning disk unit, two sCMOS Photometrics Prime 95B cameras for simultaneous dual acquisition to provide 95% quantum efficiency at 11 μm x 11 μm pixels and a Leica 100x oil objective (HC PL APO 1,4 oil immersion). For double-label experiments using OR1-EGFP and OR3-IRFP, we excited EGFP and IRFP with a 491-nm (~100mw) and a 635-nm laser (>28mW), respectively. We detected green and far red fluorescence with an emission filter using a 525/50 nm and a 708/75 nm detection window, respectively. A thermostated heater (Tokai Hit Stage Top Incubator) allowed for heating at 37°C, humidity and CO_2 control (5%). Time-lapse analysis of GFP and IRFP foci was performed in 2D acquiring 241 time points at a 0.5 s time interval. For triple-label experiments using OR1-EGFP, OR3-IRFP and MCP-mScarletI, we first imaged EGFP and IRFP as above and then, at identical z-position, we excited mScarletI with 561-nm laser and we detected red fluorescence with an emission filter using a 609/54 nm detection window. Time-lapse analysis of GFP, IRFP and mScarletI foci was performed in 2D acquiring 181 time points at a 1 s time interval.
The system was controlled using Metamorph 7.10 software. The steps of the subsequent image analysis are depicted in **Figure 29.**

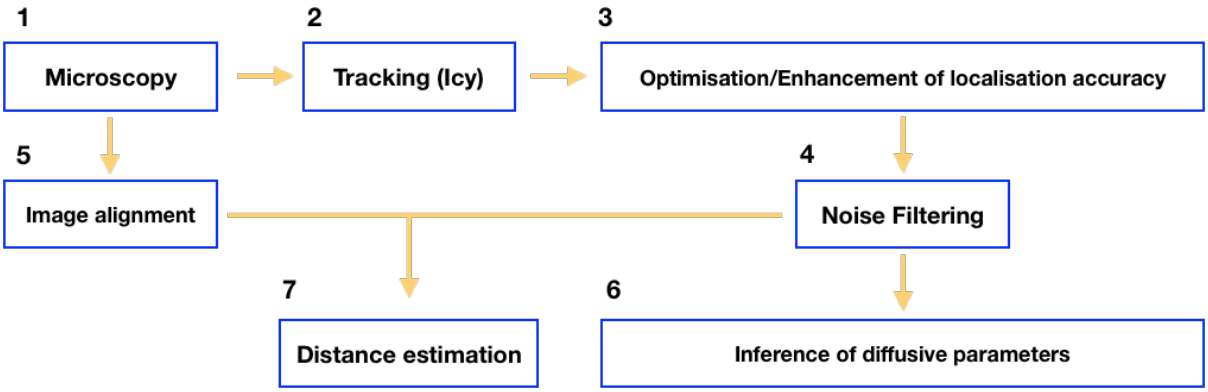


Figure 29: Steps of image processing and data analysis.

Image processing and data analysis

⇒ Tracking (Icy)

Raw microscopy images were loaded into ICY, an image analysis software (De Chaumont et al., 2012) (http://icy.bioimageanalysis.org), and saved as a single hyper-stack OME-TIF file compressed using LZW algorithm to reduce storage data size. All relevant metadata were stored in the same file. During the same session, particles were detected and linked using Icy plugins and exported in XML format individually per channel.

⇒ Optimizing / Enhancing localization accuracy

Despite an excellent overall performance of Icy, we frequently found that detected points were off-centered to detected particles and false positives were present (**Figure 30**). Trying to mitigate their effects, we model each particle using a 2D bell curve as follows:

$$S_{x,y} = I_o \exp\left\{-\frac{1}{2}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}^T \begin{bmatrix} L_x^2 & \theta L_x L_y \\ \theta L_x L_y & L_y^2 \end{bmatrix}^{-1} \begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix}\right\} + B_G.$$

The reference signal level is given by B_G, the spot has a maximum signal I_o and center of mass represented by μ. Parameters L_x and L_y represent spot size in respective directions and a degree of rotation is denotated by parameter θ.

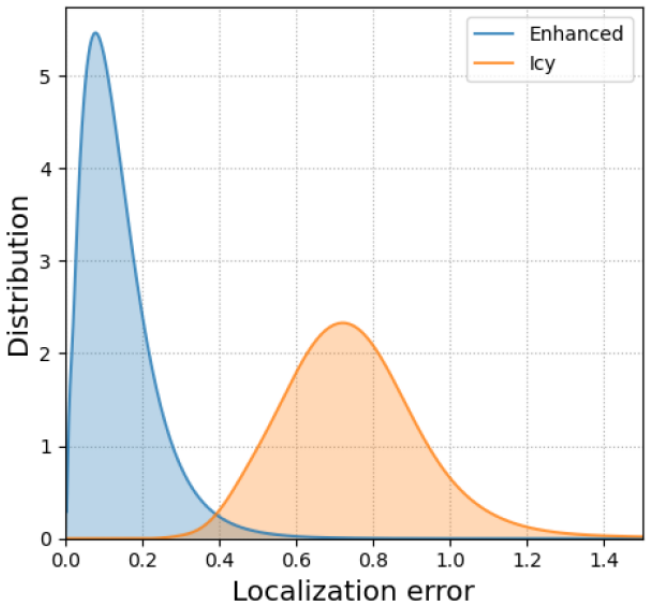


Figure 30: Comparison between results obtained using Icy and further optimized with aforementioned method. 500 particles were simulated to reproduce particles observed in microscopy sessions (Oliveira et al., 2021).

⇒ **Noise filtering**

Using signal intensity, spot size and localization error for all the particles tracked, we determine outliers by calculating median and inter-quartile ranges (IQR) for these 3 parameters. Any detection in which any of these parameters is superior/inferior to 2 IQR above/below median were considered as outliers and removed from remaining analysis.

⇒ **Image alignment**

The set-up using two cameras helped us with the measurement of distances and diffusive parameters. Unfortunately, it inserted non-negligible alignment discrepancies between channels. Furthermore, due to the usage of two different wavelengths, we could also observe chromatic aberration. Dealing with these inaccuracies is vital for the estimation of average distances between spots.

To correct these problems, we use a set of generic affine transformations to perform a digital post-alignment in 2 steps (Oliveira et al., 2021). The first step handles more grotesque errors associated with the dual camera setup, whilst the second will rescale the image and correct chromatic aberration (**Figure 31).**

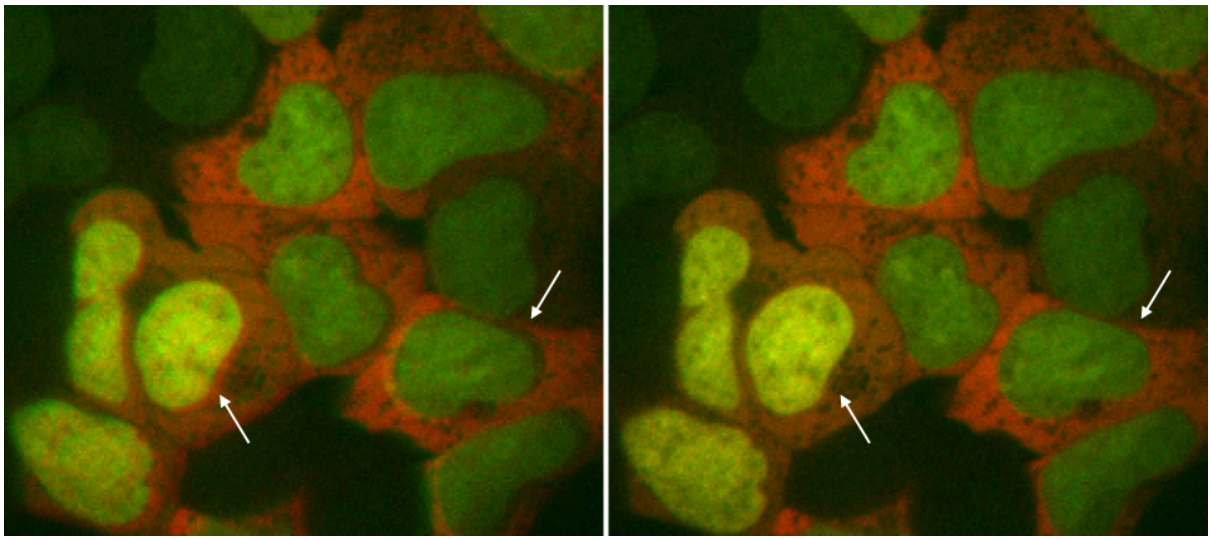


Figure 31: Image alignment using the GP-Tool program. White arrows point to regions with pronounced effects caused by chromatic aberrations. After alignment algorithm is deployed, both channels are properly superposed.

⇒ **Inference of diffusive parameters**

GP-FBM (Oliveira et al., 2021) allows measurement noise, misdetections and occlusions to be easily integrated into the workflow. By choosing GP-FBM, we assume that chromatin local dynamics follows a fractional Brownian type of motion (Oliveira et al., 2021), that is, the dynamics can be modelled by the covariance matrix:

$$\Sigma_{D,\alpha}(t,t') = D\left(|t|^{\alpha} + |t'|^{\alpha} - |t-t'|^{\alpha}\right).$$

This kernel produces a generalized Brownian motion, with mean squared displacement <r^2> = 2 n D t ^ α, where n corresponds to the number of degrees of freedom. In addition, we can easily add localization errors σ^2 to the diagonal terms of this matrix, assuming that these errors are decorrelated and normally distributed.

As a last point, GP-FBM considers confounded background movement generated by cell displacement, membrane fluctuations and/or chromatin reallocation by model implicit correlation present in the trajectories of 2 or more particles moving under similar context. If overlooked, background movement will be the cause of over-estimated dynamics parameters. More details are found in (Oliveira et al., 2021).

⇒ **Calculating distances between particles**

Finally, we can estimate 2D distance between particles in a Pythagorean framework once coordinates are properly calibrated by the alignment matrix as earlier introduced.

⇒ **Defining bursting windows**

For the triple-label movies, the ANCHOR labels were robustly detected in nearly all frames, whereas the MCP signal appeared to frequently be lost in the z-plane. Visual inspection of the MCP signal presence/absence across the movie frames (see Results) showed clear differences between prolonged absence of transcriptional bursting and intermittent loss of signal due to z-movement. “Bursting” movies were defined as the subsets of frames forming contiguous sections of at least 49 MCP-positive frames, with no periods of >7 missing frames; “non-bursting” movies were defined as the subsets of frames forming contiguous sections of at least 98 MCP-negative frames. These subsets were extracted for re-analysis by GPTool.

Circular chromatin conformation capture sequencing (4C-seq)

4C-seq using the SCR as a bait was performed as previously described (Van De Werken et al., 2012). Briefly, 5 million cells per biological condition were cross-linked with 2% formaldehyde in PBS and incubated on a rocker for 10 minutes at room temperature (RT). Nuclei were isolated in Lysis buffer (10 mM Tris-HCl pH8.0, 10 mM NaCl, 0.2% NP-40/Igepal CA-630, 1X protease inhibitor). Purified nuclei were digested overnight (O/N) at 37˚C with the first restriction enzyme (DpnII, 1500U final; New England Biolabs) and posteriorly subjected to an overnight proximity ligation at 16 ˚C with 20,000 U of T4 DNA Ligase (New England Biolabs). After ligation, chromatin was de-crosslinked and purified after proteinase K

and RNAse A treatment. Circularized DNA was then linearized by a second digestion with 5 U of Csp6I (Thermo Fisher) per µg of DNA, O/N at 37˚C, followed by final O/N ligation (5 ng/µl final DNA concentration) at 16˚C with T4 DNA Ligase (200 U per µg of DNA) and DNA purification using phenol/chloroform extraction.

Allele-specific 4C-seq was performed as previously described (Splinter et al., 2011) with slight modifications. The procedure is based on the presence of SNPs creating allele-specific restriction sites on one of the two alleles of hybrid F1 ESCs, thereby enabling amplification and direct analysis of either the 129 or the CAST allele. Instead of replacing the second restriction enzyme (Csp6I) with the "new" allele-specific restriction enzyme, we performed a normal 4C (DpnII and Csp6I digestion) and before the PCR amplification step we digested the 4C template with the appropriate restriction enzyme and purified the DNA using phenol/chloroform extraction. For the "Distal SCR" bait, AvaII restriction enzyme was used for 129-specific 4C-seq. A second viewpoint ("Proximal SCR") was used for the Anchor_ΔSRR111$^{129-/CAST+}$ cell line, since the "Distal SCR" region was deleted from the genome. For the "Proximal SCR" bait, Alw26I restriction enzyme was used for 129-specific 4C-seq and BveI for CAST-specific 4C-seq.

Typically, 600 ng of the resultant 4C DNA template were used to generate 4C-seq libraries by performing a PCR using Expand Long Template PCR System (Roche) with target-specific designed reading and non-reading primers (see **Table 4**) containing Illumina sequencer adapters. The 4C template was PCR amplified for 27 cycles and 6 reactions were pooled together. Optimal PCR conditions were found using a control template 4C material, comprising genomic DNA that is digested with the secondary enzyme and re-digested to make circularized 4C templates containing only the contiguous genomic sequence linked to the bait (Karasu & Sexton, 2021). Then, generated 4C-seq libraries were purified with SPRI select beads (Beckman Coulter) to discard primer dimer DNA products and 4C-seq DNA template were quantified using Bioanalyzer and pooled equimolarly for sequencing on Hiseq 4000 sequencer (IGBMC Genomeast platform) using single-end 50 bp reads (100 bp for the "Proximal SCR" bait).

4C-seq data analysis

Fastq sequencing files were demultiplexed and the bait sequence, up to but not including the DpnII site, were trimmed with the Sabre tool (https://github.com/najoshi/sabre) before mapping to the mm9 genome with Bowtie (Langmead, Trapnell, Pop, & Salzberg, 2009). The mapped reads were then processed and visualized by the 4See tool (Ben Zouari et al., 2020).

Virtual 4C plots

All fastq files from ES Hi-C data (Bonev et al., 2017)were downloaded and re-mapped to the mm9 genome assembly and normalized using FAN-C (Kruse, Hug, & Vaquerizas, 2020). The normalized matrix spanning the 2 Mb around the Sox2 gene at 4 kb resolution was extracted, and matrix rows containing all interactions with the designated "bait" were taken. Running means (window of five bins) were plotted against genomic coordinate.

Single molecule RNA Fluorescence *in situ* Hybridization (smFISH)

All smFISH experiments were carried out by Marit de Kort, a PhD student in Tineke Lenstra's team. All the probes (MS2v6, Sox2, Hygromycin) were designed via Stellaris Probe Designer and labeled with Cy3 or Cy5 on the 3' ends. Glass coverslips (VWR) were placed into a 12-well plate and coated O/N with 1 ml Poly-D-lysine (Merck-Millipore) (50 µl Poly-D-lysine (stock 50mg/ml) + 950 µl H_2O). 400,000 cells were seeded per well and allowed to attach for approximately three hours, until they reach 70-80% confluency. Cells were washed three times with pre-warmed HBSS (Hanks' Balanced Salt Solution; Thermo Fisher) buffer (no calcium, no magnesium, no phenol red). HBSS buffer was aspirated and cells were fixed with 4% PFA (Paraformaldehyde; Electron Microscopy Sciences) in PBS for 10 minutes at RT, followed by two 10-minutes washes with 1xPBS at RT. Cells were permeabilized with 70% EtOH at 4°C, O/N. The following day, combinations of two probes (2.5 µM of each probe) were mixed and resuspended in hybridization buffer (5 g Dextran sulfate sodium salt (Sigma-Aldrich), 35 ml of Ultrapure nuclease-free H_2O, 5 ml 20x SSC (saline-sodium citrate; Thermo Fisher), 5 ml deionized formamide (Sigma-Aldrich)) (2.2 µl probe mix + 52.8 µl hybridization buffer). Cells

were removed from EtOH and washed with 1-2 ml freshly prepared pre-warmed wash buffer (40 ml of Ultrapure nuclease-free H_2O, 5 ml of 20x SSC, 5 ml of deionized formamide) for 5 minutes at 37°C. Wash buffer was aspirated and coverslips were let dry for few minutes. 50 µl of resuspended probe hybridization mix was applied on a parafilm (placed inside a Petri dish) and coverslips were gently placed on top of each drop (cells facing the drop) to avoid bubbles. A humid kimwipe was placed next to the coverslips, the Petri dish was sealed with parafilm and placed O/N at 37°C. After the overnight incubation, coverslips were moved into a new 12-well containing 2 ml of pre-warmed wash buffer and incubated for 30 minutes at 37°C. The washing step was repeated for three times in total, each wash at 37°C for 30 minutes. Coverslips were then rinsed with 2xSSC and washed with 1xPBS for 5 minutes at RT. Samples were mounted in Prolong Gold with DAPI (Invitrogen) on glass slides (VWR), let to dry for 24 hours in the dark, and then transferred to -20°C or directly imaged.

Imaging was performed on Zeiss AxioObserver 7 inverted wide-field fluorescence microscope with LED illumination (SpectraX, Lumencor) and sCMOS ORCA Flash 4.0 V3 717 (Hamamatsu). A 40x oil objective lens (NA 1.4) with 1.6x Optovar was used. 27 z-stacks were imaged from -4 to 4 µM with 0.3 µM steps and 1x1 binning. An exposure time of 500ms was used for Cy3 (100% LED power), 750ms for Cy5 (100% LED power) and 25ms for DAPI (10% LED power). Micromanager 1.4 software was used. Image quantification was carried out using a custom Python pipeline. The scripts are available upon request to Lenstra Lab. Images were compressed to 3D image z stacks -4 to 4 μM. Cell and nuclear masks were determined using a custom Python algorithm. Spots corresponding to Sox2, MS2v6 and Hygromycin transcripts were then counted for cells and nuclei.

Western Blot

Western blots were performed using standard procedures. In brief, all whole cell extract lysates were prepared in RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris-HCl, pH 8.0, 1mM EDTA). Cell lysate was sonicated with Covaris E220 in AFA microtubes (PI:175, DF:10%, C/B:200, 60 s) to shear the genomic DNA and the lysate was cleared by centrifugation at 13,000 rpm, for 20 min, at 4°C. Protein concentration was measured with Bio-Rad Protein Assay. 30 μg of protein were diluted in 4x loading buffer (100 mM Tris-HCl pH 6,8, 30% glycerol, 4% SDS, 0,2% bromophenol blue, 400 mM DTT) and

boiled at 95°C, for 5 min. Protein lysate was loaded onto a Bis-Tris 10% polyacrylamide gel and electrophoresis was carried out using the Mini-Protean cell (Bio-Rad) system. Protein was transferred to a PVDF membrane. Membranes were blocked for 1 hour, at RT, with 4% milk PBS Tween (PBST). Membrane was incubated in primary antibody overnight in 4% milk PBST at 4°C, then washed four times, 15 min, at RT in PBST and incubated in secondary antibody in 4% milk PBST for 1 hour, at RT. After secondary incubation, membranes were washed four times, 5 min, at RT in PBST, incubated in Pierce ECL Western Blotting Substrate (ThermoFisher), and visualized by film exposure. Antibodies used were anti-Sox2 (Santa Cruz), anti-tubulin (IGBMC, 1TUB2A2) and anti-histone H3 (IGBMC, H31HH.3EI).

Table 3: Cell lines.

Cell Line	Clone	Parental line	Description	gRNA sequence	Location (mm9)	System used
Anchor_Sox2-SCR	36	F1 ESCs	Sox2-5.5C$^{ANCH3/+}$	CCCCTGGACAGCAACAGCAG	chr3:34,542,624 - 34,544,520	CRISPR-Cas9
			Sox2-109.4T$^{ANCH1/+}$	CCCAGCCTACCTCGAACTCA	chr3:34,657,473-34,659,303	
Anchor_Control	18	F1 ESCs	Sox2-55C$^{ANCH3/+}$	GTTCAAAAACTAGAAACA	chr3:34,603,689-34,605,222	CRISPR-Cas9
			Sox2-170T$^{ANCH1/+}$	CCTTGCAAGCACAAGGACGC	chr3:34,718,340-34,720,139	
Anchor_Alexander	B56	F1 ESCs	Sox2-8C$^{ANCH3/+}$	TATGTTAAGCTAGTTTCTGA	chr3:34,539,890-34,541,689	CRISPR-Cas9
			Sox2-117T$^{ANCH1/+}$	GTAAGCTATCTCATTGCCCG	chr3:34,665,107-34,666,618	
Anchor_ΔSRR111	3	Anchor_Sox2-SCR	ΔSRR111$^{129/CAST+}$	TAGCATCTGGCCAAGGAATG	chr3:34,659,726-34,661,602	CRISPR-Cas9
				CCCAACGTACATGTTTTTGT		
Anchor_ΔSRR107.111	30	Anchor_ΔSRR111	ΔSRR107.111$^{129/CAST+}$	AGCACAAAATAAAATTTAAG	chr3:34,656,540-34,658,026	Cas9 nickase
				CTATGCACATGCTGGGACCA		
				ACTAGAGCTCAACCTTGGCC		
				GGTTAGTTCTCTTCAGCAAG		
Anchor+MS2+HygroR	H3H1	Anchor_Sox2-SCR	24xMS2 loops introduced in the *Sox2* locus	GCCAGCCCTCACATGTGCGAC	chr3:34,549,639-34,550,797	CRISPR-Cas9
Anchor+ MS2-HygroR	H3H1A1	Anchor+MS2+HygroR	Removed resistance (Hygro^{R}) marker			Cre recombinase
Anchor_Sox2-MS2		H3H1A1	Cells stably express PiggyBac vectors epB-MCP-mScarletI; epB-EF1α-NLS-OR1-GFP; epB-EF1α-NLS-OR3-IRFP			PiggyBac transposon system

Table 4: Primers.

Primer Name	Sequence
“Anchor_Sox2-SCR” PCR screening	
Whole_Sox2_F	5’-TCAGGGAAAGCTGACGTTCT
Anchor3R3	5’-TCAACTCGCCAATCTCACCT
Anchor3R11	5’-CCTTGGCTGTGAACGGCAG
Anchor3S4	5’-TGACGCTGTTTTAAGTGCCC
Anchor3S10	5’-GCCGCAAACCAGGGAAAA
Whole_Sox2_R	5’-GGGAATGGCCCATGCATCTG
whole_SCR_F1	5’-TGCTAACTACCCACCCCTTG
Anchor1R6	5’-CATTGCCTCTGCAGTCGGC
Block Anchor1 R	5’-AAGCTTAAGTTTAAACGCTAGCTGCC
Anchor1S2	5’-CCGAGGATCCAGGCAATGT
whole_SCR_R	5’-GAGTCTTTGCCCAAGCTGTC
Anchor1S4	5’-GCCGACTGCAGAGGCAATG
“Anchor_Control” PCR screening	
Whole_ Control_ Sox2_F	5’-CCTTGCAGAGGACCCACACT
Anchor3R2	5’-CGTTGCAGTCACAAACACCCA
Anchor3S	5’-TGGGTGTTTGTGACTGCAAC
Whole_ Control_Sox2_R	5’-ACACAGTTGCCCTCAGAGGC
Whole_ Control_ SCR_F	5’-AGGCCAGGCCTTTCCATACA
Anchor1R6	5’-CATTGCCTCTGCAGTCGGC
Anchor1S2	5’-CCGAGGATCCAGGCAATGT
Whole_ Control_ SCR_R	5’-AGGAGTTCCAGGTCAACCAGC
“Anchor_Alexander” PCR screening	
Whole_Sox2_8C _For	5’-AAAGCTCTACGCTCGACCTC

Anchor3R	5’-TCAACTCGCCAATCTCACCT
Anchor3S	5’-TGGGTGTTTGTGACTGCAAC
Whole_Sox2_8C_Rev	5’-TGAGGACTTGGCCTGGACTC
Whole_Sox2_117T_For	5’-ACGCTGGGTTTGACAGCA
Anchor1R2	5’-ACATTGCCTGGATCCTCGG
Anchor1S4	5’-GCCGACTGCAGAGGCAATG
Whole_Sox2_117T_Rev	5’-CCTCACTGAGCCTAAGCCTCTA
RT-qPCR for expression analysis	
mSDHA For	5’-GCTCCTGCCTCTGTGGTTGA
mSDHA Rev	5’-AGCAACACCGATGAGCCTG
Oct4 For	5’-TGGAAAGCAACTCAGAGGGA
Oct4 Rev	5’-TTCTGCAGGGCTTTCATGTC
Nanog For	5’-AGCAGAAGTACCTCAGCCTC
Nanog Rev	5’-CCGCTTGCACTTCATCCTTT
Sox2 For	5’-GCTCGCAGACCTACATGAAC
Sox2 Rev	5’-TGGAGTGGGAGGAAGAGGTA
Sox2 (129) Rev	5’-CGCCTAACGTACCACTAGAACTTT
Sox2 (CAST) Rev	5’-CGCCTAACGTACCACTAGAACTTA
Actb For	5’-CTAAGGCCAACCGTGAAAAGAT
Actb Rev	5’-CACAGCCTGGATGGCTACGT
Myc For	5’-TACAATCTGCGAGCCAGGAC
Myc Rev	5’-AAGTTCACGTTGAGGGGCAT
RT-qPCR for screening CRISPR deleted clones	
SRR111_int_129_F	5’-GAAAGGCAGGGAGTTGCAGT
SRR111_int_R	5’-TCCCTCACACCTGAAGGGTA
SRR111_int_CAST_F	5’-AGATGAAAGGCAGGGAGTAGTG

111delF	5’-ACCACGCCTAGTCATTGATATGT
111delR	5’-GAACTGAAGTCTGTCTGGTCCAC
SRR111delscrn_F_129	5’-CTGCTGAGCTGAGACCCTGA
SRR111_CAST_ext_Fc	5’-CCCTGCAGGTCTCCTATTCC
SRR111_CAST_ext_Rc2	5’-ACTTGCTTGTAATCTAGCTCCTG
111del_leftg_intact_R	5’-ATGTCCCACCTTTATAGCACTCA
111del_rightg_intact_F	5’-TTGCAGATGAGGTATTTTTGACA
SRR107_ext_F	5’-CCGGGTCTTATTCTGTAGCAC
107del_129_extR	5’-GACGGGTCATGGTGACACAC
107_extscreen_R_CAST	5’-GACGGGTCATGGTGACACAG
inside_SNP_test_For_1	5’-CGAGCCGGGAGGGAGTCCAA
inside_SNP_test_Rev_1	5’-CCTTTGATCCCAGCACTTGGAGGGC
PCR for screening mESCs with 24 × MS2 loops	
438 SOX2_FOR	5’-CGCCCAGTAGACTGCACAT
439 SOX2_REV	5’-CCCTCCCAATTCCCTTGTAT
CE-27-20	5’-AGATGCAGCCGATGCACCGA
CE-28-20	5’-CGCCTAACGTACCACTAGAACTTA
CE-29-20	5’-AGATGCAACCGATGCACCGC
CE-30-20	5’-CGCCTAACGTACCACTAGAACTTT
CE-31-20	5’-CAGCCTGATTCCAATAACAGAGCCG
CE-34-20	5’-GCATGATGCAGGAGCAGCTGG
CE-38-20	5’-CACACGGAAGATCTATCGATCTCGAGATTG
CE-39-40	5’-GTTGCTATGGCCGCGAGAACG
4C-seq primers (Distal SCR bait)	
4C DpnII SCR distal Pr2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXXXXGGGGAGGTCAGACACCTGATC

4C CSP SCR distal Pr1_ada	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTTTCCGGTAGGGGTGGAGC
4C-seq primers (Proximal SCR bait)	
4C DpnII BfuAI Pr2	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXXGCAAGAGCCAGGTGTGGCTC
4C Csp BfuAI Pr2	CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTCCTGGTGCTTTGCCCAGCAC
Illumina adapter sequence (red); Index (blue); Sequence specific primer (black)	

Table 5: Reagents.

Reagent	Source	Identifier
18 mm round coverslips	VWR	Cat#631-0153
32% Paraformaldehyde	Electron Microscopy Sciences	Cat#15714-S
35mm glass bottom petri dish	MatTek Corporation	Cat#P35G-1.5-20-C
anti-Sox2	Santa Cruz	Cat# sc-365823,Lot#K1414
BioLaminin	BioLamina	Cat#LN511
CHIR99021	Axon Medchem	Cat#1386
Csp6I	Thermo Fisher	Cat#ER0211
Dextran sulfate sodium salt	Sigma-Aldrich	Cat#67578
DpnII	New England Biolabs	Cat#R0543M
Expand Long Template PCR System	Roche	Cat#11759060001
Flavopiridol	Sigma-Aldrich	Cat#F3055
Formamide	Sigma-Aldrich	Cat#F9037
Gibson Assembly Master Mix	New England Biolabs	Cat#E2611
HBSS	Thermo Fisher	Cat#14175095
Lipofectamine 2000 Transfection Reagent	Invitrogen	Cat#11668019
Microscope slides, SuperFrost	VWR	Cat#631-0117
Nucleospin RNA purification kit	Macherey-Nagel	Cat#740955.50
PD0325901	Axon Medchem	Cat#1408
Pierce ECL Western Blotting Substrate	Thermo Fisher	Cat#32109
Poly-D-lysine	Merck-Millipore	Cat#A-003-E
ProLong Gold Antifade Mountant with DAPI	Invitrogen	Cat#P36935
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat#M0492

QIAquick PCR Purification Kit	Qiagen	Cat#28104
QuantitTect SYBR Green PCR kit	Qiagen	Cat#204345
Random hexamer primers	Thermo Fisher	Cat#SO142
SPRIselect	Beckman Coulter	Cat# B23319
SuperScript IV	Invitrogen	Cat#18090050
T4 DNA Ligase	New England Biolabs	Cat#M0202M
Triptolide	Sigma-Aldrich	Cat#T3652
UltraPure, SSC 20x	Thermo Fisher	Cat#15557044

References

Abudayyeh, O. O., Gootenberg, J. S., Essletzbichler, P., Han, S., Joung, J., Belanto, J. J., … Zhang, F. (2017). RNA targeting with CRISPR-Cas13. *Nature*, *550*(7675), 280–284. https://doi.org/10.1038/nature24049

Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature Reviews. Genetics*, *13*(10), 720–731. https://doi.org/10.1038/nrg3293

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., … Weiner, O. D. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *ELife*, *8*, 1–42. 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., … 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

Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., … Lacroute, P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, *491*(7422), 56–65. https://doi.org/10.1038/nature11632

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., & Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Developmental Cell*, *16*(1), 47–57. https://doi.org/10.1016/j.devcel.2008.11.011

Amarillo, I. E., Dipple, K. M., & Quintero-Rivera, F. (2013). Familial Microdeletion of 17q24.3 Upstream of SOX9 Is Associated With Isolated Pierre Robin Sequence Due to Position Effect. *American Journal of Medical Genetics Part A*, *161*(5), 1167–1172. https://doi.org/https://doi.org/10.1002/ajmg.a.35847

Anderson, E., & Hill, R. E. (2014). Long range regulation of the sonic hedgehog gene. *Current Opinion in Genetics & Development*, *27*, 54–59. https://doi.org/10.1016/j.gde.2014.03.011

Andrey, G., Montavon, T., Mascrez, B., Federico, G., Noordermeer, D., Leleu, M., … Duboule, D. (2013). A Switch Between Topological Domains Underlies HoxD Genes Collinearity in Mouse Limbs. *Science*, *13*(9), 613–626. https://doi.org/10.1038/nrg3207

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*, *339*(6123), 1074–1077. https://doi.org/10.1126/science.1232542

Bagheri-Fam, S., Barrionuevo, F., Dohrmann, U., Günther, T., Schüle, R., Kemler, R., … Scherer, G. (2006). Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Developmental Biology*, *291*(2), 382–397. https://doi.org/10.1016/j.ydbio.2005.11.013

Bahr, C., Von Paleske, L., Uslu, V. V., Remeseiro, S., Takayama, N., Ng, S. W., … Spitz, F. (2018). A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature*, *553*(7689), 515–520. https://doi.org/10.1038/nature25193

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., & Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO Journal*, *28*(24), 3785–3798. https://doi.org/10.1038/emboj.2009.340

Banerji, J., Olson, L., & Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell*, *33*(3), 729–740. https://doi.org/10.1016/0092-8674(83)90015-6

Banerji, J., Rusconi, S., & Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, *27*(2 PART 1), 299–308. https://doi.org/10.1016/0092-8674(81)90413-X

Barakat, T. S., Halbritter, F., Zhang, M., Rendeiro, A. F., Perenthaler, E., Bock, C., & Chambers, I. (2018). Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell*, *23*(2), 276-288.e8. https://doi.org/10.1016/j.stem.2018.06.014

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

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., … Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, *543*(7646), 519–524. https://doi.org/10.1038/nature21411

Becskei, A., Kaufmann, B. B., & Van Oudenaarden, A. (2005). Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics*, *37*(9), 937–944. https://doi.org/10.1038/ng1616

Beliveau, B. J., Boettiger, A. N., Avendaño, M. S., Jungmann, R., McCole, R. B., Joyce, E.

F., … Wu, C. T. (2015). Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using Oligopaint FISH probes. *Nature Communications*, *6*(May). https://doi.org/10.1038/ncomms8147

Ben Zouari, Y., Platania, A., Molitor, A. M., & Sexton, T. (2020). 4See: A Flexible Browser to Explore 4C Data. *Frontiers in Genetics*, *10*(January), 1–13. https://doi.org/10.3389/fgene.2019.01372

Benabdallah, N. S., Gautier, P., Hekimoglu-Balkan, B., Lettice, L. A., Bhatia, S., & Bickmore, W. A. (2016). SBE6: A novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells. *Open Biology*, *6*(11). https://doi.org/10.1098/rsob.160197

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Grimes, G. R., Therizols, P., & Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell*, 1–12. https://doi.org/10.1016/j.molcel.2019.07.038

Bender, M. A., Bulger, M., Close, J., & Groudine, M. (2000). β-globin Gene switching and DNase I sensitivity of the endogenous β-globin locus in mice do not require the locus control region. *Molecular Cell*, *5*(2), 387–393. https://doi.org/10.1016/S1097-2765(00)80433-5

Bender, M. A., Roach, J. N., Halow, J., Close, J., Alami, R., Bouhassira, E. E., … Fiering, S. N. (2001). Targeted deletion of 5′HS1 and 5′HS4 of the β-globin locus control region reveals additive activity of the DNaseI hypersensitive sites. *Blood*, *98*(7), 2022–2027. https://doi.org/10.1182/blood.V98.7.2022

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., … Lyonnet, S. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics*, *41*(3), 359–364. https://doi.org/10.1038/ng.329

Benoist, C., & Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. *Nature*, *290*(5804), 304–310. https://doi.org/10.1038/290304a0

Bensaude, O. (2011). Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription*, *2*(3), 103–108. https://doi.org/10.4161/trns.2.3.16172

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., … Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome.

Proceedings of the National Academy of Sciences of the United States of America, *99*(2), 757–762. https://doi.org/10.1073/pnas.231608898

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., & Long, R. M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular Cell*, *2*(4), 437–445. https://doi.org/10.1016/S1097-2765(00)80143-4

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, M. T., De Marco, R., … Kleinjan, D. A. (2013). Disruption of autoregulatory feedback by a mutation in a remote, ultraconserved PAX6 enhancer causes aniridia. *American Journal of Human Genetics*, *93*(6), 1126–1134. https://doi.org/10.1016/j.ajhg.2013.10.028

Bibel, M., Richter, J., Lacroix, E., & Barde, Y.-A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nature Protocols*, *2*(5), 1034–1043. https://doi.org/10.1038/nprot.2007.147

Bintu, B., Mateo, L. J., Su, J. H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., … Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, *362*(6413). https://doi.org/10.1126/science.aau1783

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., … De Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, *447*(7146), 799–816. https://doi.org/10.1038/nature05874

Blake, W. J., Kærn, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, *422*(6932), 633–637. https://doi.org/10.1038/nature01546

Blow, M. J. (2010). ChIP-seq Identification of Weakly Conserved Heart Enhancers. *Atherosclerosis*, *280*(9), 183–191. https://doi.org/10.1016/j.atherosclerosis.2018.11.031

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A. S., Yu, T., Marie-Nelly, H., … Zweckstetter, M. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. *Nature Structural and Molecular Biology*, *25*(9), 833–840. https://doi.org/10.1038/s41594-018-0112-y

Boettiger, A. N., & Levine, M. (2009). Synchronous and stochastic patterns of gene activation in the drosophila embryo. *Science*, *325*(5939), 471–473. https://doi.org/10.1126/science.1173976

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. V., … Young, R. A. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell*, *175*(7), 1842-1855.e16.

https://doi.org/10.1016/j.cell.2018.10.042

Bonev, B., & Cavalli, G. (2016). Organization and function of the 3D genome. *Nature Reviews Genetics*, *17*(11), 661–678. https://doi.org/10.1038/nrg.2016.112

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., … Cavalli, G. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, *171*(3), 557-572.e24. https://doi.org/10.1016/j.cell.2017.09.043

Bonn, S., Zinzen, R. P., Girardot, C., Gustafson, E. H., Perez-Gonzalez, A., Delhomme, N., … Furlong, E. E. M. (2012). Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics*, *44*(2), 148–156. https://doi.org/10.1038/ng.1064

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., … Crawford, G. E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, *132*(2), 311–322. https://doi.org/10.1016/j.cell.2007.12.014

Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R., & Nasmyth, K. (1985). Characterization of a “silencer” in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell*, *41*(1), 41–48. https://doi.org/10.1016/0092-8674(85)90059-5

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., … Wysocka, J. (2014). Reorganization of enhancer patterns in transition from naive to primed pluripotency. *Cell Stem Cell*, *14*(6), 838–853. https://doi.org/10.1016/j.stem.2014.04.003

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–198. https://doi.org/10.1038/nature08450

Cao, S. X., Gutman, P. D., Dave, H. P. G., & Schechter, A. N. (1989). Identification of a transcriptional silencer in the 5'-flanking region of the human ε-globin gene. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(14), 5306–5309. https://doi.org/10.1073/pnas.86.14.5306

Capel, B. (2017). Vertebrate sex determination: Evolutionary plasticity of a fundamental switch. *Nature Reviews Genetics*, *18*(11), 675–689. https://doi.org/10.1038/nrg.2017.60

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J. B., Espinola, S. M., Gurgo, J., Messina, O., … Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell*, *74*(1), 212-222.e5.

https://doi.org/10.1016/j.molcel.2019.01.011

Carter, D., Chakalova, L., Osborne, C. S., Dai, Y. feng, & Fraser, P. (2002). Long-range chromatin regulatory interactions in vivo. *Nature Genetics*, *32*(4), 623–626. https://doi.org/10.1038/ng1051

Cattoni, D. I., Gizzi, A. M. C., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., … Nollmann, M. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01962-x

Chao, S. H., & Price, D. H. (2001). Flavopiridol Inactivates P-TEFb and Blocks Most RNA Polymerase II Transcription in Vivo. *Journal of Biological Chemistry*, *276*(34), 31793–31799. https://doi.org/10.1074/jbc.M102306200

Chen, B. C., Legant, W. R., Wang, K., Shao, L., Milkie, D. E., Davidson, M. W., … Betzig, E. (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. *Science*, *346*(6208). https://doi.org/10.1126/science.1257998

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., … Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479–1491. https://doi.org/10.1016/j.cell.2013.12.001

Chen, B., Zou, W., Xu, H., Liang, Y., & Huang, B. (2018). Efficient labeling and imaging of protein-coding genes in living cells using CRISPR-Tag. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-07498-y

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018a). 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, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J. B., & Gregor, T. (2018b). 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., … 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, M., Ma, Z., Wu, X., Mao, S., Yang, Y., Tan, J., … Chen, A. K. (2017). A molecular beacon-based approach for live-cell imaging of RNA transcripts with minimal target engineering at the single-molecule level. *Scientific Reports*, *7*(1), 1–11.

https://doi.org/10.1038/s41598-017-01740-1

Chew, J.-L., Loh, Y.-H., Zhang, W., Chen, X., Tam, W.-L., Yeap, L.-S., … Ng, H.-H. (2005). Reciprocal Transcriptional Regulation of Pou5f1 and Sox2 via the Oct4/Sox2 Complex in Embryonic Stem Cells. *Molecular and Cellular Biology*, *25*(14), 6031–6046. https://doi.org/10.1128/mcb.25.14.6031-6046.2005

Cho, W. K., Jayanth, N., English, B. P., Inoue, T., Andrews, J. O., Conway, W., … Cisse, I. I. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. *ELife*, *5*(MAY2016), 1–31. https://doi.org/10.7554/eLife.13617

Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., & Cisse, I. I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science*, *361*(6400), 412–415. https://doi.org/10.1126/science.aar4199

Chodosh, L. A., Fire, A., Samuels, M., & Sharp, P. A. (1989). 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole inhibits transcription elongation by RNA polymerase II in vitro. *Journal of Biological Chemistry*, *264*(4), 2250–2257. https://doi.org/10.1016/s0021-9258(18)94169-4

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G. M., Cattoglio, C., … Tjian, R. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science*, *361*(6400). https://doi.org/10.1126/science.aar2555

Chubb, J. R., Trcek, T., Shenoy, S. M., & Singer, R. H. (2006). Transcriptional Pulsing of a Developmental Gene. *Current Biology*, *16*(10), 1018–1025. https://doi.org/10.1016/j.cub.2006.03.092

Cisse, I. I., Izeddin, I., Causse, S. Z., Boudarene, L., Senecal, A., Muresan, L., … Hajj, B. (2013). Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells. *Science*, *245*(August), 664–667.

Collis, P., Antoniou, M., & Grosveld, F. (1990). Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *The EMBO Journal*, *9*(1), 233–240. https://doi.org/10.1002/j.1460-2075.1990.tb08100.x

Conic, S., Desplancq, D., Ferrand, A., Fischer, V., Heyer, V., Martin, B. R. S., … Tora, L. (2018). Imaging of native transcription factors and histone phosphorylation at high resolution in live cells. *Journal of Cell Biology*, *217*(4), 1537–1552. https://doi.org/10.1083/jcb.201709153

Costlow, N., & Lis, J. T. (1984). High-resolution mapping of DNase I-hypersensitive sites of

Drosophila heat shock genes in Drosophila melanogaster and Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *4*(9), 1853–1863. https://doi.org/10.1128/mcb.4.9.1853

Cox, D. B. T., Gootenberg, J. S., Abudayyeh, O. O., Franklin, B., Kellner, M. J., Joung, J., & Zhang, F. (2017). RNA editing with CRISPR-Cas13. *Science*, *358*(6366), 1019–1027. https://doi.org/10.1126/science.aaq0180

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., … Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21931–21936. https://doi.org/10.1073/pnas.1016071107

Dao, L. T. M., Galindo-Albarrán, A. O., Castro-Mondragon, J. A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., … Spicuglia, S. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nature Genetics*, *49*(7), 1073–1081. https://doi.org/10.1038/ng.3884

Dar, R. D., Razooky, B. S., Singh, A., Trimeloni, T. V., McCollum, J. M., Cox, C. D., … Weinberger, L. S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(43), 17454–17459. https://doi.org/10.1073/pnas.1213530109

De Bruin, D., Zaman, Z., Liberatore, R. A., & Ptashne, M. (2001). Telomere looping permits gene activation by a downstream UAS in yeast. *Nature*, *409*(6816), 109–113. https://doi.org/10.1038/35051119

De Chaumont, F., Dallongeville, S., Chenouard, N., Hervé, N., Pop, S., Provoost, T., … Olivo-Marin, J. C. (2012). Icy: An open bioimage informatics platform for extended reproducible research. *Nature Methods*, *9*(7), 690–696. https://doi.org/10.1038/nmeth.2075

De Gobbi, M., Viprakasit, V., Hughes, J. R., Fisher, C., Buckle, V. J., Ayyub, H., … Higgs, D. R. (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science*, *312*(5777), 1215–1217. https://doi.org/10.1126/science.1126431

de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., … Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biology*, *8*(5). https://doi.org/10.1371/journal.pbio.1000384

de Wit, E., Bouwman, B. a M., Zhu, Y., Klous, P., Splinter, E., Verstegen, M. J. a M., … de Laat, W. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. *Nature*, *501*(7466), 227–231. https://doi.org/10.1038/nature12420

Dekker, J. (2006). The three "C" s of chromosome conformation capture: controls, controls, controls. *Nature Methods*, *3*(1), 17–21. https://doi.org/10.1038/nmeth823

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

Del Bene, F., Ettwiller, L., Skowronska-Krawczyk, D., Baier, H., Matter, J. M., Birney, E., & Wittbrodt, J. (2007). In vivo validation of a computationally predicted conserved Ath5 target gene set. *PLoS Genetics*, *3*(9), 1661–1671. https://doi.org/10.1371/journal.pgen.0030159

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P. D., … 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., … 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

Denholtz, M., Bonora, G., Chronis, C., Splinter, E., de Laat, W., Ernst, J., … Plath, K. (2013). Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell*, *13*(5), 602–616. https://doi.org/10.1016/j.stem.2013.08.013

Denker, A., & De Laat, W. (2016). The second decade of 3C technologies: Detailed insights into nuclear organization. *Genes and Development*, *30*(12), 1357–1382. https://doi.org/10.1101/gad.281964.116

Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., … Ren, B. (2017). A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells. *Nature Methods*, *14*(6), 629–635. https://doi.org/10.1038/nmeth.4264

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-bourget, J. E., Lee, A. Y., … Ecker, J. R. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*(7539), 331–336. https://doi.org/10.1038/nature14222

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., … Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, *489*(7414), 101–108. https://doi.org/10.1038/nature11233

Doni Jayavelu, N., Jajodia, A., Mishra, A., & Hawkins, R. D. (2020). Candidate silencer elements for the human and mouse genomes. *Nature Communications*, *11*(1), 1–15. https://doi.org/10.1038/s41467-020-14853-5

Dorighi, K. M., Swigut, T., Henriques, T., Bhanu, N. V., Scruggs, B. S., Nady, N., … Wysocka, J. (2017). Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell*, *66*(4), 568-576.e4. https://doi.org/10.1016/j.molcel.2017.04.018

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., … Dekker, J. (2006). Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Research*, *16*(10), 1299–1309. https://doi.org/10.1101/gr.5571506

Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C., & Taddei, A. (2011). Tight protein-DNA interactions favor gene silencing. *Genes and Development*, *25*(13), 1365–1370. https://doi.org/10.1101/gad.611011

Dubarry, N., Pasta, F., & Lane, D. (2006). ParABS systems of the four replicons of Burkholderia cenocepacia: New chromosome centromeres confer partition specificity (Journal of Bacteriology (2006) 188, 4 (1489-1496)). *Journal of Bacteriology*, *188*(8), 3164. https://doi.org/10.1128/JB.188.8.3164.2006

Dufourt, J., Trullo, A., Hunter, J., Fernandez, C., Lazaro, J., Dejean, M., … Lagha, M. (2018). Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. *Nature Communications*, *9*(1), 1–13. https://doi.org/10.1038/s41467-018-07613-z

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Esnault, C., Ghavi-Helm, Y., Brun, S., Soutourina, J., Van Berkum, N., Boschiero, C., … Werner, M. (2008). Mediator-Dependent Recruitment of TFIIH Modules in Preinitiation Complex. *Molecular Cell*, *31*(3), 337–346. https://doi.org/10.1016/j.molcel.2008.06.021

Espinola, S. M., Gotz, M., Fiche, J.-B., Bellec, M., Houbron, C., Cardozo-Gizzi, A. M., … Nollmann, M. (2021). Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nature Genetics*, *53*(April). https://doi.org/10.1038/s41588-021-00816-z

Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., …

Brangwynne, C. P. (2016). Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell*, *165*(7), 1686–1697. https://doi.org/10.1016/j.cell.2016.04.047

Filonov, G. S., Moon, J. D., Svensen, N., & Jaffrey, S. R. (2014). Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society*, *136*(46), 16299–16308. https://doi.org/10.1021/ja508478x

Finn, E. H., Pegoraro, G., Brandão, H. B., Valton, A. L., Oomen, M. E., Dekker, J., … Misteli, T. (2019). Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, *176*(6), 1502-1515.e10. https://doi.org/10.1016/j.cell.2019.01.020

Flyamer, I. M., Gassler, J., Imakaev, M., Brandão, H. B., Ulianov, S. V, Abdennur, N., … Tachibana-konwalski, K. (2017). Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature Publishing Group*, *544*(7648), 110–114. https://doi.org/10.1038/nature21711

Franke, M., Ibrahim, D. M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., … Pombo, A. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature Publishing Group*, *538*(7624), 265–269. https://doi.org/10.1038/nature19800

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., & Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature*, *466*(7305), 490–493. https://doi.org/10.1038/nature09158

Fromm, M., & Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Molecular and Cellular Biology*, *3*(6), 991–999. https://doi.org/10.1128/mcb.3.6.991

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

Fulco, C. P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S. R., Perez, E. M., … Engreitz, J. M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, *354*(6313), 769–773. https://doi.org/10.1126/science.aag2445

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., … Engreitz, J. M. (2019). Activity-by-Contact model of enhancer specificity from thousands of CRISPR perturbations (CRISPRi FlowFISH). *Nature Genetics*,

51(December), 529990. https://doi.org/10.1101/529990

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. Bin, … Ruan, Y. (2009). An oestrogen-receptor-α-bound human chromatin interactome. *Nature*, *462*(7269), 58–64. https://doi.org/10.1038/nature08497

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J.-M., Singer, R. H., & Bertrand, E. (2003). Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells associated with localized mRNA and are required for their movements [5-8]. In yeast, ASH1 mRNPs are asso-ciated with a specific myosin motor, and their move. *Current Biology*, *13*(02), 161–167.

Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I., & Bai, Y. (2018). Emerging roles of linker histones in regulating chromatin structure and function. *Nature Reviews Molecular Cell Biology*, *19*(3), 192–206. https://doi.org/10.1038/nrm.2017.94

Gasperini, M., Hill, A. J., McFaline-Figueroa, J. L., Martin, B., Kim, S., Zhang, M. D., … Shendure, J. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell*, *176*(1–2), 377-390.e19. https://doi.org/10.1016/j.cell.2018.11.029

Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H. A., Sellou, H., … Bystricky, K. (2017). Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophysical Journal*, *113*(7), 1383–1394. https://doi.org/10.1016/j.bpj.2017.08.014

Ghavi-Helm, Y., Klein, F. A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., & Furlong, E. E. M. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. *Nature*, *512*(7512), 96–100. https://doi.org/10.1038/nature13417

Gilchrist, D. A., Dos Santos, G., Fargo, D. C., Xie, B., Gao, Y., Li, L., & Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell*, *143*(4), 540–551. https://doi.org/10.1016/j.cell.2010.10.004

Gilchrist, D. A., Nechaev, S., Lee, C., Ghosh, S. K. B., Collins, J. B., Li, L., … Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes and Development*, *22*(14), 1921–1933. https://doi.org/10.1101/gad.1643208

Gillies, S. D., Morrison, S. L., Oi, V. T., & Tonegawa, S. (1983). A tissue-specific

transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell*, *33*(3), 717–728. https://doi.org/10.1016/0092-8674(83)90014-4

Giorgetti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., … Heard, E. (2014). Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell*, *157*(4), 950–963. https://doi.org/10.1016/j.cell.2014.03.025

Giorgetti, L., & Heard, E. (2016). Closing the loop: 3C versus DNA FISH. *Genome Biology*, *17*(1), 215. https://doi.org/10.1186/s13059-016-1081-2

Gisselbrecht, S. S., Palagi, A., Kurland, J. V., Rogers, J. M., Ozadam, H., Zhan, Y., … Bulyk, M. L. (2020). Transcriptional Silencers in Drosophila Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Molecular Cell*, *77*(2), 324-337.e8. https://doi.org/10.1016/j.molcel.2019.10.004

Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), 1025–1036. https://doi.org/10.1016/j.cell.2005.09.031

Gonen, N., Futtner, C. R., Wood, S., Alexandra Garcia-Moreno, S., Salamone, I. M., Samson, S. C., … Lovell-Badge, R. (2018). Sex reversal following deletion of a single distal enhancer of Sox9. *Science*, *360*(6396), 1469–1471. https://doi.org/10.1126/science.aas9408

Gregor, T., Tank, D. W., Wieschaus, E. F., & Bialek, W. (2007). Probing the Limits to Positional Information. *Cell*, *130*(1), 153–164. https://doi.org/10.1016/j.cell.2007.05.025

Gröschel, S., Sanders, M. A., Hoogenboezem, R., De Wit, E., Bouwman, B. A. M., Erpelinck, C., … Delwel, R. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in Leukemia. *Cell*, *157*(2), 369–381. https://doi.org/10.1016/j.cell.2014.02.019

Grosschedl, R., & Birnstiel, M. A. X. L. (1980). *Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo*. *77*(12), 7102–7106.

Grosveld, F., van Assendelft, G. B., Greaves, D. R., & Kollias, G. (1987). Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, *51*(6), 975–985. https://doi.org/10.1016/0092-8674(87)90584-8

Gu, B., Swigut, T., Spencley, A., Bauer, M. R., Chung, M., Meyer, T., & Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory

elements. *Science*, *359*(6379), 1050–1055. https://doi.org/10.1126/science.aao3136

Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hager, G. L., McNally, J. G., & Misteli, T. (2009). Transcription Dynamics. *Molecular Cell*, *35*(6), 741–753. https://doi.org/10.1016/j.molcel.2009.09.005

Hah, N., Benner, C., Chong, L. W., Yu, R. T., Downes, M., & Evansa, R. M. (2015). Inflammation-sensitive super enhancers form domains of coordinately regulated enhancer RNAs. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(3), E297–E302. https://doi.org/10.1073/pnas.1424028112

Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2011). A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, *145*(4), 622–634. https://doi.org/10.1016/j.cell.2011.03.042

Hah, N., Murakami, S., Nagari, A., Danko, C. G., & Lee Kraus, W. (2013). Enhancer transcripts mark active estrogen receptor binding sites. *Genome Research*, *23*(8), 1210–1223. https://doi.org/10.1101/gr.152306.112

Harlen, K. M., & Churchman, L. S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, *18*(4), 263–273. https://doi.org/10.1038/nrm.2017.10

Harper, C. V., Finkenstädt, B., Woodcock, D. J., Friedrichsen, S., Semprini, S., Ashall, L., … White, M. R. H. (2011). Dynamic analysis of stochastic transcription cycles. *PLoS Biology*, *9*(4). https://doi.org/10.1371/journal.pbio.1000607

Hatzis, P., & Talianidis, I. (2002). Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Molecular Cell*, *10*(6), 1467–1477. https://doi.org/10.1016/S1097-2765(02)00786-4

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., … Higgs, D. R. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*, *48*(8), 895–903. https://doi.org/10.1038/ng.3605

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., … 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

Heist, T., Fukaya, T., & Levine, M. (2019). Large distances separate coregulated genes in living Drosophila embryos. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(30), 15062–15067. https://doi.org/10.1073/pnas.1908962116

Hendrix, D. A., Hong, J. W., Zeitlinger, J., Rokhsar, D. S., & Levine, M. S. (2008). Promoter elements associated with RNA Pol II stalling in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(22), 7762–7767. https://doi.org/10.1073/pnas.0802406105

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S. A., Belver, L., Wendorff, A. A., … Ferrando, A. A. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nature Medicine*, *20*(10), 1130–1137. https://doi.org/10.1038/nm.3665

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., … Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, *155*(4). https://doi.org/10.1016/j.cell.2013.09.053

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., … Young, R. A. (2015). Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Molecular Cell*, *58*(2), 362–370. https://doi.org/10.1016/j.molcel.2015.02.014

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

Hong, J. W., Hendrix, D. A., & Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science*, *321*(5894), 1314. https://doi.org/10.1126/science.1160631

Hsieh, T. H. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Tjian, R., & Darzacq, X. (2020). Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell*, *78*(3), 539-553.e8. https://doi.org/10.1016/j.molcel.2020.03.002

Hughes, J. R., Roberts, N., Mcgowan, S., Hay, D., Giannoulatou, E., Lynch, M., … Higgs, D.

R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, *46*(2), 205–212. https://doi.org/10.1038/ng.2871

Iborra, F. J., Pombo, A., Jackson, D. A., & Cook, P. R. (1996). Active RNA polymerases are localized within discrete transcription "factories" in human nuclei. *Journal of Cell Science*, *109*(6), 1427–1436.

Jacome, A., & Fernandez-Capetillo, O. (2011). Lac operator repeats generate a traceable fragile site in mammalian cells. *EMBO Reports*, *12*(10), 1032–1038. https://doi.org/10.1038/embor.2011.158

Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M., & Epstein, D. J. (2006). A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development*, *133*(4), 761–772. https://doi.org/10.1242/dev.02239

Jiang, J., Cai, H., Zhou, Q., & Levine, M. (1993). Conversion of a dorsal-dependent silencer into an enhancer: Evidence for dorsal corepressors. *EMBO Journal*, *12*(8), 3201–3209. https://doi.org/10.1002/j.1460-2075.1993.tb05989.x

Jin, Q., Yu, L. R., Wang, L., Zhang, Z., Kasper, L. H., Lee, J. E., … Ge, K. (2011). Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *EMBO Journal*, *30*(2), 249–262. https://doi.org/10.1038/emboj.2010.318

Johnson, D. S., Mortazavi, A., & Myers, R. M. (2007). *Genome-Wide Mapping of in Vivo Protein-DNA Interactions*. (June), 1497–1503.

Jonkers, I., Kwak, H., & Lis, J. T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *ELife*, *2014*(3), 1–25. https://doi.org/10.7554/eLife.02407

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., … Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), 430–435. https://doi.org/10.1038/nature09380

Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., & Chen, L. (2011). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology*, *30*(1), 90–98. https://doi.org/10.1038/nbt.2057

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., … Huang, B. (2016). Versatile protein tagging in cells with split fluorescent protein. *Nature*

Communications, *7*, 1–9. https://doi.org/10.1038/ncomms11046

Kan, Y. W., Dozy, A. M., Varmus, H. E., Taylor, J. M., Holland, J. P., Lie-Injo, L. E., … Todd, D. (1975). Deletion of α-globin genes in haemoglobin-H disease demonstrates multiple α-globin structural loci. *Nature*, *255*(5505), 255–256. https://doi.org/10.1038/255255a0

Karasu, N., & Sexton, T. (2021). 4C-Seq: Interrogating Chromatin Looping with Circular Chromosome Conformation Capture. In B. Bodega & C. Lanzuolo (Eds.), *Capturing Chromosome Conformation: Methods and Protocols* (pp. 19–34). https://doi.org/10.1007/978-1-0716-0664-3_3

Karpova, T. S., Chen, T. Y., Sprague, B. L., & McNally, J. G. (2004). Dynamic interactions of a transcription factor with DNA are accelerated by a chromatin remodeller. *EMBO Reports*, *5*(11), 1064–1070. https://doi.org/10.1038/sj.embor.7400281

Karpova, T. S., Kim, M. J., Spriet, C., Nalley, K., Stasevich, T. J., Kherrouche, Z., … McNally, J. G. (2008). Concurrent fast and slow cycling of a transcriptional activator at an endogenous promoter. *Science*, *319*(5862), 466–469. https://doi.org/10.1126/science.1150559

Kheradpour, P., Stark, A., Roy, S., & Kellis, M. (2007). Reliable prediction of regulator targets using 12 Drosophila genomes. *Genome Research*, *17*(12), 1919–1931. https://doi.org/10.1101/gr.7090407

Kim, J. H., Rege, M., Valeri, J., Dunagin, M. C., Metzger, A., Titus, K. R., … Phillips-Cremins, J. E. (2019). LADL: light-activated dynamic looping for endogenous gene expression control. *Nature Methods*, *16*(7), 633–639. https://doi.org/10.1038/s41592-019-0436-5

Kim, S., & Shendure, J. (2019). Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Molecular Cell*, *76*(2), 306–319. https://doi.org/10.1016/j.molcel.2019.08.010

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., … Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), 182–187. https://doi.org/10.1038/nature09033

Kim, Y. J., Björklund, S., Li, Y., Sayre, M. H., & Kornberg, R. D. (1994). A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell*, *77*(4), 599–608. https://doi.org/10.1016/0092-8674(94)90221-6

Kimura, H., & Cook, P. R. (2001). Kinetics of core histones in living human cells: Little exchange of H3 and H4 and some rapid exchange of H2B. *Journal of Cell Biology*, *153*(7), 1341–1353. https://doi.org/10.1083/jcb.153.7.1341

Kininis, M., Chen, B. S., Diehl, A. G., Isaacs, G. D., Zhang, T., Siepel, A. C., … Kraus, W. L. (2007). Genomic Analyses of Transcription Factor Binding, Histone Acetylation, and Gene Expression Reveal Mechanistically Distinct Classes of Estrogen-Regulated Promoters. *Molecular and Cellular Biology*, *27*(14), 5090–5104. https://doi.org/10.1128/mcb.00083-07

Kioussis, D., Vanin, E., Delange, T., Flavell, R. A., & Grosveld, F. G. (1983). β-Globin gene inactivation by DNA translocation in γβ-thalassaemi. *Nature*, *306*(5944), 662–666. https://doi.org/10.1038/306662a0

Klann, T. S., Black, J. B., Chellappan, M., Safi, A., & Song, L. (2017). *CRISPR–Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome*. *35*(6), 561–568. https://doi.org/10.1038/nbt.3853.CRISPR

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., … Andrau, J. C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural and Molecular Biology*, *18*(8), 956–963. https://doi.org/10.1038/nsmb.2085

Kornberg, R. (1974). Chromatin Structure : A Repeating Unit of Histones and DNA. *Science*, *184*, 868–871.

Kornberg, R., & Thomas, J. (1973). *Chromatin Structure: Oligomers of the Histones*. *637*, 0–3.

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T. H. S., … Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, *78*(3), 554-565.e7. https://doi.org/10.1016/j.molcel.2020.03.003

Krijger, P. H. L., & De Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. *Nature Reviews Molecular Cell Biology*, *17*(12), 771–782. https://doi.org/10.1038/nrm.2016.138

Krumm, A., Meulia, T., Brunvand, M., & Groudine, M. (1992). The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes and Development*, *6*(11), 2201–2213. https://doi.org/10.1101/gad.6.11.2201

Kruse, K., Hug, C. B., & Vaquerizas, J. M. (2020). FAN-C: a feature-rich framework for the analysis and visualisation of chromosome conformation capture data. *Genome Biology*,

21(1), 1–19. https://doi.org/10.1186/s13059-020-02215-9

Lam, M. T. Y. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *J Autism Dev Disord*, *47*(3), 549–562. https://doi.org/10.1097/CCM.0b013e31823da96d.Hydrogen

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., … Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650–665. https://doi.org/10.1016/j.cell.2018.01.029

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3). https://doi.org/10.1186/gb-2009-10-3-r25

Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A., & Singer, R. H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science*, *332*(6028), 475–478. https://doi.org/10.1126/science.1202142

Lauderdale, J. D., Wilensky, J. S., Oliver, E. R., Walton, D. S., & Glaser, T. (2000). 3' deletions cause aniridia by preventing PAX6 gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(25), 13755–13759. https://doi.org/10.1073/pnas.240398797

Lee, Y. H., & Saint-Jeannet, J. P. (2011). Sox9 function in craniofacial development and disease. *Genesis*, *49*(4), 200–208. https://doi.org/10.1002/dvg.20717

Leighton J. Core, Waterfall, J. J., & Lis, J. T. (2008). Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Reform and Revolt in the City of Dreaming Spires*, *322*(DECEMBER), 277–313. https://doi.org/10.2307/j.ctv9zcj2n.53

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., … de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), 1725–1735. https://doi.org/10.1093/hmg/ddg180

Lettice, L. A., Williamson, I., Wiltshire, J. H., Peluso, S., Devenney, P. S., Hill, A. E., … Hill, R. E. (2012). Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. *Developmental Cell*, *22*(2), 459–467. https://doi.org/10.1016/j.devcel.2011.12.010

Li, C., Cesbron, F., Oehler, M., Brunner, M., & Höfer, T. (2018). Frequency Modulation of Transcriptional Bursting Enables Sensitive and Rapid Gene Regulation. *Cell Systems*,

6(4), 409-423.e11. https://doi.org/10.1016/j.cels.2018.01.012

Li, J., Dong, A., Saydaminova, K., Chang, H., Wang, G., Ochiai, H., … Pertsinidis, A. (2019). Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell*, *178*(2), 491-506.e28. https://doi.org/10.1016/j.cell.2019.05.029

Li, Jieru, Hsu, A., Hua, Y., Wang, G., Cheng, L., Ochiai, H., … Pertsinidis, A. (2020). Single-gene imaging links genome topology, promoter–enhancer communication and transcription control. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/s41594-020-0493-6

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., … Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, *498*(7455), 516–520. https://doi.org/10.1038/nature12210

Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G., & Lieb, J. D. (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, *484*(7393), 251–255. https://doi.org/10.1038/nature10985

Lieberman-aiden, E., Berkum, N. L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Mirny, L. A. (2009). *Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome*. *33292*(October), 289–294.

Lim, B. (2018). Imaging transcriptional dynamics. *Current Opinion in Biotechnology*, *52*(Pol II), 49–55. https://doi.org/10.1016/j.copbio.2018.02.008

Lim, B., Heist, T., Levine, M., & Fukaya, T. (2018). Visualization of Transvection in Living Drosophila Embryos. *Molecular Cell*, *70*(2), 287-296.e6. https://doi.org/10.1016/j.molcel.2018.02.029

Lim, B., & Levine, M. S. (2021). Enhancer-promoter communication : hubs or loops ? *Current Opinion in Genetics & Development*, *67*, 5–9. https://doi.org/10.1016/j.gde.2020.10.001

Lin, Y. C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., … Murre, C. (2012). Global changes in the nuclear positioning of genes and intra-and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunology*, *13*(12), 1196–1204. https://doi.org/10.1038/ni.2432

Lin, Yi, Mori, E., Kato, M., Xiang, S., Wu, L., Kwon, I., & McKnight, S. L. (2016). Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers. *Cell*, *167*(3), 789-802.e12. https://doi.org/10.1016/j.cell.2016.10.003

Lin, Yihan, Sohn, C. H., Dalal, C. K., Cai, L., & Elowitz, M. B. (2015). Combinatorial gene regulation by modulation of relative pulse timing. *Nature*, *527*(7576), 54–58. https://doi.org/10.1038/nature15710

Ling, J., Ainol, L., Zhang, L., Yu, X., Pi, W., & Tuan, D. (2004). HS2 enhancer function is blocked by a transcriptional terminator inserted between the enhancer and the promoter. *Journal of Biological Chemistry*, *279*(49), 51704–51713. https://doi.org/10.1074/jbc.M404039200

Liu, N. Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M. M. G. A., Teunissen, H., … de Wit, E. (2021). WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics*, *53*(1), 100–109. https://doi.org/10.1038/s41588-020-00744-4

Liu, Y., Kung, C., Fishburn, J., Ansari, A. Z., Shokat, K. M., & Hahn, S. (2004). Two Cyclin-Dependent Kinases Promote RNA Polymerase II Transcription and Formation of the Scaffold Complex. *Molecular and Cellular Biology*, *24*(4), 1721–1735. https://doi.org/10.1128/mcb.24.4.1721-1735.2004

Liu, Z., Legant, W. R., Chen, B. C., Li, L., Grimm, J. B., Lavis, L. D., … Tjian, R. (2014). 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *ELife*, *3*, e04236. https://doi.org/10.7554/eLife.04236

Liu, Z., & Tjian, R. (2018). Visualizing transcription factor dynamics in living cells. *Journal of Cell Biology*, *217*(4), 1181–1191. https://doi.org/10.1083/jcb.201710038

Loffreda, A., Jacchetti, E., Antunes, S., Rainone, P., Daniele, T., Morisaki, T., … Mazza, D. (2017). Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00398-7

Long, H. K., Osterwalder, M., Welsh, I. C., Hansen, K., Davies, J. O. J., Liu, Y. E., … Wysocka, J. (2020). Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell*, *27*(5), 765-783.e14. https://doi.org/10.1016/j.stem.2020.09.001

Long, H. K., Prescott, S. L., & Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell*, *167*(5), 1170–1187. https://doi.org/10.1016/j.cell.2016.09.018

Lorch, Y., Zhang, M., & Kornberg, R. D. (1999). Histone Octamer Transfer by a Chromatin-Remodeling Complex. *Cell*, *96*, 389–392.

Los, G. V, Encell, L. P., Mcdougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., … Wood, K. V. (2008). *HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis*. *3*(6), 373–382.

Lovén, J., Hoke, H. A., Lin, C. Y., Lau, A., Orlando, D. A., Vakoc, C. R., … Young, R. A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, *153*(2), 320–334. https://doi.org/10.1016/j.cell.2013.03.036

Lu, H., Yu, D., Hansen, A. S., Ganguly, S., Liu, R., Heckert, A., … Zhou, Q. (2018). Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature*, *558*(7709), 318–323. https://doi.org/10.1038/s41586-018-0174-3

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, *389*(6648), 251–260. https://doi.org/10.1038/38444

Lupiáñez, D. G., Kraft, K., Visel, A., & Mundlos, S. (2015). *Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Article Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions*. 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004

Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S., & Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3002–3007. https://doi.org/10.1073/pnas.1420024112

Ma, H., Reyes-Gutierrez, P., & Pederson, T. (2013). Visualization of repetitive DNA sequences in human chromosomes with transcription activator-like effectors. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(3), 1222. https://doi.org/10.1073/pnas.1323266111

Mahmoudi, T., Katsani, K. R., & Verrijzer, C. P. (2002). GAGA can mediate enhancer function in trans by linking two separate DNA molecules. *EMBO Journal*, *21*(7), 1775–1781. https://doi.org/10.1093/emboj/21.7.1775

Mansour, M. R., Abraham, B. J., Anders, L., Berezovskaya, A., Gutierrez, A., Durbin, A. D., … Look, A. T. (2014). An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*, *346*(6215), 1373–1377. https://doi.org/10.1126/science.1259037

Marinić, M., Aktas, T., Ruf, S., & Spitz, F. (2013). An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. *Developmental*

Cell, *24*(5), 530–542. https://doi.org/10.1016/j.devcel.2013.01.025

Marshall, N. F., & Price, D. H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. *Journal of Biological Chemistry*, *270*(21), 12335–12338. https://doi.org/10.1074/jbc.270.21.12335

Marshall, W. F., Straight, A., Marko, J. F., Swedlow, J., Dernburg, A., Belmont, A., … Sedat, J. W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. *Current Biology*, *7*(12), 930–939. https://doi.org/10.1016/S0960-9822(06)00412-X

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., … Heard, E. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell*, *145*(3), 447–458. https://doi.org/10.1016/j.cell.2011.03.032

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., & Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature*, *568*(7750), 49–54. https://doi.org/10.1038/s41586-019-1035-4

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., … Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science*, *337*(6099), 1190–1195. https://doi.org/10.1126/science.1222794

May, D., Blow, M. J., Kaplan, T., McCulley, D. J., Jensen, B. C., Akiyama, J. A., … Visel, A. (2012). Large-scale discovery of enhancers from human heart tissue. *Nature Genetics*, *44*(1), 89–93. https://doi.org/10.1038/ng.1006

Mayran, A., & Drouin, J. (2018). Pioneer transcription factors shape the epigenetic landscape. *Journal of Biological Chemistry*, *293*(36), 13795–13804. https://doi.org/10.1074/jbc.R117.001232

McCord, R. P., Kaplan, N., & Giorgetti, L. (2020). Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell*, 1–21. https://doi.org/10.1016/j.molcel.2019.12.021

McKnight, S. L., & Miller, O. L. (1979). Post-replicative nonribosomal transcription units in D. melanogaster embryos. *Cell*, *17*(3), 551–563. https://doi.org/10.1016/0092-8674(79)90263-0

McNally, J. C., Müller, W. G., Walker, D., Wolford, R., & Hager, G. L. (2000). The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science*, *287*(5456), 1262–1265. https://doi.org/10.1126/science.287.5456.1262

Melo, C. A., Drost, J., Wijchers, P. J., van de Werken, H., de Wit, E., Vrielink, J. A. F. O., … Agami, R. (2013). ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. *Molecular Cell*, *49*(3), 524–535. https://doi.org/10.1016/j.molcel.2012.11.021

Mercola, M., Wang, X. F., Olsen, J., & Calame, K. (1983). Transcriptional enhancer elements in the mouse immunoglobulin heavy chain locus. *Science*, *221*(4611), 663–665. https://doi.org/10.1126/science.6306772

Mifsud, B., Tavares-Cadete, F., Young, A. N., Sugar, R., Schoenfelder, S., Ferreira, L., … 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

Mir, M., Reimer, A., Haines, J. E., Li, X. Y., Stadler, M., Garcia, H., … Darzacq, X. (2017). Dense bicoid hubs accentuate binding along the morphogen gradient. *Genes and Development*, *31*(17), 1784–1794. https://doi.org/10.1101/gad.305078.117

Mir, M., Stadler, M. R., Ortiz, S. A., Harrison, M. M., Darzacq, X., & Eisen, M. B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *BioRxiv*, (Dv), 1–27. https://doi.org/10.1101/377812

Miyanari, Y., Ziegler-Birling, C., & Torres-Padilla, M.-E. (2013). Live visualization of chromatin dynamics with fluorescent TALEs. *Nature Structural & Molecular Biology*, *20*(11), 1321–1324. https://doi.org/10.1038/nsmb.2680

Mizuguchi, G., Shen, X., Landry, J., Wu, W. H., Sen, S., & Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science*, *303*(5656), 343–348. https://doi.org/10.1126/science.1090701

Mlynarczyk-Evans, S., Royce-Tolland, M., Alexander, M. K., Andersen, A. A., Kalantry, S., Gribnau, J., & Panning, B. (2006). X chromosomes alternate between two states prior to random X-inactivation. *PLoS Biology*, *4*(6), 0906–0916. https://doi.org/10.1371/journal.pbio.0040159

Molina, N., Suter, D. M., Cannavo, R., Zoller, B., Gotic, I., & Naef, F. (2013). Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(51), 20563–20568. https://doi.org/10.1073/pnas.1312310110

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., … Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. *Cell*,

147(5), 1132–1145. https://doi.org/10.1016/j.cell.2011.10.023

Moorthy, S. D., Davidson, S., Shchuka, V. M., Singh, G., Malek-Gilani, N., Langroudi, L., … 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 Research*, *27*(2), 246–258. https://doi.org/10.1101/gr.210930.116

Moorthy, S. D., & Mitchell, J. A. (2016a). Generating CRISPR/Cas9 mediated monoallelic deletions to study enhancer function in mouse embryonic stem cells. *Journal of Visualized Experiments*, *2016*(110), 1–2. https://doi.org/10.3791/53552

Moorthy, S. D., & Mitchell, J. A. (2016b). Generating CRISPR/Cas9 mediated monoallelic deletions to study enhancer function in mouse embryonic stem cells. *Journal of Visualized Experiments*, *2016*(110), 1–12. https://doi.org/10.3791/53552

Moreau, Hen, Wasylyk, Everett, Gaub, & Chambon, P. (1981). The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants P.Moreau,. *Nucleic Acids Research*, *9*(20), 2589–2598.

Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., … Wang, K. C. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nature Communications*, *8*(May), 1–9. https://doi.org/10.1038/ncomms15993

Mousavi, K., Zare, H., Dell'Orso, S., Grontved, L., Gutierrez-Cruz, G., Derfoul, A., … Sartorelli, V. (2013). ERNAs Promote Transcription by Establishing Chromatin Accessibility at Defined Genomic Loci. *Molecular Cell*, *51*(5), 606–617. https://doi.org/10.1016/j.molcel.2013.07.022

Mueller, B., Mieczkowski, J., Kundu, S., Wang, P., Sadreyev, R., Tolstorukov, M. Y., & Kingston, R. E. (2017). Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes and Development*, *31*(5), 451–462. https://doi.org/10.1101/gad.293118.116

Müller, H.-P., Sogo, J., & Schaffner, W. (1989). An enhancer stimulates transcription in Trans when attached to the promoter via a protein bridge. *Cell*, *58*(4), 767–777. https://doi.org/10.1016/0092-8674(89)90110-4

Muse, G. W., Gilchrist, D. A., Nechaev, S., Shah, R., Parker, J. S., Grissom, S. F., … Adelman, K. (2007). RNA polymerase is poised for activation across the genome. *Nature Genetics*, *39*(12), 1507–1511. https://doi.org/10.1038/ng.2007.21

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., … Fraser, P.

(2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. https://doi.org/10.1038/nature12593

Narlikar, G. J., Fan, H., & Kingston, R. E. (2002). Cooperation between Complexes that Regulate Chromatin Structure and Transcription Chromatin structure creates barriers for each step. *Cell*, *108*, 475–487.

Neuberger, M. S. (1983). Expression and regulation of immunoglobulin heavy chain gene transfected into lymphoid cells. *The EMBO Journal*, *2*(8), 1373–1378. https://doi.org/10.1002/j.1460-2075.1983.tb01594.x

Neumann, F. R., Dion, V., Gehlen, L. R., Tsai-Pflugfelder, M., Schmid, R., Taddei, A., & Gasser, S. M. (2012). Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Genes and Development*, *26*(4), 369–383. https://doi.org/10.1101/gad.176156.111

Ni, Z., Saunders, A., Fuda, N. J., Yao, J., Suarez, J.-R., Webb, W. W., & Lis, J. T. (2008). P-TEFb Is Critical for the Maturation of RNA Polymerase II into Productive Elongation In Vivo. *Molecular and Cellular Biology*, *28*(3), 1161–1170. https://doi.org/10.1128/mcb.01859-07

Nir, G., Farabella, I., Estrada, C. P., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., … Wu, C. ting. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. In *bioRxiv*. https://doi.org/10.1101/374058

Nojima, T., Gomes, T., Grosso, A. R. F., Kimura, H., Dye, M. J., Dhir, S., … Proudfoot, N. J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell*, *161*(3), 526–540. https://doi.org/10.1016/j.cell.2015.03.027

Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., … Heard, E. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, *485*(7398), 381–385. https://doi.org/10.1038/nature11049

Nora, P., Goloborodko, A., Valton, A., Dekker, J., Mirny, L. A., Bruneau, B. G., … Abdennur, N. (2017). *Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Article Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization*. 930–944. https://doi.org/10.1016/j.cell.2017.05.004

Nozaki, T., Imai, R., Tanbo, M., Nagashima, R., Tamura, S., Tani, T., … Maeshima, K. (2017). Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Molecular Cell*, *67*(2), 282-293.e7.

https://doi.org/10.1016/j.molcel.2017.06.018

Ochiai, H., Sugawara, T., & Yamamoto, T. (2015). Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Research*, *43*(19), e127. https://doi.org/10.1093/nar/gkv624

Olins, A. L., & Olins, D. E. (1974). Spheroid chromatin units (v bodies). *Science*, *183*(4122), 330–332. https://doi.org/10.1126/science.183.4122.330

Oliveira, G. M., Oravecz, A., Kobi, D., Maroquenne, M., Bystricky, K., Sexton, T., & Molina, N. (2021). *Precise measurements of chromatin diffusion dynamics by modeling using Gaussian processes*. 1–27. Retrieved from https://doi.org/10.1101/2021.03.16.435699

Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., … Pennacchio, L. A. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, *554*(7691), 239–243. https://doi.org/10.1038/nature25461

Oudelaar, A. M., Davies, J. O. J., Hanssen, L. L. P., Telenius, J. M., Schwessinger, R., Liu, Y., … 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

Paige, J., Wu, K., & Jaffrey, S. R. (2011). RNA Mimics of Green Fluorescent Protein. *Science*, *333*(July), 201–217. https://doi.org/10.4159/harvard.9780674333987.c22

Palstra, R.-J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., & de Laat, W. (2003). The β-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics*, *35*(2), 190–194. https://doi.org/10.1038/ng1244

Passot, F. M., Calderon, V., Fichant, G., Lane, D., & Pasta, F. (2012). Centromere binding and evolution of chromosomal partition systems in the Burkholderiales. *Journal of Bacteriology*, *194*(13), 3426–3436. https://doi.org/10.1128/JB.00041-12

Pengelly, A. R. (2013). A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb. *Journal of Chemical Information and Modeling*, *9*(3), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., … Rubin, E. M. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature*, *444*(7118), 499–502. https://doi.org/10.1038/nature05295

Perry, M. W., Boettiger, A. N., Bothma, J. P., & Levine, M. (2010). Shadow enhancers foster

robustness of drosophila gastrulation. *Current Biology*, *20*(17), 1562–1567. https://doi.org/10.1016/j.cub.2010.07.043

Perry, M. W., Boettiger, A. N., & Levine, M. (2011). Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13570–13575. https://doi.org/10.1073/pnas.1109873108

Petesch, S. J., & Lis, J. T. (2008). Rapid, Transcription-Independent Loss of Nucleosomes over a Large Chromatin Domain at Hsp70 Loci. *Cell*, *134*(1), 74–84. https://doi.org/10.1016/j.cell.2008.05.029

Petrovic, J., Zhou, Y., Fasolino, M., Goldman, N., Schwartz, G. W., Mumbach, M. R., … Faryabi, R. B. (2019). Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. *Molecular Cell*, *73*(6), 1174-1190.e12. https://doi.org/10.1016/j.molcel.2019.01.006

Phair, R. D., Scaffidi, P., Elbi, C., Vecerová, J., Dey, A., Ozato, K., … Misteli, T. (2004). Global Nature of Dynamic Protein-Chromatin Interactions In Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins. *Molecular and Cellular Biology*, *24*(14), 6393–6402. https://doi.org/10.1128/mcb.24.14.6393-6402.2004

Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., … Corces, V. G. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, *153*(6), 1281–1295. https://doi.org/10.1016/j.cell.2013.04.053

Pichon, X., Lagha, M., Mueller, F., & Bertrand, E. (2018). A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges. *Molecular Cell*, *71*(3), 468–480. https://doi.org/10.1016/j.molcel.2018.07.022

Plaschka, C., Larivière, L., Wenzeck, L., Seizl, M., Hemann, M., Tegunov, D., … Cramer, P. (2015). Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*, *518*(7539), 376–380. https://doi.org/10.1038/nature14229

Platania, A., & Sexton, T. (2020). Chapter 4 - Chromatin architecture and topology in pluripotent stem cells. In E. Meshorer & G. Testa (Eds.), *Stem Cell Epigenetics* (pp. 93–113). https://doi.org/https://doi.org/10.1016/B978-0-12-814085-7.00004-0

Portz, B., Lu, F., Gibbs, E. B., Mayfield, J. E., Rachel Mehaffey, M., Zhang, Y. J., … Gilmour, D. S. (2017). Structural heterogeneity in the intrinsically disordered RNA

polymerase II C-terminal domain. *Nature Communications*, *8*(May), 1–12. https://doi.org/10.1038/ncomms15231

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C. A., Schneider, R., & Bickmore, W. A. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. *Nature Genetics*, *48*(6), 681–686. https://doi.org/10.1038/ng.3550

Proudfoot, N. J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science*, *324*(5923), 89–91. https://doi.org/10.1126/science

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, *470*(7333), 279–285. https://doi.org/10.1038/nature09692

Rafael Galupa, A., ge Pierre Nora, E., Worsley-Hunt, R., Ohler, U., Giorgetti, L., Heard Correspondence, E., … Heard, E. (2020). A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary Molecular Cell Article A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Molecular Cell*, *77*, 1–16. https://doi.org/10.1016/j.molcel.2019.10.030

Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., … Young, R. A. (2010). C-Myc regulates transcriptional pause release. *Cell*, *141*(3), 432–445. https://doi.org/10.1016/j.cell.2010.03.030

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biology*, *4*(10), 1707–1719. https://doi.org/10.1371/journal.pbio.0040309

Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banerjee, B., … Sherwood, R. I. (2016). High-throughput mapping of regulatory DNA. *Nature Biotechnology*, *34*(2), 167–174. https://doi.org/10.1038/nbt.3468

Rao, S. S. P., Huang, S. C., Glenn St Hilaire, B., Engreitz, J. M., Perez, E. M., Kieffer-Kwon, K. R., … Aiden, E. L. (2017). Cohesin Loss Eliminates All Loop Domains. *Cell*, *171*(2), 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J.

T., … Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Raser, J. M., & O'Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science*, *304*(5678), 1811–1814. https://doi.org/10.1126/science.1098641

Redolfi, J., Zhan, Y., Valdes-Quezada, C., Kryzhanovska, M., Guerreiro, I., Iesmantavicius, V., … Giorgetti, L. (2019). DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nature Structural and Molecular Biology*, *26*(6), 471–480. https://doi.org/10.1038/s41594-019-0231-0

Rhee, H. S., & Pugh, B. F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*, *147*(6), 1408–1419. https://doi.org/10.1016/j.cell.2011.11.013

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–329. https://doi.org/10.1038/nature14248

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., … Jones, S. (2007). Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods*, *4*(8), 651–657. https://doi.org/10.1038/nmeth1068

Robin, P. (1994). A fall of the base of the tongue considered as a new cause of nasopharyngeal respiratory impairment: Pierre Robin sequence, a translation. 1923. *Plastic and Reconstructive Surgery*, *93*(6), 1301—1303. Retrieved from http://europepmc.org/abstract/MED/8171154

Robinett, C. C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., & Belmont, A. S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *Journal of Cell Biology*, *135*(6 II), 1685–1700. https://doi.org/10.1083/jcb.135.6.1685

Roh, T. Y., Cuddapah, S., & Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes and Development*, *19*(5), 542–552. https://doi.org/10.1101/gad.1272505

Rougvie, A. E., & Lis, J. T. (1988). The RNA polymerase II molecule at the 5′ end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, *54*(6), 795–804. https://doi.org/10.1016/S0092-8674(88)91087-2

Saad, H., Gallardo, F., Dalvai, M., Tanguy-le-gac, N., Lane, D., & Bystricky, K. (2014). *DNA Dynamics during Early Double-Strand Break Processing Revealed by Non-Intrusive Imaging of Living*. *10*(3). https://doi.org/10.1371/journal.pgen.1004187

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science*, *361*(6400), eaar3958. https://doi.org/10.1126/science.aar3958

Sagai, T., Amano, T., Tamura, M., Mizushina, Y., Sumiyama, K., & Shiroishi, T. (2009). A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development*, *136*(10), 1665–1674. https://doi.org/10.1242/dev.032714

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., & Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development*, *132*(4), 797–803. https://doi.org/10.1242/dev.01613

Sanchez, A., Cattoni, D. I., Walter, J. C., Rech, J., Parmeggiani, A., Nollmann, M., & Bouet, J. Y. (2015). Stochastic Self-Assembly of ParB Proteins Builds the Bacterial DNA Segregation Apparatus. *Cell Systems*, *1*(2), 163–173. https://doi.org/10.1016/j.cels.2015.07.013

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

Scalettar, B. A., Hearst, J. E., & Klein, M. P. (1989). FRAP and FCS Studies of Self-Diffusion and Mutual Diffusion in Entangled DNA Solutions. *Macromolecules*, *22*(12), 4550–4559. https://doi.org/10.1021/ma00202a030

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

Schoenfelder, S., Furlan-magaril, M., Mifsud, B., Tavares-cadete, F., Sugar, R., Javierre, B., … Fraser, P. (2015). *The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements Stefan*. 582–597. https://doi.org/10.1101/gr.185272.114.Freely

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., … Fraser, P. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*, *42*(1), 53–61. https://doi.org/10.1038/ng.496

Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F. E., Zimmer, C., … Darzacq, X. (2014). Transcription factors modulate c-Fos transcriptional bursts. *Cell Reports*, *8*(1), 75–83. https://doi.org/10.1016/j.celrep.2014.05.053

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., … 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

Shaban, H. A., Barth, R., & Bystricky, K. (2018). Formation of correlated chromatin domains at nanoscale dynamic resolution during transcription. *Nucleic Acids Research*, *46*(13), e77. https://doi.org/10.1093/nar/gky269

Shaban, H. A., Barth, R., Recoules, L., & Bystricky, K. (2020). Hi-D: Nanoscale mapping of nuclear dynamics in single living cells. *Genome Biology*, *21*(1), 1–21. https://doi.org/10.1186/s13059-020-02002-6

Shaban, H. A., & Seeber, A. (2020). Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Research*, *48*(7), 3423–3434. https://doi.org/10.1093/nar/gkaa135

Shang, Y., Hu, X., Direnzo, J., Lazar, M. A., & Brown, M. (2000). Cofactor Dynamics and Sufficiency in Estrogen Receptor-Regulated Transcription ciated factor, pCAF (Blanco et al., 1998). Coactivators such as CBP, p300, pCAF, and possibly SRC1 and AIB1 possess intrinsic histone acetyltransferase (HAT) activi. *Cell*, *103*, 843–852.

Sharova, L. V., Sharov, A. A., Nedorezov, T., Piao, Y., Shaik, N., & Ko, M. S. H. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Research*, *16*(1), 45–58. https://doi.org/10.1093/dnares/dsn030

Shen, Y., Yue, F., Mc Cleary, D. F., Ye, Z., Edsall, L., Kuan, S., … Lobanenkov, V. V. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, *488*(7409), 116–120. https://doi.org/10.1038/nature11243

Shin, H. Y., Willi, M., Yoo, K. H., Zeng, X., Wang, C., Metser, G., & Hennighausen, L. (2016). Hierarchy within the mammary STAT5-driven Wap super-enhancer. *Nature Genetics*, *48*(8), 904–911. https://doi.org/10.1038/ng.3606

Shin, Y., Berry, J., Pannucci, N., Haataja, M. P., Toettcher, J. E., & Brangwynne, C. P. (2017). Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell*, *168*(1–2), 159-171.e14. https://doi.org/10.1016/j.cell.2016.11.054

Shin, Y., & Brangwynne, C. P. (2017). Liquid phase condensation in cell physiology and

disease. *Science*, *357*(6357). https://doi.org/10.1126/science.aaf4382

Shin, Y., Chang, Y. C., Lee, D. S. W., Berry, J., Sanders, D. W., Ronceray, P., … Brangwynne, C. P. (2018). Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell*, *175*(6), 1481-1491.e13. https://doi.org/10.1016/j.cell.2018.10.057

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

Shore, D., Stillman, D. J., Brand, A. H., & Nasmyth, K. A. (1987). Identification of silencer binding proteins from yeast: possible roles in SIR control and DNA replication. *The EMBO Journal*, *6*(2), 461–467. https://doi.org/10.1002/j.1460-2075.1987.tb04776.x

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., De Wit, E., … 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

Simpson, R. T. (1978). Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry*, *17*(25), 5524–5531. https://doi.org/10.1021/bi00618a030

Slattery, M. (2014). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Bone*, *23*(1), 1–7. https://doi.org/10.1038/jid.2014.371

Smith, E., & Shilatifard, A. (2014). Enhancer biology and enhanceropathies. *Nature Structural and Molecular Biology*, *21*(3), 210–219. https://doi.org/10.1038/nsmb.2784

Splinter, E., de Wit, E., Nora, E. P., Klous, P., van de Werken, H. J. G., Zhu, Y., … de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes and Development*, *25*(13), 1371–1383. https://doi.org/10.1101/gad.633311

Stenoien, D. L., Patel, K., Mancini, M. G., Dutertre, M., Smith, C. L., O'Malley, B. W., & Mancini, M. A. (2001). FRAP reveals that mobility of oestrogen receptor-α is ligand- and proteasome-dependent. *Nature Cell Biology*, *3*(1), 15–23. https://doi.org/10.1038/35050515

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C., & Doudna, J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, *507*(7490), 62–67. https://doi.org/10.1038/nature13011

Stevens, T. J., Lando, D., Basu, S., Atkinson, L. P., Cao, Y., Lee, S. F., … Laue, E. D. (2017). 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, *544*(7648), 59–64. https://doi.org/10.1038/nature21429

Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B., & Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell*, *182*(6), 1641-1659.e26. https://doi.org/10.1016/j.cell.2020.07.032

Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., & Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science*, *332*(6028), 472–474. https://doi.org/10.1126/science.1198817

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., … Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell*, *165*(3), 593–605. https://doi.org/10.1016/j.cell.2016.02.067

Szabo, Q., Jost, D., Chang, J. M., Cattoni, D. I., Papadopoulos, G. L., Bonev, B., … Cavalli, G. (2018). TADs are 3D structural units of higher-order chromosome organization in Drosophila. *Science Advances*, *4*(2), 1–13. https://doi.org/10.1126/sciadv.aar8082

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. https://doi.org/10.1016/j.cell.2006.07.024

Thakore, P. I., Song, L., Safi, A., Shivakumar, K., Kabadi, A. M., Reddy, T. E., … Gersbach, C. A. (2015). Highly Specific Epigenome Editing by CRISPR/Cas9 Repressors for Silencing of Distal Regulatory Elements. *Nature Methods*, *12*(12), 1143–1149. https://doi.org/10.1038/nmeth.3630.Highly

Thomas A. Johnson, & Cem Elbi, Bhavin S. Parekh,§ Gordon L. Hager, and S. J. (2008). Chromatin Remodeling Complexes Interact Dynamically with a Glucocorticoid Receptor–regulated Promoter. *Molecular Biology of the Cell*, *20*(August), 2673–2683. https://doi.org/10.1091/mbc.E08

Thomas, H. F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., … Buecker, C. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Molecular Cell*, *81*(0), 1–14. https://doi.org/10.1016/j.molcel.2020.12.047

Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., … Stamatoyannopoulos, J. A. (2012). The accessible chromatin landscape of the human genome. *Nature*, *489*(7414), 75–82. https://doi.org/10.1038/nature11232

Tie, F., Banerjee, R., Stratton, C. A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., … Harte, P. J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development*, *136*(18), 3131–3141. https://doi.org/10.1242/dev.037127

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. *Molecular Cell*, *10*(6), 1453–1465. https://doi.org/10.1016/S1097-2765(02)00781-5

Tortora, M. M., Salari, H., & Jost, D. (2020). Chromosome dynamics during interphase: a biophysical perspective. *Current Opinion in Genetics and Development*, *61*, 37–43. https://doi.org/10.1016/j.gde.2020.03.001

Tsai, A., Muthusamy, A. K., Lavish, L. D., Singer, R. H., Stern, D. L., & Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. *BioRxiv*, 1–18. https://doi.org/10.1101/128280

Tuan, D., Kong, S., & Hu, K. (1992). Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(23), 11219–11223. https://doi.org/10.1073/pnas.89.23.11219

Tuan, D. Y. H., Solomon, W. B., London, I. M., & Lee, D. P. (1989). An erythroid-specific, developmental-stage-independent enhancer far upstream of the human "β-like globin" genes. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(8), 2554–2558. https://doi.org/10.1073/pnas.86.8.2554

Tunnacliffe, E., & Chubb, J. R. (2020). What Is a Transcriptional Burst? *Trends in Genetics*, *36*(4), 288–297. https://doi.org/10.1016/j.tig.2020.01.003

Ulianov, S. V, Velichko, A. K., Magnitov, M. D., Luzhin, A. V., Golov, A. K., Ovsyannikova, N., … Razin, S. V. (2021). Suppression of liquid–liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells. *Nucleic Acids Research*, 1–18. https://doi.org/10.1093/nar/gkab249

Uslu, V. V., Petretich, M., Ruf, S., Langenfeld, K., Fonseca, N. A., Marioni, J. C., & Spitz, F. (2014). Long-range enhancers regulating Myc expression are required for normal facial morphogenesis. *Nature Genetics*, *46*(7), 753–758. https://doi.org/10.1038/ng.2971

Vahedi, G., Kanno, Y., Furumoto, Y., Jiang, K., Parker, S. C. J., Erdos, M. R., … O'Shea, J. J. (2015). Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*, *520*(7548), 558–562. https://doi.org/10.1038/nature14154

Van De Werken, H. J. G., Landan, G., Holwerda, S. J. B., Hoichman, M., Klous, P., Chachik,

R., … De Laat, W. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nature Methods*, *9*(10), 969–972. https://doi.org/10.1038/nmeth.2173

Vernimmen, D., & Bickmore, W. A. (2015). The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends in Genetics*, *31*(12), 696–708. https://doi.org/10.1016/j.tig.2015.10.004

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., … Pennacchio, L. A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature*, *457*(7231), 854–858. https://doi.org/10.1038/nature07730

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser - A database of tissue-specific human enhancers. *Nucleic Acids Research*, *35*(SUPPL. 1), 88–92. https://doi.org/10.1093/nar/gkl822

Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. https://doi.org/10.1038/nature08451

Visel, A., Taher, L., Girgis, H., May, D., Golonzhka, O., Hoch, R. V., … Rubenstein, J. L. R. (2013). A high-resolution enhancer atlas of the developing telencephalon. *Cell*, *152*(4), 895–908. https://doi.org/10.1016/j.cell.2012.12.041

Vispé, S., DeVries, L., Créancier, L., Besse, J., Bréand, S., Hobson, D. J., … Bailly, C. (2009). Triptolide is an inhibitor of RNA polymerase I and II-dependent transcription leading predominantly to down-regulation of short-lived mRNA. *Molecular Cancer Therapeutics*, *8*(10), 2780–2790. https://doi.org/10.1158/1535-7163.MCT-09-0549

Vos, S. M., Farnung, L., Urlaub, H., & Cramer, P. (2018). Structure of paused transcription complex Pol II–DSIF–NELF. *Nature*, *560*(7720), 601–606. https://doi.org/10.1038/s41586-018-0442-2

Wada, T., Takagi, T., Yamaguchi, Y., Ferdous, A., Imai, T., Hirose, S., … Handa, H. (1998). DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes and Development*, *12*(3), 343–356. https://doi.org/10.1101/gad.12.3.343

Wada, T., Takagi, T., Yamaguchi, Y., Watanabe, D., & Handa, H. (1998). Evidence that P-TEFb alleviates the negative effect of DSIF on RNA polymerase II-dependent transcription in vitro. *EMBO Journal*, *17*(24), 7395–7403. https://doi.org/10.1093/emboj/17.24.7395

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Held, M., Zimmer, J., … Scherer, G. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and

around the SRY-related gene SOX9. *Cell*, *79*(6), 1111–1120. https://doi.org/10.1016/0092-8674(94)90041-8

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., … Fu, X. D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, *474*(7351), 390–397. https://doi.org/10.1038/nature10006

Wang, H., Nakamura, M., Abbott, T. R., Zhao, D., Luo, K., Yu, C., … Russa, M. La. (2019). *CRISPR-mediated live imaging of genome editing and transcription*. (September), 2–6.

Wang, Q., Carroll, J. S., & Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell*, *19*(5), 631–642. https://doi.org/10.1016/j.molcel.2005.07.018

Wang, S., Su, J., Beliveau, B. J., Bintu, B., Moffitt, J. R., & Wu, C. (2016). *Spatial organization ofchromatin domains and compartments in single chromosomes*. *544*(2011), 2955–2960.

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., … 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

Williams, L. H., Fromm, G., Gokey, N. G., Henriques, T., Muse, G. W., Burkholder, A., … Adelman, K. (2015). Pausing of RNA Polymerase II Regulates Mammalian Developmental Potential through Control of Signaling Networks. *Molecular Cell*, *58*(2), 311–322. https://doi.org/10.1016/j.molcel.2015.02.003

Williamson, I., Kane, L., Devenney, P. S., Flyamer, I. M., Anderson, E., Kilanowski, F., … Lettice, L. A. (2019). Developmentally regulated Shh expression is robust to TAD perturbations. *Development (Cambridge)*, *146*(19). https://doi.org/10.1242/dev.179523

Williamson, I., Lettic, L. A., Hill, R. E., & Bickmore, W. A. (2016). Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development (Cambridge)*, *143*(16), 2994–3001. https://doi.org/10.1242/dev.139188

Wissink, E. M., Vihervaara, A., Tippens, N. D., & Lis, J. T. (2019). Nascent RNA analyses: tracking transcription and its regulation. *Nature Reviews Genetics*, *20*(12), 705–723. https://doi.org/10.1038/s41576-019-0159-6

Wit, E. De, & Laat, W. De. (2012). A decade of 3C technologies-insights into nuclear organization. *Genes & Development*, 11–24. https://doi.org/10.1101/gad.179804.111.GENES

Wollman, A. J. M., Shashkova, S., Hedlund, E. G., Friemann, R., Hohmann, S., & Leake, M. C. (2017). Transcription factor clusters regulate genes in eukaryotic cells. *ELife*, *6*, 1–36. https://doi.org/10.7554/eLife.27451

Workman, J. L., & Kingston, R. E. (1992). Nucleosome core displacement in vitro via a metastable transcription factor-nucleosome complex. *Science*, *258*(5089), 1780–1784. https://doi.org/10.1126/science.1465613

Wu, C. (1980). The 5' ends of drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature*, *286*(5776), 854–860. https://doi.org/10.1038/286854a0

Wu, C. H., Yamaguchi, Y., Benjamin, L. R., Horvat-Gordon, M., Washinsky, J., Enerly, E., … Gilmour, D. (2003). NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes and Development*, *17*(11), 1402–1414. https://doi.org/10.1101/gad.1091403

Yamada, S., Whitney, P. H., Huang, S. K., Eck, E. C., Garcia, H. G., & Rushlow, C. A. (2019). The Drosophila Pioneer Factor Zelda Modulates the Nuclear Microenvironment of a Dorsal Target Enhancer to Potentiate Transcriptional Output. *Current Biology*, *29*(8), 1387-1393.e5. https://doi.org/10.1016/j.cub.2019.03.019

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., … Handa, H. (1999). NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell*, *97*(1), 41–51. https://doi.org/10.1016/S0092-8674(00)80713-8

Yáñez-Cuna, J. O., Dinh, H. Q., Kvon, E. Z., Shlyueva, D., & Stark, A. (2012). Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Research*, *22*(10), 2018–2030. https://doi.org/10.1101/gr.132811.111

Yang, L.-Z., Wang, Y., Li, S.-Q., Yao, R.-W., Luan, P.-F., Wu, H., … Chen, L.-L. (2019). Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Molecular Cell*, 981–997. https://doi.org/10.1016/j.molcel.2019.10.024

Yao, Y., Minor, P. J., Zhao, Y. T., Jeong, Y., Pani, A. M., King, A. N., … Epstein, D. J. (2016). Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nature Genetics*, *48*(5), 575–580. https://doi.org/10.1038/ng.3542

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., … Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature*, *453*(7194), 519–523. https://doi.org/10.1038/nature06968

Yudkovsky, N., Ranish, J. A., & Hahn, S. (2000). A transcription reinitiation intermediate

that is stabilized by activator. *Nature*, *408*(6809), 225–229. https://doi.org/10.1038/35041603

Yunger, S., Rosenfeld, L., Garini, Y., & Shav-Tal, Y. (2010). Single-allele analysis of transcription kinetics in living mammalian cells. *Nature Methods*, *7*(8), 631–633. https://doi.org/10.1038/nmeth.1482

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: Establishing competence for gene expression. *Genes and Development*, *25*(21), 2227–2241. https://doi.org/10.1101/gad.176826.111

Zeitlinger, J., Stark, A., Kellis, M., Hong, J. W., Nechaev, S., Adelman, K., … Young, R. A. (2007). RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo. *Nature Genetics*, *39*(12), 1512–1516. https://doi.org/10.1038/ng.2007.26

Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, *21*(8), 1273–1283. https://doi.org/10.1101/gr.122382.111

Zhang, H., Roberts, D. N., & Cairns, B. R. (2005). Genome-wide dynamics of Htz1, a histone H2A variant that poises repressed/basal promoters for activation through histone loss. *Cell*, *123*(2), 219–231. https://doi.org/10.1016/j.cell.2005.08.036

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., … 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

Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M., & Furlong, E. E. M. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, *462*(7269), 65–70. https://doi.org/10.1038/nature08531

Appendix 2

frontiers in Genetics

TECHNOLOGY AND CODE
published: 21 January 2020
doi: 10.3389/fgene.2019.01372

4See: A Flexible Browser to Explore 4C Data

Yousra Ben Zouari[1,2,3,4], *Angeliki Platania*[1,2,3,4], *Anne M. Molitor*[1,2,3,4] *and Tom Sexton*[1,2,3,4]*

[1] Institute of Genetics and Molecular and Cellular Biology (IGBMC), Illkirch, France, [2] CNRS UMR7104, Illkirch, France, [3] INSERM U1258, Illkirch, France, [4] University of Strasbourg, Illkirch, France

It is established that transcription of many metazoan genes is regulated by distal regulatory sequences beyond the promoter. Enhancers have been identified at up to megabase distances from their regulated genes, and/or proximal to or within the introns of unregulated genes. The unambiguous identification of the target genes of newly identified regulatory elements can thus be challenging. Well-studied enhancers have been found to come into direct physical proximity with regulated genes, presumably by the formation of chromatin loops. Chromosome conformation capture (3C) derivatives that assess the frequency of proximity between different genetic elements is thus a popular method for exploring gene regulation by distal regulatory elements. For studies of chromatin loops and promoter-enhancer communication, 4C (circular chromosome conformation capture) is one of the methods of choice, optimizing cost (required sequencing depth), throughput, and resolution. For ease of visual inspection of 4C data we present 4See, a versatile and user-friendly browser. 4See allows 4C profiles from the same bait to be flexibly plotted together, allowing biological replicates to either be compared, or pooled for comparisons between different cell types or experimental conditions. 4C profiles can be integrated with gene tracks, linear epigenomic profiles, and annotated regions of interest, such as called significant interactions, allowing rapid data exploration with limited computational resources or bioinformatics expertise.

Keywords: 4C, epigenomics, browser, chromatin loops, quantile normalization, biological replicates

OPEN ACCESS

Edited by:
Karim Mekhail,
University of Toronto, Canada

Reviewed by:
Mikhail Spivakov,
Babraham Institute (BBSRC), United Kingdom
Alexander Kouzmenko,
Tokiwa Foundation, Japan

***Correspondence:**
Tom Sexton
sexton@igbmc.fr

Specialty section:
This article was submitted to Epigenomics and Epigenetics, a section of the journal Frontiers in Genetics

Received: *23 September 2019*
Accepted: *16 December 2019*
Published: *21 January 2020*

Citation:
Ben Zouari Y, Platania A, Molitor AM and Sexton T (2020) 4See: A Flexible Browser to Explore 4C Data. Front. Genet. 10:1372. doi: 10.3389/fgene.2019.01372

INTRODUCTION

Since early transgenic studies it has been clear that promoter sequences are insufficient to regulate the spatiotemporal expression patterns of many developmental genes. "Remote control" is additionally conferred by distal activating sequences, termed enhancers, which have been intensively studied over the last years (Schoenfelder and Fraser, 2019). Genome-wide profiling of histone modifications and protein binding sites by ChIP-seq have uncovered a general chromatin signature of enhancer regions: DNase-hypersensitive, bound by the transcriptional coactivator p300, and marked by the monomethylation of lysine-4 of histone H3 (H3K4me1) (Heintzman et al., 2009). Follow-on studies refined these findings further by identifying chromatin features that were characteristic of different enhancer properties. For example, the strongest-acting enhancers are also accompanied by acetylation of lysine-27 of histone H3 (H3K27ac) (Creyghton et al., 2010;

Rada-Iglesias et al., 2011) and/or acetylation on globular histone domains (Pradeepa et al., 2016), recruit RNA polymerase II, and general transcriptional machinery (Koch et al., 2011), and are even transcriptionally active, producing non-coding RNA (eRNAs) (Kim et al., 2010). Enhancers lacking these extra features, and sometimes even encompassing repressive marks, such as H3K27 trimethylation (H3K27me3), are proposed to be "poised" enhancers, which may become activated at later developmental stages. Interestingly, the chromatin states at enhancer sequences vary much more across cell types than those of gene promoters (Roadmap Epigenomics et al., 2015), suggesting that much of the regulatory potential is epigenetically carried by enhancers. However despite advances in identifying enhancers genome-wide, both through epigenomic profiling and high-throughput reporter assays (Arnold et al., 2013; Roadmap Epigenomics et al., 2015), unambiguous identification of their target genes is still a major challenge. Important developmental enhancers have been found at megabase distances from target genes, and/or within the introns of unregulated genes (Lettice et al., 2003; Amano et al., 2009; Herranz et al., 2014); previous studies estimate that up to ~90% of enhancers may indeed skip the closest genes on the linear chromosome fiber (Sanyal et al., 2012; Schoenfelder et al., 2015).

Since the advent of the chromosome conformation capture method (3C) (Dekker et al., 2002) and its variants to measure relative spatial proximity of pairwise genomic regions, many enhancers have been found to physically interact with their target genes, often with "looping out" of the intervening chromatin (Palstra et al., 2003); the resultant "active chromatin hub" has been proposed to provide the permissive regulatory environment for transcription initiation, although the exact mechanism remains unclear. In many studied cases, looping is concomitant with transcriptional induction, whereas in others, the loop is pre-formed to poise the gene for subsequent activation (Schoenfelder and Fraser, 2019). Recent reports using microscopy methods have also been made of enhancers and promoters being well separated on gene activation (Alexander et al., 2019; Benabdallah et al., 2019), although enhancer-promoter interactions were previously reported in the studied loci, raising questions as to whether interactions may completely precede transcription and/or be very transient events. In any case, physical proximity measured by 3C-based methods is becoming a popular means of ascribing target genes to otherwise cryptic distal regulatory elements, or of identifying novel candidate regulatory regions of specific genes of interest. For example, intergenic sequence variants associated with diseases have been better characterized once their target genes were identified by 3C-based approaches (Herranz et al., 2014; Schoenfelder and Fraser, 2019).

With the advent of next-generation sequencing, several higher throughput variants of 3C have been developed to obtain genome-wide chromatin interaction maps. Hi-C is an "all-to-all" method, systematically assessing all pairwise chromatin contacts (Lieberman-Aiden et al., 2009). However, due to the great complexity of the sequenced material, calling specific looping interactions requires prohibitively expensive sequencing depth (Rao et al., 2014; Bonev et al., 2017), and Hi-C loop calling algorithms have been demonstrated to not be very robust (Forcato et al., 2017). A recent modification, Capture Hi-C, incorporates capture with a pool of thousands of oligonucleotides, allowing the complexity of sequenced Hi-C material to be reduced sufficiently to assess the chromatin looping interactions with all promoters (Hughes et al., 2014; Sahlen et al., 2015; Schoenfelder et al., 2015). However, capture libraries can be expensive, and their design still represents a trade-off between coverage of assessed promoters and resolution of the identified loops. For the highest resolution profiling of smaller numbers of candidate regions, the method of choice is the "one-to-all" 4C (circular chromosome conformation capture), which assesses all the chromatin interactions with one specific bait of interest (Simonis et al., 2006; van de Werken et al., 2012) (**Figure 1A**). In brief, nuclei are fixed in their native topologies with formaldehyde, digested with a restriction enzyme and re-ligated, as for 3C, such that chimeric DNA sequences are generated between restriction fragments which may be unlinked on the linear chromosome fiber but are physically proximal at the time of fixation. The purified DNA is then circularized by digestion with a secondary restriction enzyme and re-ligation under dilute conditions, allowing an inverse PCR strategy to amplify all the chimeric DNA linked to a specific bait restriction fragment of interest. The much reduced complexity of a 4C library, compared to that of Hi-C, means that promoter interactomes can be reliably profiled with just a few million sequence reads, and ~20 baits can readily be multiplexed into a sequencing run, making it a much more cost-effective method (van de Werken et al., 2012). The major limitations of 4C are the relatively small throughput in baits that can be assessed at a time, and that the direct sequencing of PCR products confounds results with large numbers of PCR duplicates that cannot be distinguished from counts of true 3C ligation events. However, *in silico* approaches can minimize the impact of PCR duplicates (de Wit et al., 2015), and "unique molecular identifier" variants of 4C have also been developed (Schwartzman et al., 2016).

Due to the growing popularity of 4C experiments, several algorithms have been developed to call significant interactions (van de Werken et al., 2012; Thongjuea et al., 2013; Williams et al., 2014; Klein et al., 2015; Raviram et al., 2016; Geeven et al., 2018); recent benchmarking shows that all methods work well on simulated data, but no single method is optimum for all experimental setups (Walter et al., 2019). However, whereas most of these methods has an in-built tool to plot the static results after data processing, a simple, flexible browser allowing a user to rapidly visualize their 4C results is currently lacking (see **Figure 2** and summary of the different plotting options currently available in **Table 1**). Moreover, while some methods allowed raw and/or smoothed 4C data to be exported as files that can be opened and visualized alongside epigenomic profiles on genome browsers, they offered no flexibility in plotting the epigenomic profiles directly alongside the 4C plot while different smoothing or peak calling parameters are being trialed. We recently developed ChiCMaxima, a suite of tools to analyze Capture Hi-C data, which includes a GUI (graphical user interface) to flexibly visualize data sets alongside gene annotations and epigenomic profiles (Ben Zouari et al., 2019). Here we report

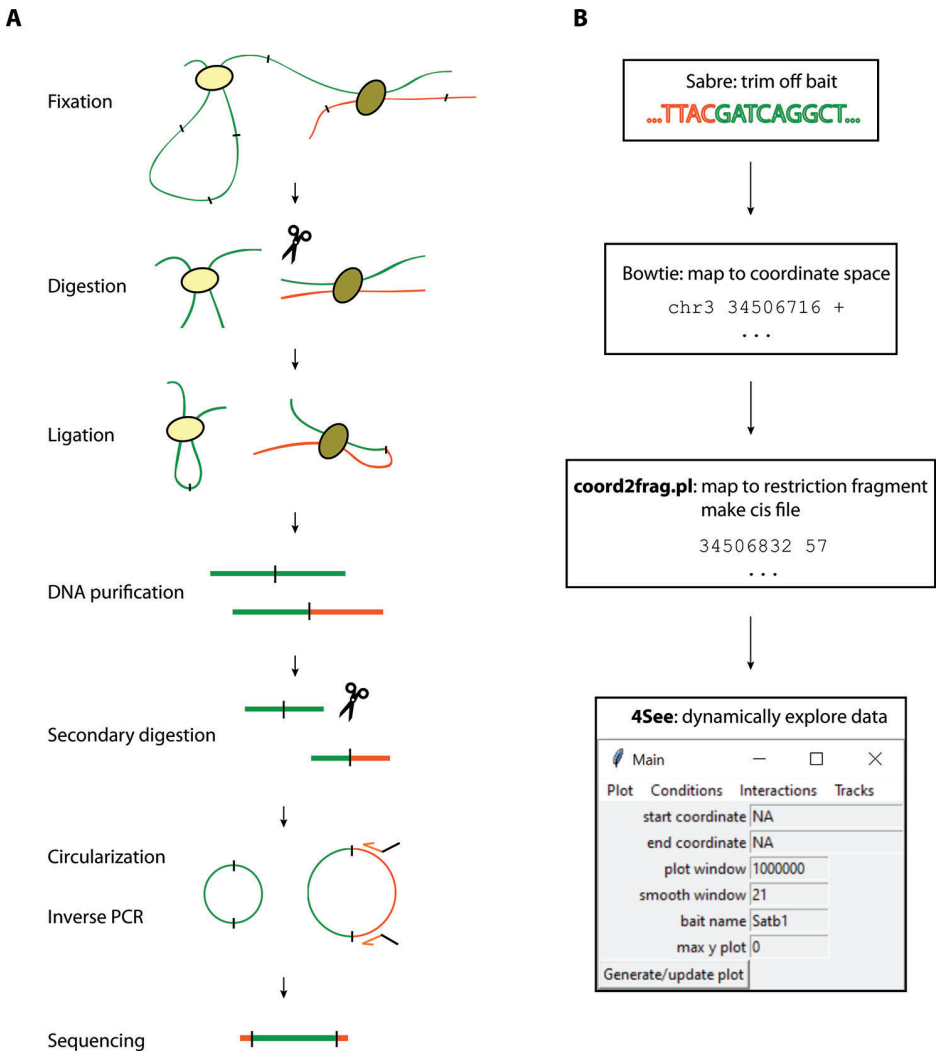


FIGURE 1 | Overview of the 4C method and analysis. **(A)** The 4C method entails chromatin fixation, restriction digestion and re-ligation to generate hybrid sequences between fragments that were physically proximal during fixation. The DNA is purified, digested with a secondary restriction enzyme and re-ligated under dilute conditions to generate DNA circles. Chimeric products linked to a specific bait fragment of interest (orange) are amplified by inverse PCR with bait-specific primers (orange arrows) flanked by Illumina sequencing adapters (black overhangs). The PCR products are then directly loaded onto Illumina flow-cells for high-throughput sequencing. **(B)** Pre-processing steps before 4See; tools denoted in bold accompany this manuscript. The fastq sequences are first trimmed to remove bait restriction fragment sequence (orange), leaving just the prey DNA sequence (green) for mapping to the reference genome with Bowtie. The mapped genomic coordinates are converted to restriction fragment space by a custom perl script, which counts the total number of reads mapping to each fragment on the same chromosome as the bait. This "cis" file can then be directly input into 4See.

4See, the adaptation of ChiCMaxima tools for the user-friendly integrative exploration of 4C data sets. 4See provides flexibility to compare different replicates side by side, or to average them together when comparing experimental conditions, and to visualize 4C profiles at different smoothing window sizes, necessary to identify putative interactions at different distances from the bait, without the need to reload or re-process the initial data. 4See utilizes quantile normalization to allow different plotted profiles to be fairly compared during the visualization. 4See also allows 4C profiles to be easily plotted alongside gene annotations and linear epigenomic tracks, as well as for specific regions (e.g. interactions called by other algorithms) to be highlighted. We anticipate that 4See will be a useful tool to the community for quick and easy exploration of 4C data, particularly when used in conjunction with existing interaction calling tools.

MATERIALS AND METHODS

Pre-Processing

4C

J1 mouse embryonic stem (ES) cells were grown on gamma-irradiated mouse embryonic fibroblast cells under standard

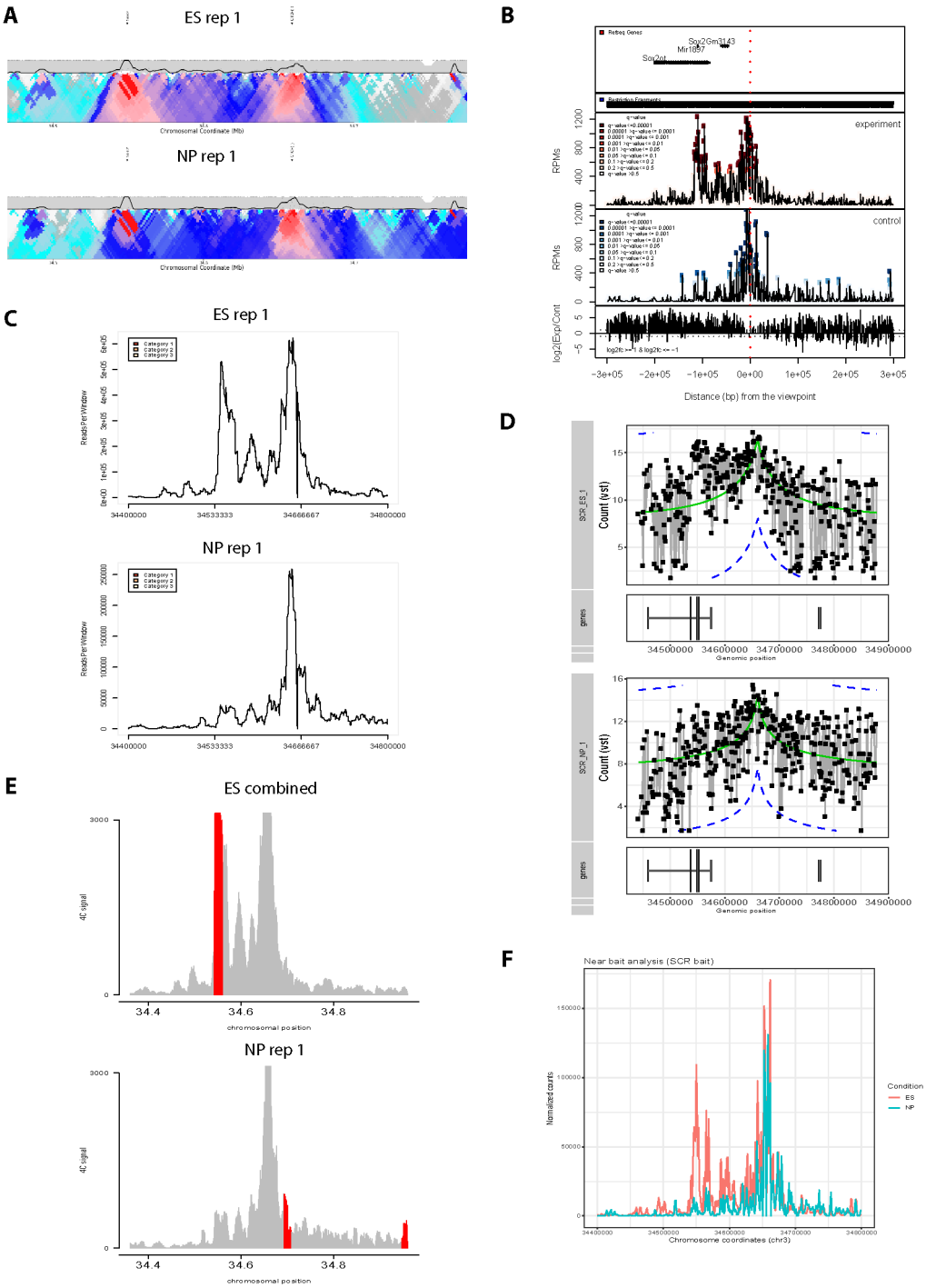


FIGURE 2 | Overview of graphing options from existing 4C analysis methods. All methods have been applied to two ES replicates and one NPC 4C data set for the *Sox2* SCR bait (see also **Figure 3**). **(A)** 4Cseqpipe (van de Werken et al., 2012) results shown independently for one ES replicate and the NPC data set. Running median scores are plotted as a line graph (5 kb resolution), with domainograms plotted underneath as a heat map for median scores at steadily increasing window sizes. Positions of the CTCF site within the SCR and the *Sox2* gene are indicated by arrows. Note that the independent normalization means that the SCR-*Sox2* interaction differences between the two cell types is not evident, compared to other methods. **(B)** r3Cseq (Thongjuea et al., 2013) results for the combined three data sets, showing panels, from top to bottom: positions of Refseq genes; restriction fragment coverage; averaged profile for the "experiment" (ES) condition, with called interactions at different confidence levels highlighted; profile for the "control" (NPC) condition, with called interactions at different confidence levels highlighted; plot of the log2-ratio of experiment vs control 4C signal. The position of the SCR bait is indicated by a red dashed line. Note that r3Cseq appears to call a very large number of

(*continued*)

FIGURE 2 | interactions. **(C)** fourSig (Williams et al., 2014) results shown independently for one ES and one NPC replicate, showing smoothed 4C plots (21-fragment windows). Called interactions ("Categories" 1, 2 or 3, for different confidence levels) would be highlighted on the plot, but none were called by fourSig. **(D)** FourCSeq (Klein et al., 2015) results shown independently for one ES and one NPC replicate, showing normalized and processed 4C signal plots (gray line graphs and black points), alongside the positions of known genes. The green line indicates the centralized 4C value, and the dashed blue lines indicate the threshold values for a z-score difference > 2. Significant interactions would be highlighted, but were not detected by FourCSeq. Note that differences between the two cell types is not so evident, compared to most other methods. **(E)** peakC (Geeven et al., 2018) results shown for the combined analysis of the two ES replicates, independently of the NP replicate, giving smoothed 4C plots (21-fragment windows) as histograms. The red regions indicate called interactions. **(F)** 4C-ker (Raviram et al., 2016) results shown for the combined analysis, with line plots of combined ES (red) and NPC (blue) 4C signal. Note that many interactions were called by 4C-ker within the plotted window for both ES and NP, but that the documentation did not provide a means to plot them alongside the shown line plots.

conditions (4.5 g/L glucose-DMEN, 15% FCS, 0.1 mM non-essential amino acids, 0.1 mM beta-mercaptoethanol, 1 mM glutamine, 500 U/mL LIF, gentamicin), then passaged onto feeder-free 0.2% gelatin-coated plates for at least two passages to remove feeder cells. For *in vitro* differentiation to neural precursor cells (NPCs), F1 ES cells were cultured in the same medium supplemented with 1 μM PD03259010 and 3 μM CHIR99021 ("2i" conditions) and without feeders. The cells were then cultured for six days with medium without LIF or 2i and with 10% FCS, and with 5 μM retinoic acid for the final four days, to generate embryoid bodies (Bibel et al., 2007). J1/F1 ES or differentiated cells were detached with trypsin, then washed by centrifugation in PBS before fixation. Mouse $CD4^+$ $CD8^+$ double-positive (DP) thymocytes were obtained from 4 week old mouse thymus by FACS with anti-CD4-PE and anti-CD8a-FITC antibodies (eBioScience). Both cell preparations were fixed with 2% formaldehyde in mES culture medium for 10 min at 23°C. The fixation was quenched with cold glycine at a final concentration of 125 mM, then cells were washed with PBS and permeabilized on ice for 1 h with 10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% NP-40, and protease inhibitors. Nuclei were resuspended in *Dpn*II restriction buffer at 10 million nuclei/mL concentration, and 5 million nuclei aliquots were further permeabilized by treatment for either 1 h with 0.4% SDS at 37°C (ES cells), or for 20 min with 0.7% SDS at 65°C, then for 40 min at 37°C (DP cells). The SDS was then neutralized by incubating for a further 1 h with either 2.6% (ES) or 3.3% (DP) Triton-X100 at 37°C. Nuclei were digested overnight with 1000 U *Dpn*II at 37°C, then washed twice by centrifuging and resuspending in T4 DNA ligase buffer. *In situ* ligation was performed in 400 μL T4 DNA ligase buffer with 20,000 U T4 DNA ligase overnight at 16°C. DNA was purified by reverse cross-linking with an overnight incubation at 65°C with proteinase K, followed by RNase A digestion, phenol/chloroform extraction, and isopropanol precipitation. The DNA was digested with 5 U/μg *Csp*6I at 37°C overnight, then re-purified by phenol/chloroform extraction and isopropanol precipitation. The DNA was then circularized by ligation with 200 U/μg T4 DNA ligase under dilute conditions (5 ng/μL DNA), and purified by phenol/chloroform extraction and isopropanol precipitation. 50 ng aliquots of this DNA were used as template for PCR with bait-specific primers containing Illumina adapter termini (primer sequences and optimal PCR conditions available on request). PCR reactions were pooled, primers removed by washing with 1.8× AMPure XP beads, then quantified on a Bioanalyzer (Agilent) before sequencing with a HiSeq 4000 (Illumina).

Pre-Processing 4C Data for 4See

All bait sequence (including and downstream of the primer sequence, up to but not including the GATC *Dpn*II site) are trimmed by the demultiplexing Sabre tool (https://github.com/najoshi/sabre), allowing two mismatches, before mapping to the mm9 genome with Bowtie (Langmead et al., 2009) (**Figure 1B**). Interaction calling was done with peakC (Geeven et al., 2018) with different window sizes as specified by the parameter wSize.

TABLE 1 | Comparison of graphing options from existing 4C analysis methods.

Method	GUI	Preprocessing	Handles conditions/replicates	Plotting options	Annotations
4Cseqpipe	No	Several custom scripts to convert fastq files to multiple formats and intermediate files. These failed for test data and intermediate files had to be made manually.	No	Can alter trendline resolution and plotting window (coordinate space); domainogram parameters fixed	Limited: manually curated bed file gives arrows on plot
r3Cseq	No	Processes bam files directly	Yes, but restricted to pairwise comparison of "experiment" and "control" conditions	Can only alter plotting window	RefSeq genes
fourSig	No	Custom script converts bam to input format	No	Can only alter plotting window	None
FourCSeq	No	Need to set up metadata table in R, which points to processed bam files	Handled in one combined object, but default is to plot each data set individually, and documentation does not say how to do otherwise.	Can only alter plotting window	Positions of genes from transcriptome (unlabeled)
4C-ker	No	Requires bed file of restriction fragments and bedgraph of 4C coverage per observed fragment. Custom scripts to generate from sam files failed and input files had to be made manually.	Yes	In principle, many settings can be changed in the R command prompt (ggplot2 call settings), but is not documented or user-friendly	None
peakC	No	Essentially the same as this manuscript, but utility scripts not provided	Handles replicates but not different conditions	Can only alter plotting window	None

For previously published 4C results (Narendra et al., 2015), fastq files were downloaded from the Gene Expression Omnibus and processed just like the other data sets.

Analysis and Plotting of 4C Data Sets by Other Methods

Three 4C data sets (two replicates from ES cells; one replicate from *in vitro* differentiation of ES cells towards NPCs) were analyzed and plotted by 4Cseqpipe (van de Werken et al., 2012), r3Cseq (Thongjuea et al., 2013), fourSig (Williams et al., 2014), FourCSeq (Klein et al., 2015), 4C-ker (Raviram et al., 2016), and peakC (Geeven et al., 2018), using the default or recommended parameters given within the documentation accompanying the tools.

4See

System Requirements

4See is a GUI written in R (version > = 3.2), with the following packages (and their dependencies) additionally required, found on Bioconductor and/or CRAN: tcltk2, tkrplot, limma, caTools, rtracklayer. All scripts and test data are available under the terms of the GNU General Public License, version 3, on Github: https://github.com/TomSexton00/4See. The GUI is launched by sourcing the main script, 4See.r, from within an R environment. From then on, all manipulation is performed from a windows interface, and does not require use of command prompts. A full user manual in pdf format is also found with the scripts on Github, and is provided as **Supplementary Data**.

Input Format

4See deals with a simple text format, which we term the "cis" format, entailing a header with three tab-delimited fields (data set name, bait chromosome, bait coordinate) followed by a two-column table, denoting the coordinate of the mid-point of every restriction fragment found on the same chromosome as the bait, and its corresponding number of supporting sequence reads. The cis format is generated by a perl script, coord2frag.pl, provided with 4See, which maps the genomic coordinates of 4C sequencing results into their corresponding restriction fragments and then counts the number of reads for each fragment. The perl script accepts any non-headed text format for sequences, as long as a column for chromosome, coordinate, and strand can be specified. The restriction fragment information is provided by "frag" tables, headed four-column tables, giving a unique integer identifier, the chromosome, coordinate, and fragment length for each restriction fragment. These in turn are generated by a provided perl script, makefrags.pl, requiring a user input for the sequence of the primary restriction enzyme cutting site, and a folder containing the sequences in fasta format for each chromosome of the genome assembly used. The header of the cis file provides the required information on the bait location and 4C data set name, but is also used to ensure that only 4C data sets for the same bait (with identical bait chromosome and coordinate) are treated together.

Managing Conditions and Replicates

After loading one or more cis files, 4See opens a dialog box allowing the user to determine how to handle different conditions and replicates by assigning a value to each data set (**Figures 3A, C**). All 4C data sets assigned a non-zero integer are quantile normalized for fairer comparison across data sets (Ritchie et al., 2015). Data sets given the same value are averaged together before plotting; those assigned zero are omitted from normalization and plotting. Additional options allow the plotting color and data label to be specified by the user, and these can be re-run *via* the "Conditions" drop-down menu. Thus a user can rapidly compare different replicates side by side, or average them into one plot for comparison with different cell types or conditions, without needing to reload the data (**Figures 3B, D**).

Plot Settings

4C profiles and the chromatin interactions they uncover differ with bait and experimental condition. In particular, the ease of distinguishing peaks of 4C signal above background depends on the distance of the interaction, since background signal of random chromosome collisions is much higher at shorter ranges (Dekker et al., 2002; Lieberman-Aiden et al., 2009). Other factors, such as whether the interaction is sharp with a single regulatory element, or is broadened across larger regions, such as "super-enhancers", or the extent to which very short-range contacts dominate the plot and hide longer-range loops (which may be a consequence of the 4C digestion efficiency and/or relative compaction of the assessed locus), mean that features of chromatin topology are often overlooked with one fixed plot setting. The control panel of 4See includes options for the user to alter the region plotted, up to ± 1.5 Mb of the bait position, and to set a maximum plotted value on the y-axis (4C signal), to better visualize certain aspects of the 4C profile (**Figure 4A**). However a major confounding factor in visualizing 4C data is the need to smooth the plots, since "spikes" from spurious PCR duplicates make them appear very noisy at single restriction fragment resolution. Most analytical approaches counter this by taking running means (or medians) of sliding windows, but the results can be heavily influenced by the choice of window size. Reflecting this challenge, some 4C analytical tools adopt a "domainogram" approach, whereby averages are taken over many sliding windows of many different sizes, and the results are pooled together in a heat map (de Wit et al., 2008; van de Werken et al., 2012; see also **Figure 2A**), although the visual interpretation of these results is often challenging. To aid user choice in setting appropriate parameters for their particular 4C profile, 4See allows the window size (in numbers of restriction fragments) to be altered, and the appropriate running mean is calculated on the quantile normalized (and averaged, if replicates are pooled) data before plotting. In this manner, different aspects of chromosome topology can be readily explored (**Figures 4B, C**).

Annotations

To put the 4C profiles into a wider biological context, 4See supports the inclusion of three different types of annotations:

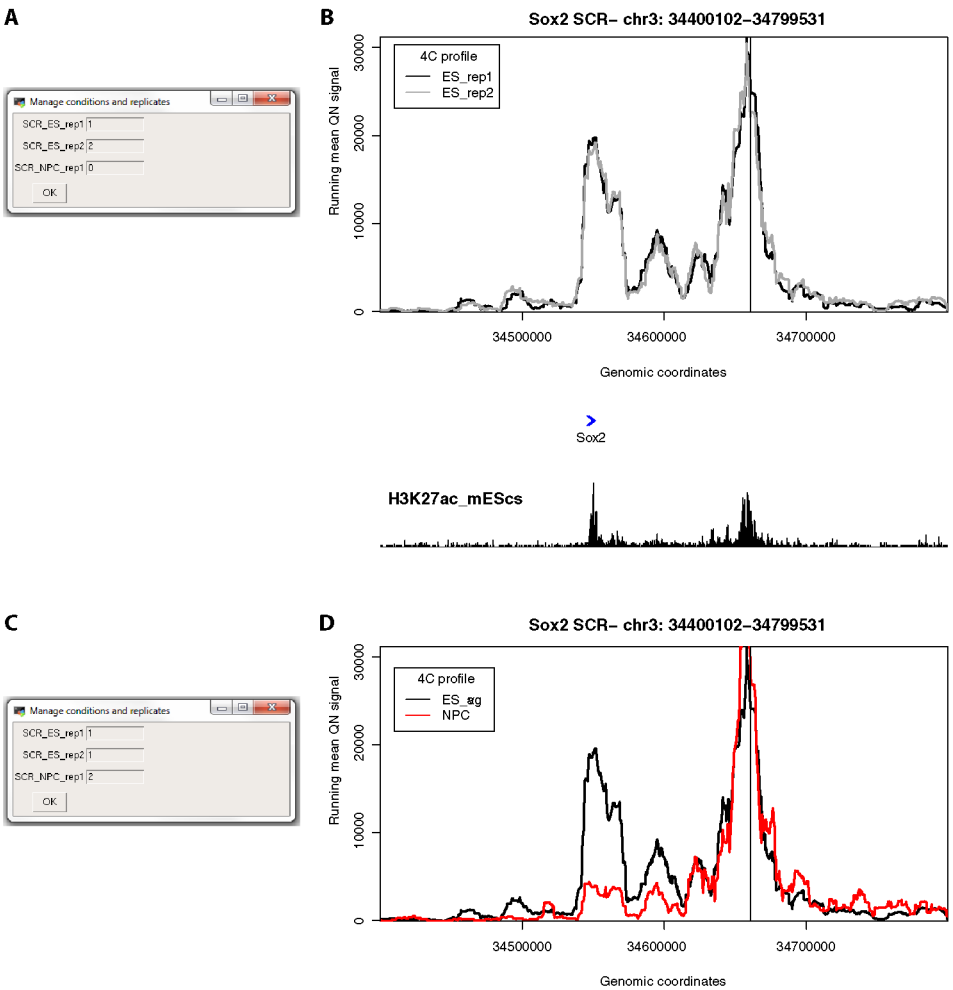


FIGURE 3 | 4See provides flexibility in handling multiple replicates and/or experimental conditions. **(A)** The 4See dialog box for conditions settings automatically opens when cis files are first loaded, in this instance two ES replicates and one NPC 4C data set for the *Sox2* SCR bait. The two ES replicates have been assigned different integers to be treated independently, and the NPC data set has been omitted by assigning it 0. **(B)** The resultant 4See plot from the conditions set in **(A)**, whereby the two ES 4C replicates are quantile normalized and plotted separately, one in black, the other gray. The plot has normalized 4C signal as the y-axis and genomic coordinate of the interacting fragment as the x-axis. The position of the SCR bait is denoted by a black vertical line, and gene position (blue arrows) and the ES H3K27ac ChIP-seq profile (black) is shown underneath the 4C plot. The profiles are highly consistent between replicates, with a strong interaction peak centered on the *Sox2* gene; note that both the gene and enhancer have a strong enrichment for H3K27ac. **(C)** As for **(A)**, but in this case the two ES replicates are given the same value to be averaged together, and the NPC data set is included as a different integer to the ES data sets. **(D)** As for **(B)**, but with the settings conditions of **(C)**, and the redundant gene and H3K27ac tracks omitted. The averaged ES 4C plot is given in black and the NPC 4C plot in red, showing a strong perturbation of the SCR-*Sox2* interaction on differentiation.

genes, linear epigenomic profiles (termed "tracks") and called interactions. Gene information is provided as a tab-delimited headed text file with the following fields: Name, Chr (with the prefix "chr"), Start, End, Strand (as "+" or "–"). When selected, the gene track is plotted in blue directly underneath the 4C profile. Only one gene track can be loaded at a time. Management of epigenomic profiles is more flexible. Any format supported by the *import* function of the rtracklayer package can be supported, but for running time efficiency we recommend loading bigWig files. As for the 4C profiles, the color and plotting level for each individual track can be altered by the user in an automatically loaded dialog box. As before, the plotting levels can be 0 (not plotted), or consecutive, positive integers. When tracks have the same level, their plots are auto-scaled to the maximum value of all of the included data sets within the plotted window. This feature allows fairer comparison for the same epigenetic mark across different conditions/tissue types (**Figure 5**). Technically, the numbers of tracks that can be loaded is only limited by system memory, although the plots become difficult to visually interpret after more than four tracks are loaded at a time.

To better highlight interactions called by existing peak-calling methods, or indeed to test how different methods and/or their parameters perform on specific 4C profiles, interactions (as bed

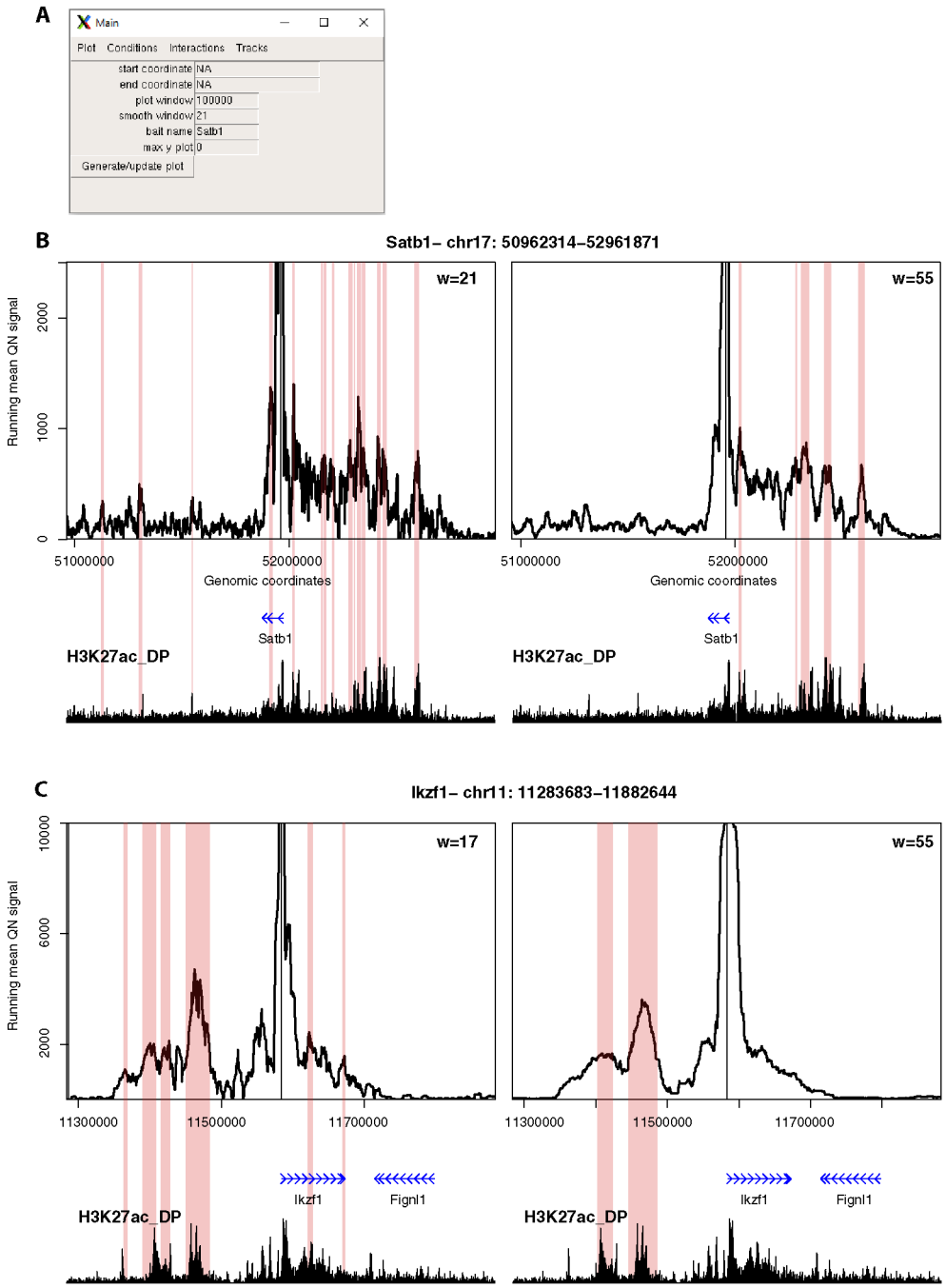


FIGURE 4 | 4See provides flexibility in running mean window sizes. **(A)** The main control panel of 4See, including options to set the x-axis ("start coordinate," "end coordinate," and "plot window") and y-axis ("max y plot") plot limits, to choose a bait name for the plot title ("bait name"), and to set the running mean window size ("smooth window") in number of restriction fragments. **(B)** 4See plots for a 4C data set in DP thymocytes with the *Satb1* gene promoter as bait. Two instances of the same x- and y-axis limits are shown, with a running mean window of 21 (left) or 55 (right) fragments. The position of the *Satb1* bait is denoted by a black vertical line, and gene position (blue arrows) and the DP H3K27ac ChIP-seq profile (black) is shown underneath the 4C plots. Pink rectangles denote regions called as interacting by peakC for the equivalent window size as the plot. For the long-range interaction, the smaller window size appears to have more spurious called interactions, less evidently linked to H3K27ac peaks; the link is better seen with greater smoothing from a larger window size. **(C)** As for **(B)**, but with the *Ikzf1* gene promoter as bait. In this case, the smaller window size (17 fragments) seems to give better resolution of specific interactions with distinct putative enhancers, which are merged into one at larger window sizes.

files or similar, with headed "chr", "start," and "end" columns) can also be loaded, and these are represented as rectangles flanking the relevant region on the 4C profile (**Figures 4B, C**). A dialog box allows the user to alter the color of the annotation and to check whether or not it is plotted. The latter feature is useful since simultaneous plotting of more than one interactions set, which often overlap, can be difficult to visually interpret. Note that whereas these are labeled "Interactions" by 4See, any region described by a bed file can be highlighted in this manner. The user can thus use this setting to highlight any feature of interest, such as called differential interactions between two 4C data sets, or the presence of specific sequence motifs (e.g. CTCF) that may be expected to be enriched at interactions.

Exporting 4See Results

Once the user settings have been finalized, a pull-down menu option allows the plot to be saved in.eps format, where it can be further processed in preparation of a figure for publication or presentation. Alternatively, the data that are actually plotted in the current 4See window (one or more quantile-normalized 4C profiles with a running mean of a specified window size applied) can be exported as bedGraph files, ready for integration into other browsers, such as local instances of UCSC (Kent et al., 2002) or IGV (Robinson et al., 2011).

RESULTS

We demonstrate the usefulness of 4See on different original and previously published 4C data sets. First, we investigate the interaction between the mouse *Sox2* gene and an established cluster of enhancers (the "SCR", or *Sox2* control region), which has been shown to be essential for *Sox2* expression in pluripotent cells (Zhou et al., 2014). Using bait primers at the SCR, we generated two biological replicates for ES cells and one after *in vitro* differentiation ("NPC"; **Figure 3**). As expected, we observed a strong interaction with the *Sox2* gene which is greatly reduced on differentiation. After loading the three data sets into the 4See browser, only changing the options within one dialog box is required to switch the view from plotting the two ES biological replicates side by side (omitting the differentiated data set) to confirm that they have consistent profiles, to comparing the averaged ES profile with the differentiated one.

Second, we explored different distance ranges of promoter-enhancer interactions at key developmental genes in mouse $CD4^+/CD8^+$ (double-positive, DP) thymocytes, namely the distal (~500 kb) enhancer cluster for *Satb1*, and the shorter-range (~50 kb) enhancer for *Ikzf1* (**Figure 4**). Comparing 4C plots at different running mean window sizes, it is apparent that different insights can be gained, and that no one window size is optimum for all profiles. For *Satb1*, shorter window sizes create what appear to be noisy profiles at the large genomic span assessed, and specific interactions are harder to discern. When the running mean window size is increased, the profile becomes smoother, and apparent peaks line up well with putative enhancers, as denoted by the presence of H3K27ac. Conversely, at the shorter distances assessed at the *Ikzf1* locus, a smaller window size allows interactions with specific enhancers to be resolved, whereas they merge into one large peak at larger window sizes. In support of this observation, we called interactions using the peakC algorithm (Geeven et al., 2018) at different window sizes, and found a good visual corroboration between discernible peaks and called interactions. 4See allows rapid re-plotting of 4C profiles with different window sizes, and also has the functionalities for adding the epigenomic profile and highlighting called interactions directly on the plot.

Third, we compared the same *Satb1* promoter-enhancer interaction between DP thymocytes, where the gene is highly expressed, and ES cells, where the gene is silent (**Figure 5**). As expected, the gene does not make any specific contacts with the thymocyte enhancer in ES cells. This locus is largely devoid of H3K27ac in ES cells, but a common problem with some browsers is that an automatic scaling creates some apparent peaks from noise on a small range of the y-axis (**Figure 5B**). 4See counters this by providing flexibility with how the epigenomic tracks are handled. By coercing the two tracks to the same scale, the difference between the two tissues is much more evident (**Figure 5C**).

Finally, we used 4See to re-analyze published 4C data, namely comparing profiles from the *Hoxa5* gene between wild-type ES cells and those where one or more key CTCF insulator sites have been deleted (Narendra et al., 2015). In this study, the authors reported that CTCF site loss caused topological defects during differentiation to neurons, with inappropriate spreading of H3K27me3. However, their analyses concluded that the topology of the *Hoxa* locus was largely unchanged in pluripotent cells (**Figure 6A**). Plotting the same data with 4See, it appears that ectopic looping interactions are formed between *Hoxa5* and more caudal regions of the locus (**Figure 6B**). Different loop calling algorithms with different parameter choices were inconsistent in calling these apparent interactions as "significant", and only one biological replicate was available, so the importance of this observation is yet to be confirmed. In any case, the CTCF site deletion did not alter H3K27me3 patterning or *Hoxa* gene expression within undifferentiated ES cells (Narendra et al., 2015), so any potential topological changes do not appear to be borne out in other phenotypes. However, we wish to use this example to highlight how the use of a flexible browser like 4See facilitates exploration of the data, potentially identifying new features that "one size fits all" algorithms may overlook.

DISCUSSION

Using novel and previously published data sets for demonstration, we have shown the flexibility and utility of 4See in exploring 4C data. With limited processing of sequencing results, and one line of code in the R prompt, a user-friendly windows-based interface is available for a broader community to explore chromatin interaction profiles. As a consequence, we envisage that 4See will be of great use to the

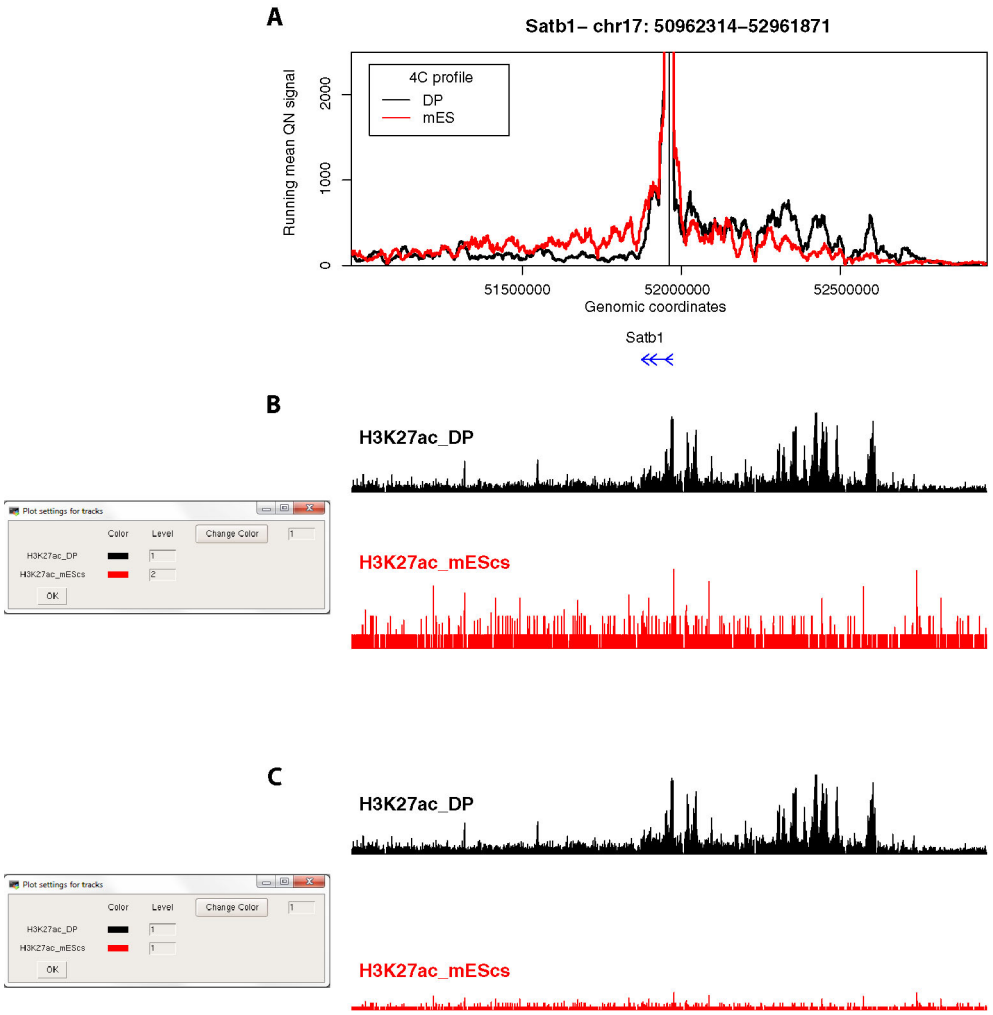


FIGURE 5 | 4See provides flexibility in handling epigenomic track scales. **(A)** The same 4C profile as **Figure 4B** (55-fragment window size; black) is plotted alongside the 4C profile for ES cells, where the locus is silent, and thymocyte enhancer interactions are not evident. **(B)** The 4See dialog box for managing epigenomic tracks defines how the different tracks are scaled. In this case, the ChIP-seq tracks for H3K27ac in ES and DP cells are treated independently, so the autoscaling of the ES track creates some spurious peaks from noise above background. **(C)** As for **(B)**, but this time the two H3K27ac tracks have been set the same integer, making the lower ES signal much more visually apparent in the plot.

chromatin field. The input cis files are not very large (~ 4 MB for mouse or human), so the browser can be run on most desktop computers and laptops. The major systems limitation comes from the importing of epigenomic tracks (which can be >500 MB) with the rtracklayer package, which is the slowest step and may overload some standalone computers if too many tracks are imported at once. If the user is interested in only a specific set of baits, the system load can be reduced by restricting imports to chromosome-specific tracks. Due to the reliance of 4See plotting on quantile normalization, which is confounded by an excessive number of zeros or very small values, 4See is not an appropriate tool for visualizing very long-range (>1.5 Mb) or interchromosomal interactions; although their built-in graphical capabilities are more limited, the tools linked to algorithms such as fourSig should be used instead (Williams et al., 2014; Walter et al., 2019). It should also be noted that 4See does not replace the existing suite of interaction calling algorithms (Walter et al., 2019). Indeed, due to its capacity to incorporate these algorithms' results into the plots, 4See should be viewed as a complementary tool for comprehensive 4C analysis, whereby the results of the algorithms can be readily visualized and compared to epigenomic tracks for validation and obtaining biological insight. Overall, 4See, in conjunction with other analytical tools, promises to facilitate chromatin interaction exploration, and will thus be of use to the epigenetics community.

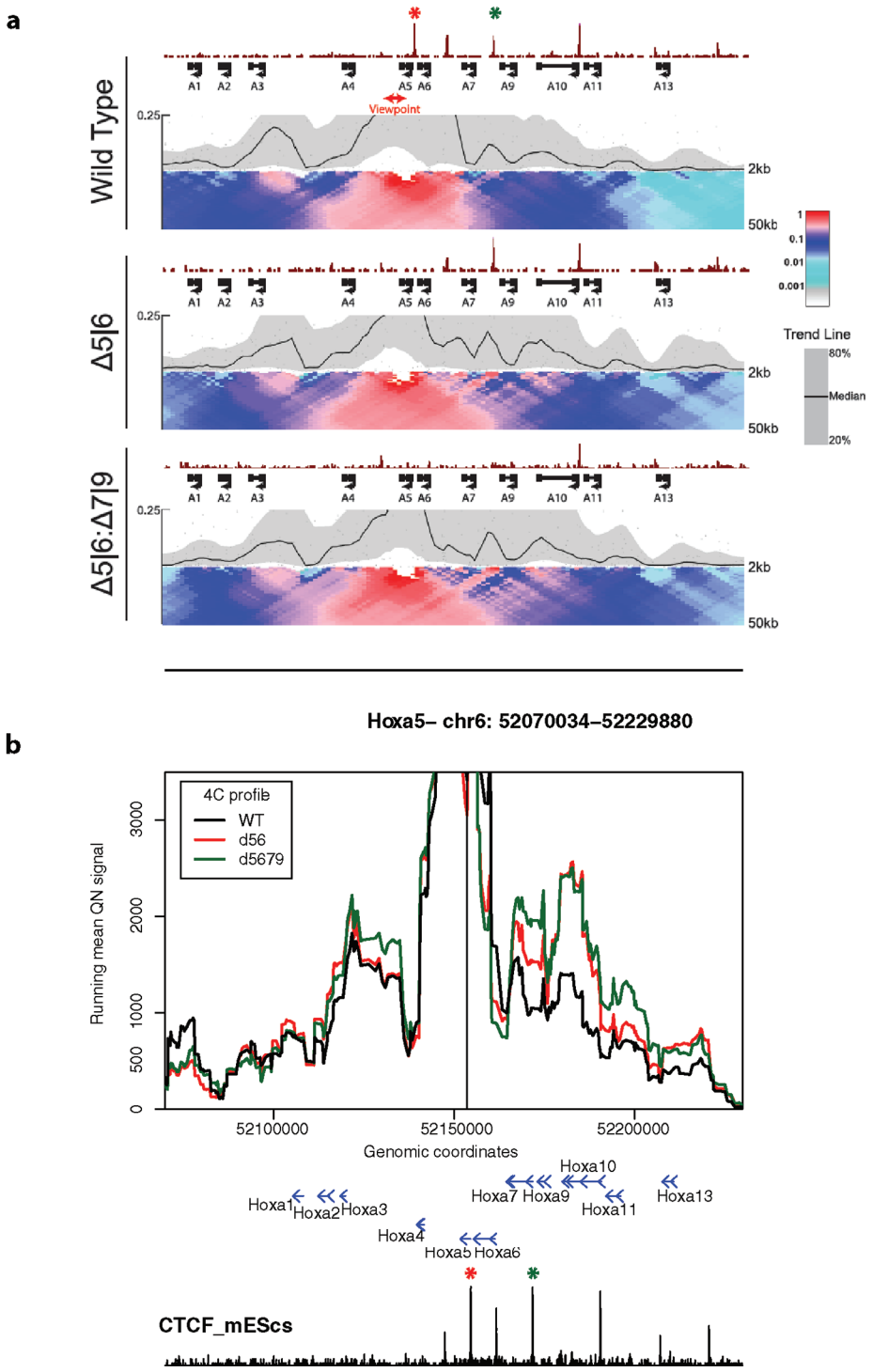


FIGURE 6 | 4See exploration can uncover previously overlooked features of chromatin topology. **(A)** Reproduced Supplemental Fig 7A from Narendra et al. (2015), used with permission. 4C profiles from the *Hoxa5* bait are shown as domainogram heat maps for wild-type ES cells (top), as well as lines that have had deletions of one CTCF site (middle; site of deletion denoted by red asterisk) or two (bottom; additional deletion site denoted by green asterisk). The CTCF ChIP-seq profile is shown above the 4C sets. No chromatin topology differences are apparent between these cell lines, and the original study concluded that spatial phenotypes only occurred on cell differentiation (Narendra et al., 2015). **(B)** The same data, processed and plotted using 4See. The position of the *Hoxa5* bait is denoted by a black vertical line, and gene position (blue arrows) and the ES CTCF ChIP-seq profile (black) is shown underneath the 4C plots. Red and green asterisks denote the positions of the single and double CTCF site deletions, as for **(A)**. In this plot, CTCF-dependent restriction of interactions between *Hoxa5* and more caudal regions (e.g. the gene body of *Hoxa10*) seems apparent.

DATA AVAILABILITY STATEMENT

Raw 4C data and the processed cis files first described here are available on GEO (GSE137417); previously published 4C results (Narendra et al., 2015) are available as data set GSE60240. ChIP-seq tracks were obtained from GEO for mouse ES (Shen et al., 2012) (CTCF and H3K27ac; GSE29218) and DP cells (Vanhille et al., 2015) (H3K27ac; GSE63732).

AUTHOR CONTRIBUTIONS

YBZ and TS designed and wrote the 4See code and accompanying scripts. AP and AM performed 4C experiments and preprocessing. TS wrote the manuscript, with continuous feedback from all authors.

FUNDING

This study was supported by funds from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Starting Grant 678624-CHROMTOPOLOGY), the ATIP-Avenir program, and the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir ANR-10-IDEX-0002-02. YBZ is supported by funds from LabEX INRT, la Region Grand Est and the ERC. AP is supported by funds from LabEX INRT, the ERC, and the Ligue Nationale Contre le Cancer. AM is supported by funds from IDEX (University of Strasbourg) and the Institut National du Cancer. TS is supported by INSERM.

ACKNOWLEDGMENTS

Sequencing was performed by the IGBMC GenomEast platform, a member of the France Genomique consortium (ANR-10-INBS-0009). We thank Claudine Ebel and Muriel Philipps for their support within the flow cytometry platform.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01372/full#supplementary-material

REFERENCES

Alexander, J. M., Guan, J., Li, B., Maliskova, L., Song, M., Shen, Y., et al. (2019). Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, e41769. doi: 10.7554/eLife.41769

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal dynamics at the Shh locus: limb bud-specific differential regulation of competence and active transcription. *Dev. Cell* 16 (1), 47–57. doi: 10.1016/j.devcel.2008.11.011

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339 (6123), 1074–1077. doi: 10.1126/science.1232542

Ben Zouari, Y., Molitor, A. M., Sikorska, N., Pancaldi, V., and Sexton, T. (2019). ChiCMaxima: a robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C. *Genome Biol.* 20 (1), 102. doi: 10.1186/s13059-019-1706-3

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., et al. (2019). Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell.* 76 (3), 473–484. doi: 10.1016/j.molcel.2019.07.038

Bibel, M., Richter, J., Lacroix, E., and Barde, Y. A. (2007). Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.* 2 (5), 1034–1043. doi: 10.1038/nprot.2007.147

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G. L., Lubling, Y., et al. (2017). Multiscale 3D genome rewiring during mouse neural development. *Cell* 171 (3), 557–572 e524. doi: 10.1016/j.cell.2017.09.043

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., et al. (2010).). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U.S.A.* 107 (50), 21931–21936. doi: 10.1073/pnas.1016071107

de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H. J., and van Steensel, B. (2008). Global chromatin domain organization of the Drosophila genome. *PloS Genet.* 4 (3), e1000045. doi: 10.1371/journal.pgen.1000045

de Wit, E., Vos, E. S., Holwerda, S. J., Valdes-Quezada, C., Verstegen, M. J., Teunissen, H., et al. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* 60 (4), 676–684. doi: 10.1016/j.molcel.2015.09.023

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

Forcato, M., Nicoletti, C., Pal, K., Livi, C. M., Ferrari, F., and Bicciato, S. (2017). Comparison of computational methods for Hi-C data analysis. *Nat. Methods* 14 (7), 679–685. doi: 10.1038/nmeth.4325

Geeven, G., Teunissen, H., de Laat, W., and de Wit, E. (2018). peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* 46 (15), e91. doi: 10.1093/nar/gky443

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459 (7243), 108–112. doi: 10.1038/nature07829

Herranz, D., Ambesi-Impiombato, A., Palomero, T., Schnell, S. A., Belver, L., Wendorff, A. A., et al. (2014). A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* 20 (10), 1130–1137. doi: 10.1038/nm.3665

Hughes, J. R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., et al. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* 46 (2), 205–212. doi: 10.1038/ng.2871

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12 (6), 996–1006. doi: 10.1101/gr.229102

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465 (7295), 182–187. doi: 10.1038/nature09033

Klein, F. A., Pakozdi, T., Anders, S., Ghavi-Helm, Y., Furlong, E. E., and Huber, W. (2015). FourCSeq: analysis of 4C sequencing data. *Bioinformatics* 31 (19), 3085–3091. doi: 10.1093/bioinformatics/btv335

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., et al. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.* 18 (8), 956–963. doi: 10.1038/nsmb.2085

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10 (3), R25. doi: 10.1186/gb-2009-10-3-r25

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., et al. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* 12 (14), 1725–1735. doi: 10.1093/hmg/ddg180

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326 (5950), 289–293. doi: 10.1126/science.1181369

Narendra, V., Rocha, P. P., An, D., Raviram, R., Skok, J. A., Mazzoni, E. O., et al. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347 (6225), 1017–1021. doi: 10.1126/science.1262088

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W. (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35 (2), 190–194. doi: 10.1038/ng1244

Pradeepa, M. M., Grimes, G. R., Kumar, Y., Olley, G., Taylor, G. C., Schneider, R., et al. (2016). Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat. Genet.* 48 (6), 681–686. doi: 10.1038/ng.3550

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470 (7333), 279–283. doi: 10.1038/nature09692

Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159 (7), 1665–1680. doi: 10.1016/j.cell.2014.11.021

Raviram, R., Rocha, P. P., Muller, C. L., Miraldi, E. R., Badri, S., Fu, Y., et al. (2016). 4C-ker: A Method to Reproducibly Identify Genome-Wide Interactions Captured by 4C-Seq Experiments. *PloS Comput. Biol.* 12 (3), e1004780. doi: 10.1371/journal.pcbi.1004780

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi: 10.1093/nar/gkv007

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518 (7539), 317–330. doi: 10.1038/nature14248

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., et al. (2011). Integrative genomics viewer. *Nat. Biotech.* 29 (1), 24–26. doi: 10.1038/nbt.1754

Sahlen, P., Abdullayev, I., Ramskold, D., Matskova, L., Rilakovic, N., Lotstedt, B., et al. (2015). Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biol.* 16, 156. doi: 10.1186/s13059-015-0727-9

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

Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* 20 (8), 437–455. doi: 10.1038/s41576-019-0128-0

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B. M., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* 25 (4), 582–597. doi: 10.1101/gr.185272.114

Schwartzman, O., Mukamel, Z., Oded-Elkayam, N., Olivares-Chauvet, P., Lubling, Y., Landan, G., et al. (2016). UMI-4C for quantitative and targeted chromosomal contact profiling. *Nat. Methods* 13 (8), 685–691. doi: 10.1038/nmeth.3922

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* 488 (7409), 116–120. doi: 10.1038/nature11243

Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., et al. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38 (11), 1348–1354. doi: 10.1038/ng1896

Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E., and Lenhard, B. (2013). r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.* 41 (13), e132. doi: 10.1093/nar/gkt373

van de Werken, H. J., Landan, G., Holwerda, S. J., Hoichman, M., Klous, P., Chachik, R., et al. (2012). Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* 9 (10), 969–972. doi: 10.1038/nmeth.2173

Vanhille, L., Griffon, A., Maqbool, M. A., Zacarias-Cabeza, J., Dao, L. T., Fernandez, N., et al. (2015). High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun.* 6, 6905. doi: 10.1038/ncomms7905

Walter, C., Schuetzmann, D., Rosenbauer, F., and Dugas, M. (2019). Benchmarking of 4C-seq pipelines based on real and simulated data. *Bioinformatics* 35 (23), 4938–4945. doi: 10.1093/bioinformatics/btz426

Williams, R. L. Jr., Starmer, J., Mugford, J. W., Calabrese, J. M., Mieczkowski, P., Yee, D., et al. (2014). fourSig: a method for determining chromosomal interactions in 4C-Seq data. *Nucleic Acids Res.* 42 (8), e68. doi: 10.1093/nar/gku156

Zhou, H. Y., Katsman, Y., Dhaliwal, N. K., Davidson, S., Macpherson, N. N., Sakthidevi, M., et al. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* 28 (24), 2699–2711. doi: 10.1101/gad.248526.114

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Angeliki PLATANIA

Assessing the links between enhancer-promoter proximity, local chromatin dynamics and gene activity

Résumé

Le lien entre la dynamique de la chromatine et la transcription reste flou, notamment parce que très peu de loci ont été évalués. Dans cette étude, en utilisant le locus autour du gène de la pluripotence *Sox2* comme modèle, j'ai réalisé un double marquage pour suivre la communication et la mobilité du promoteur et de l'enhancer, couplé à un marquage de l'ARN naissant de *Sox2* et étudié la dynamique de la machinerie protéique transcriptionnelle. Contrairement à une étude précédente, j'ai observé que *Sox2* est très proche de son enhancer mais aussi, que leur distance de séparation n'est pas totalement corrélée avec l'activation de la transcription. La perturbation de l'expression de *Sox2* lors de la différenciation des cellules, des délétions de parties de l'enhancer et de l'inhibition de l'initiation transcriptionnelle, ont montré que la mobilité est augmentée spécifiquement au niveau du promoteur de *Sox2* en même temps que l'activation de la transcription. Cependant, les loci participant à l'activation de la transcription ne présentent pas une dynamique chromatinienne sensiblement différente lorsqu'ils sont transcriptionnellement inactifs.

Résumé en anglais

The link between chromatin dynamics and transcription remains largely unclear, not least because so few loci have been assessed. In this study, using the locus around the pluripotency gene Sox2 as a model, I performed double-label ANCHOR to track promoter and enhancer communication and mobility, coupled to labelling of nascent Sox2 RNA and studies of the dynamics of the transcriptional protein machinery. Contrary to a previous study, I observed that Sox2 is in very close proximity with its enhancer, but also that the exact separation distance does not exactly correlate with transcriptional bursting. Perturbation of Sox2 expression upon differentiation, deletions within the enhancer and inhibition of transcriptional initiation, showed that the apparent diffusive speed is increased specifically at Sox2 promoter together with transcriptional activation. However, loci participating in transcriptional bursts do not have notably different chromatin dynamics to poised loci.

Key words: Sox2, chromatin dynamics, live-imaging, transcription, RNA Pol II