Rebuilding a superenhancer to investigate the additive and synergistic effects of its individual components

Joseph William Blayney

St Cross college

A thesis submitted for the degree of Doctor in Philosophy

Wellcome Doctoral Training Programme in Chromosome and Developmental Biology

Weatherall Institute of Molecular Medicine

University of Oxford

Abstract

Superenhancers are clusters of putative enhancers, densely occupied by histone 3 lysine 27 acetylation, transcription factors, and coactivators such as the mediator complex. There is a tendency for superenhancers to regulate key cell-type-specific genes, and their dysregulation is often associated with disease. It remains unclear whether superenhancer constituent elements work independently or collaboratively and whether they have distinct roles in activating expression of their cognate genes. During my DPhil, I address these two questions by studying a model mammalian superenhancer at the mouse α-globin locus, active exclusively in erythroid cells. I sought to determine how the five constituents (R1, R2, R3, Rm, R4) of the mouse α-globin superenhancer exert control over α-globin gene expression during erythropoiesis. Firstly, I tested the sufficiency of the previously described strongest α-globin superenhancer constituent (R2) to independently activate α-globin expression; this entailed characterising, in detail, a mouse model in which the other four α-globin superenhancer elements have been removed from the native locus. Surprisingly, I show that this strong enhancer is incapable of driving the expected level of expression, independently. Secondly, following synthetic biology and genome editing techniques in mouse Embryonic Stem Cells (mESCs), I rebuilt the native α-globin superenhancer in all informative combinations, starting from an enhancer-less baseline in which all five constituents have been removed. Examination of molecular phenotypes of erythroid cells derived from the engineered mESCs revealed a complex relationship between inequivalent constituents which cooperate in additive, synergistic, and redundant fashions. I also uncovered a novel class of regulatory element within the α-globin superenhancer, which I named facilitators. Unlike canonical enhancers (R1, R2), facilitators (R3, Rm, R4) have no intrinsic enhancer activity. However, they are necessary to potentiate canonical enhancers, in order to attain optimal levels of target gene expression. Lastly, by comparing the impact of the three facilitators on gene expression, I discovered a functional hierarchy that seems to be position-dependent. Ultimately, I have rigorously dissected the α-globin superenhancer and described a new class of regulatory element, which I named facilitators, that were previously mistaken for weak enhancers. Furthermore, I unravelled a necessary mode of

cooperation manifested between two types of superenhancer constituent elements, canonical enhancers and facilitators, highlighting an emergent property of a superenhancer.

Acknowledgements:

There are many people I would like to thank for helping me through my DPhil and offering support over the past four years. Firstly, thank you to my brilliant supervisors, Doug Higgs and Mira Kassouf. Doug's guidance and advice, both relating to my project and to science in general, has been invaluable. His optimism has been infectious and has made working on this project exciting and satisfying. I would like to extend an especially big thank you to Mira, who has tirelessly supported the project and me, and who has constantly been available and enthusiastic. Mira has been a great and inspirational scientific mentor, and has also driven me to become a more efficient worker and a better communicator.

Next, I would like to thank all members of the Higgs lab, past and present, who have supported me tremendously over the last four years. A particular thanks to Hele Francis, who began the work which turned into my project.

This project wouldn't have been possible without the incredible WIMM facilities. I'd like to thank the entire FACS facility, the transgenic facility (especially Jackie Sloane-Stanley), and the genome engineering facility – in particular Philip Hublitz, whose help in generating numerous genetic models was critical.

I'd also like to thank Professor Jef Boeke and those working in his lab, especially Brendan Camellato. Without the collaboration established between our labs, this project would never have begun.

Countless others have helped with guidance and support along the way, including: Rob Beagrie, Chris Babbs, Jim Hughes, James Davies, Rob Klose, Tom Milne, and Cornelia Van Dujin. I'd also like to thank those in the lab (and beyond) who have offered moral support and friendship – I'll miss all of our two hour lunches which morphed into even longer coffee breaks! Thanks also to Ravza Gür, both for her friendship and her (very patient) bioinformatics tutoring.

Outside of the lab, I'd like to thank my amazing wife, Ane Stranger, for her support and for reminding me that there's more to life than genome engineering! Also, my

parents, my siblings, my Aunt, and Frank “Pop” Davison, for all their care and encouragement.

This work was generously funded by Wellcome Trust and the Medical Research Council UK.

Contents:

List of figures:

List of abbreviations:

3C	Chromosome conformation capture
ATAC-seq	Assay for transposase-accessible chromatin followed by sequencing
BAC	Bacterial artificial chromosome
ChIPmentation	Chromatin immunoprecipitation coupled with tagmentation
ChIP-seq	Chromatin immunoprecipitation followed by sequencing
CPM	Counts per million mapped reads
CRISPR	Clustered regularly interspaced short palindromic repeats
CTCF	CCCTC-binding factor
DNA	Deoxyribonucleic acid
E12.5/E14.5	Embryonic day 12.5/14.5
E14	E14-TG2a.IV (mESC clone)
E17.5	Embryonic day 17.5
E9.5/E10.5	Embryonic day 9.5/10.5
EB	Embryoid body
eRNA	Enhancer RNA
FACS	Fluorescence-activated cell sorting
FIMO	Find Individual Motif Occurrences
GFP	Green fluorescent protein
gRNA	guide RNA
HDR	Homology-directed repair
MACS2	Model-based Analysis of ChIP-Seq
MEME	Multiple Em for Motif Elicitation
mESC	Mouse embryonic stem cell
NGS	Next-generation sequencing
PAM	Protospacer adjacent motif
PCA	Principal component analysis
PCR	Polymerase chain reaction
Pol2	RNA polymerase II
Poly-A (-/+)	Poly-adenylation tail (-/+)

qPCR	Quantitative polymerase chain reaction
RMGR	Recombinase-Mediated Genomic Replacement
RNA	Ribonucleic acid
RNA-seq	RNA-sequenciny
RT-PCR	Real-time PCR
TF	Transcription factor
TFBS	Transcription factor binding site
Tn5	Tn5-transposase
UCSC	University of California Santa Cruz
WT	Wild type
YFP	Yellow fluorescent protein

Chapter one: Introduction:

1.1 The non-coding genome:

The human genome contains ~20,000 protein-coding genes, yet this accounts for only around 2% of the overall genome. Originally, the remaining 98% was thought to be non-functional "junk DNA"; however, studies conducted over the last two decades have shown that >80% of the genome displays biochemical characteristics associated with biological function (Dunham et al., 2012). Therefore, although the vast majority of the Eukaryotic genome is non-coding, this does not necessarily mean that it is non-functional.

This is in stark contrast to Prokaryotic genomes, which contain on average 6-14% non-coding DNA (Rogozin et al., 2002). The fact that these simple single-celled organisms contain so much less non-coding DNA compared to more complex multicellular organisms suggests that non-coding DNA may play a role in regulating spatiotemporal gene expression.

Broadly speaking, the non-coding mammalian genome can be split into three classes of regulatory elements important for controlling spatiotemporal patterns of gene expression: promoters, enhancers and boundary elements (Oudelaar & Higgs, 2021). Boundary elements such as CTCF sites are important for establishing and maintaining appropriate genome organisation. This structure is thought to delimit the function of the other two element classes, enhancers and promoters, which have direct roles in activating gene expression.

The advancement of technologies, molecular and imaging tools, has allowed us to progress from investigating genome organisation at the whole chromosome level (Flemming, 1875) to single base-pair resolution (Hua et al., 2021; Krietenstein et al., 2020). We now know that both proximal and distal regions of chromatin can preferentially interact within defined and identifiable domains, for example topologically associating domains (TADs), sub-TADs and loops (Beagan & Phillips-Cremins, 2020). Frequency of interactions between enhancers, promoters and boundary elements, within these structures, is thought to greatly influence gene expression (Ibrahim & Mundlos, 2020).

1.2 Enhancers and gene regulation:

1.2.1 What are enhancers?

Enhancers are regions of DNA which recruit transcription factors (TFs) and activate expression of target genes in a cell-type-specific manner (Long et al., 2016; Shlyueva et al., 2014; Spitz & Furlong, 2012) (**figure 1.1A**). The first enhancer was described over forty years ago (Banerji et al., 1981), and since then, more than 300,000 putative enhancers have been identified in the human and mouse genomes (H. Chen & Liang, 2020; Shen et al., 2012).

Enhancers were originally defined as sequences which "can act over very long distances, and independent of their orientation" (Banerji et al., 1981; Mercola et al., 1983). These attributes have been tested and sometimes challenged over the years. Enhancers have been shown to activate expression of target genes located far away along the linear genome (Schoenfelder & Fraser, 2019). However, the genomic distances over which different enhancers act varies widely, from a few kilobases to over one megabase, and some studies have even claimed that enhancers can activate genes located on different chromosomes (Geyer et al., 1990; Lomvardas et al., 2006; Ong & Corces, 2011). Recently, an inverse correlation between enhancer activity and enhancer-promoter linear separation has been reported, further suggesting that enhancers are not completely agnostic to the genomic distance separating them from their target genes (Rinzema et al., 2021; Zuin et al., 2022).

Enhancers were also thought to act in an orientation-independent manner (Banerji et al., 1981); however, this is subject to debate. There is scarcity in reports challenging the orientation of enhancers in their natural chromatin context (Canver et al., 2015; Kassouf et al., 2022) and some reports highlight the promiscuity of enhancers and their ability to simultaneously control more than one gene in both directions (Moorthy et al., 2017). Moreover, in the context of chromatin, the organisation of the genome can act to constrain enhancer activity in one particular direction. For example, an enhancer located at the 5' boundary of a TAD will preferentially interact with, and activate expression of, target genes located 3', within the body of the TAD (Barrington et al., 2019; Hsieh et al., 2020; Vian et al., 2018).

1.2.2 Enhancer sequence characteristics:

Enhancer sequences are enriched for transcription factor binding sites (TFBSs), with the majority of enhancers containing motifs associated with a number of different factors. The cohort of TFs that a particular enhancer recruits lends cell type-specificity (Bonn et al., 2012; Long et al., 2016; Singh et al., 2021; Spitz & Furlong, 2012; Zinzen et al., 2009). For example, an erythroid-specific enhancer may contain Gata1, Klf1 and Tal1 motifs, docking sites for erythroid-specific TFs (Doré & Crispino, 2011).

Many studies have dissected enhancer "grammar" and "syntax", investigating the relative importance of TF motif strength, orientation, spacing, diversity and order (C. D. Arnold et al., 2013; Bonn et al., 2012; Farley et al., 2016; Long et al., 2016a; Singh et al., 2021; Spitz & Furlong, 2012; Zinzen et al., 2009). At the extremes, there are two models of enhancer grammar: "enhanceosomes" and "flexible billboards" (Arnosti & Kulkarni, 2005; Spitz & Furlong, 2012) (**figure 1.1B**). The enhanceosome model describes enhancers with such strict grammar and syntax rules that any change in sequence will abolish all enhancer activity; this model was originally derived from studying the interferon β enhancer (Maniatis et al., 1998), which does exhibit extreme sensitivity to changes in sequence and syntax. This is likely due to a process of cooperative TF binding and the requirement for precise protein-protein as well as protein-DNA interactions to achieve target gene up-regulation (Panne et al., 2007). On the other extreme, the flexible billboard model refers to enhancers in which motif order, spacing and orientation are unimportant, and simply containing the appropriate motifs, to recruit the necessary TFs, is sufficient for target gene activation (Kulkarni & Arnosti, 2003; Spitz & Furlong, 2012). The majority of enhancers appear to sit somewhere in the middle, able to tolerate some alterations in sequence and syntax without losing all enhancer activity (Long et al., 2016; Singh et al., 2021; Spitz & Furlong, 2012).

1.2.3 How do enhancers work?

Enhancers recruit TFs, which in turn recruit a variety of transcriptional coactivators, such as the mediator complex (Long et al., 2016; Spitz & Furlong, 2012). Coactivators interact with the activation domains of enhancer-docked TFs; these activation domains are often intrinsically disordered regions which can form many multivalent

interactions with coactivator domains, for example, Oct4 and Gcn4 interactions with components of the mediator complex in mESCs (Boija et al., 2018). The mechanism(s) via which enhancer-bound mediator is then brought into proximity with promoters, in order to up-regulate gene expression, is unclear. Three of the most prominent theories include formation of transcriptional hubs (potentially through liquid-liquid phase separation or "TF trapping"), enhancer-promoter interaction through bi-directional loop extrusion, and direct interaction between TFs (and/or looping factors) bound to enhancers and those bound at promoters (Deng et al., 2012; Hua et al., 2021; Monfils & Barakat, 2021; Oudelaar & Higgs, 2021) (**figure 1.1C**). It is important to note that these hypotheses are not mutually exclusive, and different enhancers may well exploit different mechanisms. Several studies had proposed that mediator itself might form a physical bridge between enhancers and promoters, although recent work has suggested that mediator mainly serves as a "functional" bridge, allowing communication between enhancers and promoters, rather than tethering them together (Crump et al., 2021; Krivega & Dean, 2017). Regardless, mediator plays important roles in recruitment and stability of RNA polymerase II (Pol2) at target genes' promoters, thereby facilitating formation of the preinitiation complex and downstream gene expression. Through these processes, enhancers are thought to primarily activate target genes at the stage of transcription initiation (Cramer, 2019; Larke et al., 2021; Soutourina, 2018).

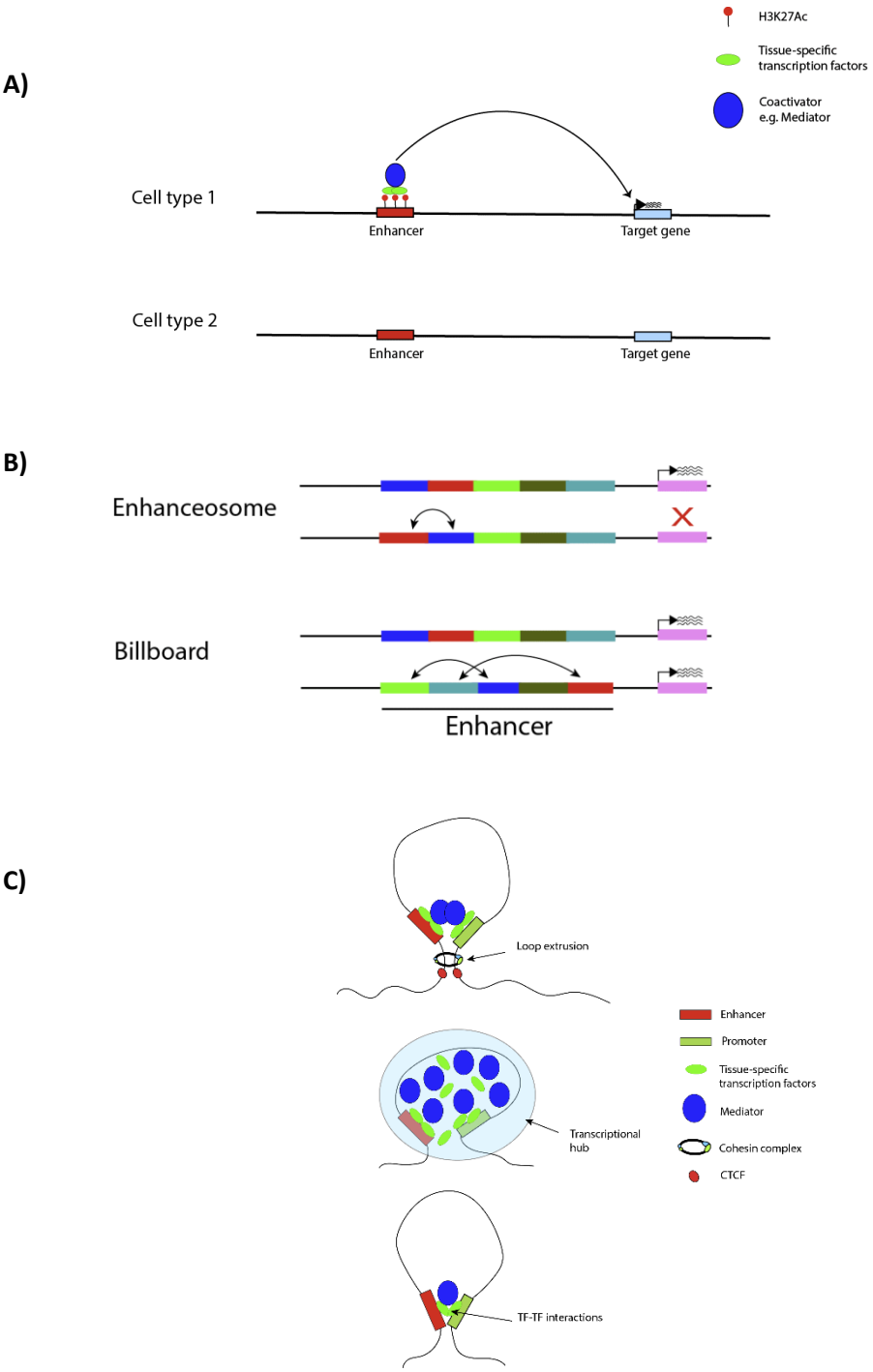


Figure 1.1 What are enhancers, and how do they impact gene expression?

Enhancers activate gene expression in a cell-type-specific fashion. In "cell type 1", the enhancer is bound by TFs and coactivators, allowing it to up-regulate the target gene, and in cell type 2, the enhancer and target gene remain inactive. (B) "Enhanceosomes" (top) are sensitive to changes in sequence; "billboards" (bottom), are more tolerant of changes to sequence. (C) There are various models of how enhancers interact with, and up-regulate their target genes: loop extrusion (top), formation of transcriptional hubs (middle), and specific TF-TF interactions (bottom). Note, despite distinct graphical representation, these models are NOT mutually exclusive.

1.2.4 Enhancer studies: genome-wide and locus-specific studies:

Identifying enhancers based on their biomolecular signatures:

Generally, enhancers are classified bioinformatically based on chromatin accessibility, TF occupancy, and enrichment for enhancer-associated histone modifications such as histone 3 lysine 4 mono-methylation (H3K4Me1) and histone 3 lysine 27 acetylation (H3K27Ac) (**figure 1.2A**). Bioinformatic detection of enhancers in this manner is high throughput and useful for identifying novel enhancer candidates, but the descriptive nature of these methods, without a functional readout to validate predictions, has led to wildly varying estimates of enhancer abundance, from ~40,000 to 300,000, in the human genome (Andersson et al., 2014; H. Chen & Liang, 2020; Dunham et al., 2012).

It can be difficult to bioinformatically distinguish enhancers from promoters. Both element types are associated with nucleosome-free regions (assayed by chromatin accessibility) and both are enriched for TF binding and H3K27Ac. Enhancers tend to be more enriched for H3K4Me1, whereas promoters are more enriched for H3K4Me3; however, it is not uncommon to find both modifications present over either element type. Additionally, a number of recent studies have shown that H3K27Ac is dispensable for enhancer activity, suggesting that some enhancers may be missed when discounting elements devoid of this modification (Bonn et al., 2012; Catarino & Stark, 2018; Pengelly et al., 2013; Pradeepa et al., 2016; T. Zhang et al., 2020). To improve chromatin annotation, several groups have developed chromatin segmentation tools which consider a broader range of epigenomic markers and are therefore better equipped to discriminate enhancers from promoters (Ernst & Kellis, 2017; Hoffman et al., 2012).

Identifying enhancers based on their activity:

Historically, the main method used to identify enhancers was to test candidate sequences for their ability to activate gene expression in reporter assays. These assays range from relatively low throughput luciferase and lacZ assays (Pennacchio et al., 2006; Thorne et al., 2010; Visel et al., 2007), to massively parallel reporter assays

(MPRAs) and other high throughput methods with NGS readouts, such as STARR-seq (C. D. Arnold et al., 2013; Inoue & Ahituv, 2015).

Each of these methods involves cloning candidate enhancer sequences into plasmids, together with specific reporter genes under the control of minimal promoters. Enhancer activity can then be determined based on reporter gene expression (**figure 1.2B**). This is measured by fluorescence (in luciferase assays), staining (in lacZ assays), or either microarrays or NGS (in high throughput assays). The fact that candidate enhancer sequences are assayed outside of their biologically relevant chromatin contexts compromises the reliability and biological relevance of these techniques. Furthermore, these techniques measure a candidate's ability to activate a single promoter sequence, despite the fact that many enhancers exhibit a degree of specificity for particular target gene promoters (X. Li & Noll, 1994; Zabidi et al., 2014). Nevertheless, high throughput reporter assays are very useful for screening large libraries of candidate enhancer sequences.

Identifying enhancers through genetic dissection:

The most reliable method for identifying an enhancer is to remove the candidate sequence from its native locus in otherwise intact cells. Changes in target gene expression can then be measured using techniques such as RT-qPCR, various RNA-sequencing methods, or imaging. This has the advantage of assaying an enhancer's ability to up-regulate its *cognate target gene* in its *native chromatin context.* Additionally, the fact that this can be conducted *in vivo* allows one to investigate the candidate enhancer's activity in the appropriate tissue type and at the appropriate developmental stage. The drawback of this approach is two-fold: one concerns interpretation; if the deletion has no effect, it can only be indicative of inactivity or redundancy. The second is technical; these genetic manipulations are low throughput and expensive.

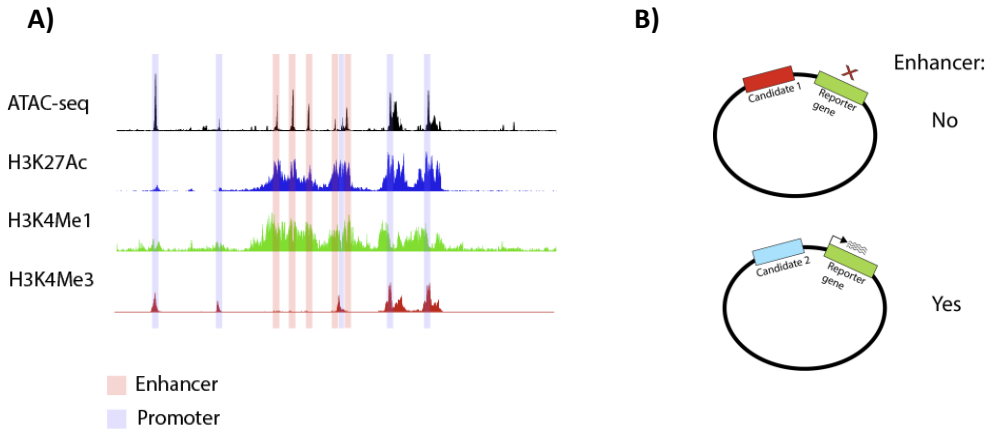


Figure 1.2 Identifying enhancer elements.

Enhancers can be identified genome-wide, using bioinformatic methods. This generally entails subsetting regions enriched for chromatin accessibility, histone 3 lysine 27 acetylation and histone 3 lysine 4 mono-methylation. (B) Transient reporter assays can also be used to identify enhancers. These entail cloning a candidate enhancer into a plasmid containing a reporter gene. If the reporter gene is expressed, one can conclude that the candidate is an active enhancer.

1.2.5 Enhancer clusters:

Most estimates suggest that the human genome contains ~20,000 genes and ~300,000 enhancers. This implies that a large number of genes may be regulated by multiple enhancers. In fact, many tissue-specific genes are regulated by "clusters" of enhancers – groups of elements, often found in close genomic proximity, which regulate the same target gene (Andersson et al., 2014; Grosveld et al., 2021; Long et al., 2016) (**figure 1.3A**).

There has been longstanding debate over whether enhancer clusters are merely collections of independent enhancers, each contributing to gene expression independently of the rest, or whether elements within clusters cooperate with one another to achieve optimal levels of gene activation. The first example of an enhancer cluster was described nearly 40 years ago (Mercola et al., 1983), and since then, many groups have reported different "flavours" of biologically significant enhancer clusters, among them: locus control regions (Grosveld et al., 1987), shadow enhancers (Hong et al., 2008), regulatory archipelagos (Montavon et al., 2011), Greek islands (Markenscoff-Papadimitriou et al., 2014), stretch enhancers (Parker et al., 2013) and superenhancers (Whyte et al., 2013). Many enhancer clusters satisfy the criteria of

multiple classes, and whether each cluster class is functionally distinct from the rest is hotly debated.

1.3 Superenhancers:

1.3.1 What is a superenhancer?

Superenhancers were initially described by the Young lab in 2013 (Whyte et al., 2013). They are dense enhancer clusters, in which constituents are separated by no more than 12.5kb, which bind especially high levels of master TFs, recruit especially high levels of mediator, and are patterned with especially high levels of H3K27Ac (Whyte et al., 2013) (**figure 1.3A**). Superenhancers are often regulators of cell identity genes and are frequently mutated in association with complex traits and genetic diseases (Dębek & Juszczyński, 2022; Harteveld & Higgs, 2010; Higgs et al., 2012; Tang et al., 2020; Yamagata et al., 2020).

Superenhancers are identified using the rank ordering of superenhancers (ROSE) algorithm. This algorithm stitches enhancers within 12.5kb of one another together; it then ranks them based on Med1 and H3K27Ac ChIP-seq data; finally, it separates superenhancers from typical enhancers based on a cut-off value, derived from enhancer rank and ChIP-seq signal (Whyte et al., 2013). 231 superenhancers have been identified in mESCs (Whyte et al., 2013) and 95 in mouse erythroid cells (Hay et al., 2016; Kassouf et al., 2022).

Superenhancers are a particularly controversial class of regulatory element; some groups argue that they are functionally indistinguishable from other enhancer clusters, and others contend that they exhibit emergent properties, marking them as a separate class of element (Blobel et al., 2021; Grosveld et al., 2021; Hnisz et al., 2015; Huang et al., 2018; Moorthy et al., 2017; Whyte et al., 2013). The key area of debate concerns whether the constituent elements comprising superenhancers function independently of one another, or whether they cooperate synergistically to drive particularly high and robust patterns of gene activation (**figure 1.3B**).

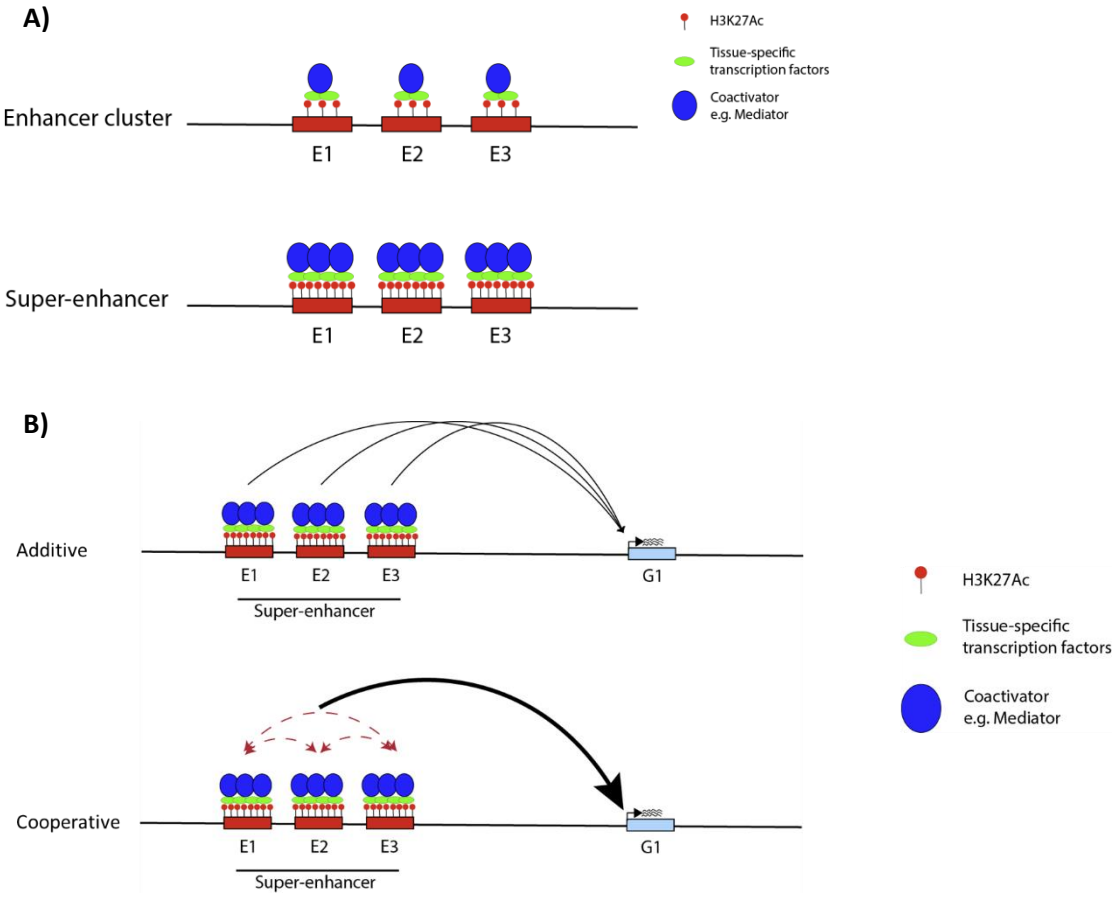


Figure 1.3 Superenhancers are clusters of elements enriched for particularly high levels of H3K27Ac, TFs, and coactivators.

Top: schematic of enhancer cluster. Bottom: schematic of superenhancer. X-axis = linear genome; E1-3 = enhancer elements; G1 = target gene. Red circle = H3K27Ac; Green oval = tissue-specific TFs; Blue oval = coactivators e.g. Mediator. (B) Top: schematic of superenhancer functioning "additively", each enhancer communicates with the target gene independently, with no enhancer-enhancer crosstalk. Bottom: schematic of superenhancer functioning "cooperatively", enhancers communicate with each other, and communicate with the target gene as a collective.

1.3.2 Do the constituents of a superenhancer combine additively or synergistically?

Multiple studies have dissected superenhancers to investigate how they function and the contributions of each of their constituents. These investigations have ranged from testing individual constituent elements in reporter assays, to *in vivo*, *in situ* deletion of constituents. Studies at different loci have presented evidence of additive relationships between constituents, redundant relationships, synergistic, and hierarchical. Experiments dissecting the α- and β-globin clusters (both of which satisfy the criteria to be considered superenhancers), deleting each constituent individually and in informative pairs, suggested that the constituents in these clusters function

independently of one another and that each cluster's capacity to activate expression of its target gene is equal to the sum of its individual components (Bender et al., 2012; Hay et al., 2016).

On the other hand, testing each constituent of the Oct4 superenhancer in luciferase assays (in mESCs) suggested that elements within this cluster combine synergistically, and this was corroborated by *in situ* deletion of each element (Hnisz et al., 2015). Several other studies have presented evidence of non-additive (synergistic and redundant) superenhancer cooperation, including experiments at the Wap cluster (Shin et al., 2016), the Fgf5 cluster (Thomas et al., 2021), the Sox2 cluster (Brosh et al., 2022), and the Fgf8 cluster (Hörnblad et al., 2021).

Hnisz et al (2015), Huang et al (2018), and Thomas et al (2021) characterised several superenhancers in more depth, and presented evidence of crosstalk between superenhancer constituents (Hnisz et al., 2015; Huang et al., 2018; Thomas et al., 2021). These three studies showed that deleting individual superenhancer constituents can cause reductions in H3K27-acetylation over neighbouring constituents, suggesting that superenhancer constituents may be able to "boost" each other's activities. Additionally, Huang et al (2018) separated superenhancer constituents into two categories: hub and non-hub enhancers, and showed that deleting hub enhancers causes a reduction in tissue-specific TF binding at their non-hub neighbours (Huang et al., 2018). Thomas et al (2021) also reported a distinct class of superenhancer constituent, referred to as amplifiers; these are elements which boost the activity of typical enhancers within the same cluster, and they are characterised by high levels of Pol2 recruitment (Thomas et al., 2021). Other studies have shown that deleting superenhancer constituents which perform poorly in luciferase assays (suggesting they have little enhancer activity) from their native loci can have dramatic effects on superenhancer activity, for example, the E8 element within the Pri-miR-290-295 superenhancer (Hnisz et al., 2015) and the HS1 element within the β-globin superenhancer (Hardison et al., 1997; Schübeler et al., 2001 & Bender et al., 2012). This could suggest that these elements play important non-enhancer roles within their cognate superenhancers.

1.3.3 Have superenhancers been rigorously investigated?

The bioinformatic criteria distinguishing superenhancers from other enhancer clusters are purely descriptive, and therefore "superenhancers" may simply represent strong enhancer clusters, with no common mode of cooperation. Genetic dissection of superenhancers to disentangle the relationships between their constituent elements is challenging, as many superenhancers regulate genes which control complex transcriptional and epigenetic programmes. Disruption of such pathways makes it difficult to separate changes in superenhancer-regulated gene expression from associated changes in cell lineage and differentiation. The majority of previous superenhancer dissection studies have drawn conclusions from deleting just one or two constituent elements; these experiments are limited in their capacity to resolve complex cooperation within multipartite superenhancers, especially if constituents cooperate with one another in different manners. Other studies have relied on artificial reporter-based assays divorced from their functionally relevant chromatin contexts.

We speculated that combinatorial reconstruction of a superenhancer from an enhancer-less baseline would allow us to fully dissect and detect even the most complex relationships between each constituent. To do this would require a well characterized, tractable superenhancer in which such comprehensive genetic engineering would be interpretable.

1.4 The murine α-globin cluster as a model mammalian superenhancer:

1.4.1 Existing characterisation of the α-globin locus:

The murine α-globin superenhancer is a cluster of five regulatory elements: R1, R2, R3, R4 and Rm, a mouse-specific element (**figure 1.4A**). The superenhancer up-regulates α-globin gene expression during erythroid differentiation. The α-globin superenhancer, and the locus in which it is situated, has been characterized in detail, from the timing of activation of each constituent element, to the locus' 3D chromatin structure, TF binding repertoire and histone modification profile (**Figure 1.4B**)

(Oudelaar et al., 2021). The α-globin superenhancer is specifically activated during erythroid differentiation, and this process coincides with the formation of the 65kb sub-TAD in which the superenhancer and downstream α-globin genes are located (Oudelaar et al., 2020). This sub-TAD is demarcated by convergently oriented CTCF sites (**Figure 1.4B**); these sites are bound by CTCF in erythroid and non-erythroid tissues; however, the sub-TAD itself is erythroid-specific. This indicates that although the CTCF sites mark the boundaries of the sub-TAD, they are not sufficient to stimulate formation of the domain. Although contentious, a number of studies have recently suggested that the cohesin complex can be loaded at active enhancers (Hua et al., 2021; Rinzema et al., 2021); subsequent bi-directional loop extrusion is thought to facilitate enhancer-promoter interactions and formation/maintenance of loops and TADs. Cohesin loading has been demonstrated at each α-globin superenhancer constituent (Hua et al., 2021). This supports a model in which enhancer activation promotes cohesin loading and domain formation, which is consistent with the fact that these phenomena occur within the same developmental window.

The fact that the α-globin superenhancer is only active in terminally differentiating erythroid cells means that manipulation of the superenhancer is extremely unlikely to affect other tissues. Additionally, sequestration of the active superenhancer within a sub-TAD means that manipulation of the superenhancer is unlikely to impact expression of genes other than α-globin. Lastly, because α-globin has no downstream function in gene regulation, any phenotype arising from genetic manipulation of the superenhancer can be attributed directly to changes in α-globin expression, rather than indirect effects on the expression of other genes. These three features make the α-globin superenhancer an attractive target for investigating superenhancer activity in more detail. It can be thought of as a tractable and isolated system, which can be dissected and studied in a rigorous and controlled manner.

1.4.2 Dissecting the α-globin superenhancer:

The Higgs lab previously dissected the α-globin superenhancer, by removing each constituent element individually and in informative pairs (Hay et al., 2016). These

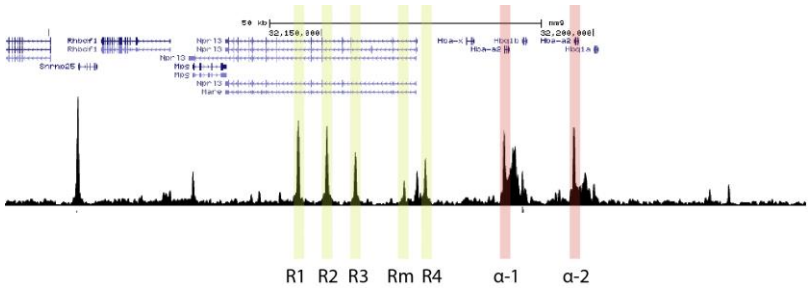
experiments suggested that the five α-globin superenhancer constituents sum as independent elements, and identified R1 and R2 as the two major activators of α-globin transcription (**figure 1.4C**). Despite showing conservation in sequence and synteny over ~70 million years of evolution (Hughes et al., 2005; Philipsen & Hardison, 2018), R3 and R4 displayed little inherent enhancer activity, and the same was true of Rm, the mouse-specific element (Hay et al., 2016).

Enhancer reporter assays supported the *in vivo* element deletion experiments (Hay et al., 2016). Each α-globin superenhancer constituent was tested in transgenic reporter assays, in which vectors containing one of the elements, a minimal promoter and the LacZ gene were stably integrated into the mouse genome. E12.5 embryos were then isolated, followed by LacZ staining. This showed that R2 has the highest enhancer potential, followed by R1, while R3, Rm and R4 were all incapable of inducing lacZ expression (Hay et al., 2016).

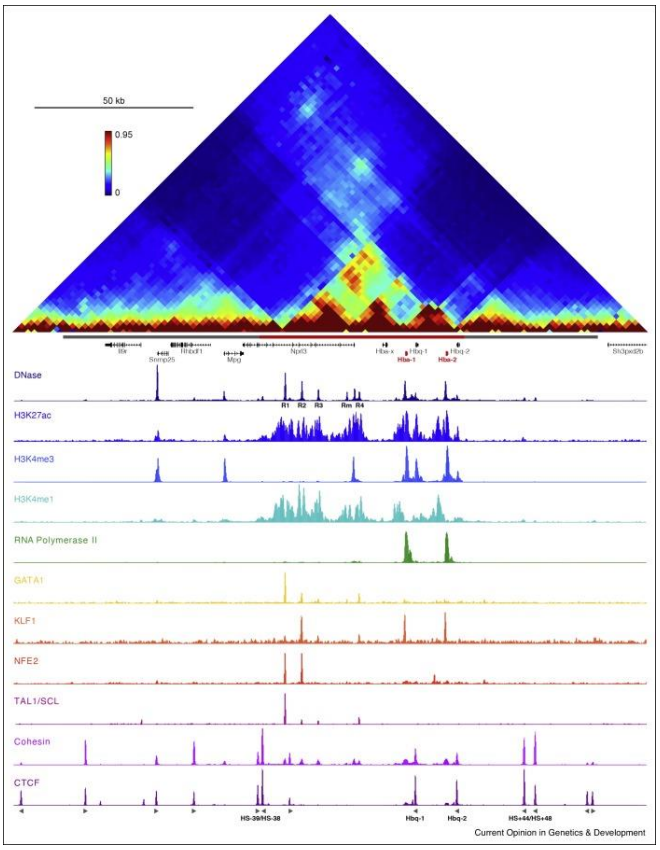
The relative importance of the R2 element is even clearer within the human α-globin superenhancer (which is made up of R1, R2, R3 and R4). Here, R2 contributes >90% to overall α-globin expression (Bernet et al., 1995; Vernimmen et al., 2007; Wallace et al., 2007); moreover, all characterised clinical cases of α-thalassaemia have presented with mutations in the R2 element and/or the α-globin genes themselves (Coelho et al., 2010; Harteveld & Higgs, 2010; Higgs & Wood, 2008; Sollaino et al., 2010; Tamary & Dgany, 2020; Viprakasit et al., 2006). R2 is a stronger enhancer than R1, R3 and R4, and consistently, it shows greater conservation in sequence and synteny throughout evolution (Hughes et al., 2005; Philipsen & Hardison, 2018).

The mouse α-globin superenhancer has been dissected and investigated in great detail. While our previous single element deletions shed light on the necessity of each individual element within the otherwise intact superenhancer, they did not report on each element's sufficiency or the functional relationships between the five constituent elements. We posited that the best test of element sufficiency and cooperation would be to rebuild the superenhancer in a combinatorial manner, within its native locus, after firstly removing all five constituents. In other words, rebuilding the superenhancer from an enhancer-less baseline.

A)

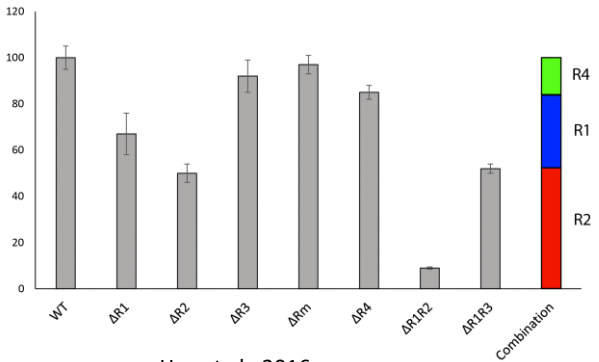


B)



Oudelaar et al., 2021

C)



Hay et al., 2016

Figure 1.4 The mouse α-globin superenhancer.

(A) ATAC-seq track of the murine α-globin locus, with the five regulatory elements (R1, R2, R3, Rm and R4), and the two downstream α-globin genes indicated. (B) Figure from Oudelaar et al., 2021, depicting characterisation of the murine α-globin locus (descending): 3D chromatin structure, DNase-seq, H3K27Ac ChIPs-seq, H3K4Me3 ChIPs-seq, H3K4Me1 ChIPs-seq, Pol2 ChIPs-seq, Gata1 ChIPs-seq, Klf1 ChIPs-seq, Nf-e2 ChIPs-seq, Tal1 ChIPs-seq, Cohesin

ChIPs-seq, CTCF ChIPs-seq, CTCF site orientation. (C) α-globin gene expression from seven mouse models harbouring homozygous single or double enhancer element deletions (adapted from Hay et al., 2016). Far right: contribution of each enhancer element, calculated by subtracting each deletion model from WT.

1.5 Thesis outline:

In this thesis I seek to understand the mode of cooperation between the five regulatory elements of the murine α-globin superenhancer.

Firstly, I ask whether the strongest enhancer of the α-globin superenhancer, R2, is sufficient to activate high levels of α-globin gene expression. To do this, I characterise a previously generated R2-only mouse, in which the R1, R3, Rm and R4 elements have been removed from the native α-globin locus. As yet, few studies have investigated the sufficiency of superenhancer constituents to drive target gene expression within their native loci. Multiple studies have tested superenhancer constituents' activities in artificial reporter assays, and several have dissected superenhancers by removing constituents individually, but hitherto, no studies have assessed individual superenhancer constituents in detail *in vivo,* and within their biologically significant contexts. In my first results chapter (chapter three), I aim to understand whether a superenhancer constituent with strong enhancer activity retains its potential to activate gene expression when separated from all other superenhancer constituents. I test the gene expression output from R2-only primary mouse erythroid cells and subsequently characterise the histone modification profile, TF and coactivator binding repertoire, 3D chromatin structure, and enhancer RNA transcription throughout the α-globin locus.

In my second results chapter (chapter four), I ask how the five α-globin superenhancer constituents cooperate with one another. To do this, I generate an "enhancer titration" series: a series of mutants, rebuilding the native α-globin superenhancer from a "blank canvas" in which all five elements have been removed. All previous *in situ* superenhancer dissection studies have deleted regulatory elements individually and/or in selective pairs. This limits the resolution with which you can draw conclusions. If a multipartite superenhancer contains five constituents which cooperate with one another in additive, redundant and synergistic manners (particularly if a single constituent cooperates differently with each other element),

then single and double element deletions are insufficient to unravel the complex interdependencies manifested within the cluster. Incomplete dissection, in this manner, has hampered previous studies, including our own lab's previous work. In this chapter, I combine efficient CRISPR-based genome editing methods with a high throughput *in vitro* erythroid differentiation system to produce erythroid cells harbouring every informative configuration of the α-globin superenhancer. This allows me to assess each constituent's role within the cluster and its relationship(s) with each other element. This revealed that the α-globin superenhancer is composed of functionally dissimilar constituents: canonical enhancers (R1 and R2) and elements with no intrinsic enhancer activity (R3, Rm and R4), which we named "facilitators" owing to their role in facilitating the full potential of canonical enhancers.

The three facilitators were inequivalent in their abilities to potentiate R1 and R2. My final results chapter (chapter 5) explores whether the hierarchy among the three α-globin facilitators is predominantly dependent on each element's position or its sequence composition. This entails the generation of additional genetic models, re-positioning elements within the native α-globin locus, to test whether this affects their abilities to potentiate the R1 and R2 enhancers, and increase α-globin expression.

My work shows that the α-globin superenhancer is indeed stronger than the sum of its parts. I show that the superenhancer is composed of two functionally distinct element classes: enhancers, which directly activate gene expression; and facilitators, which have no intrinsic enhancer activity, but instead potentiate canonical enhancers within their cognate superenhancer. I show that without facilitators, the strongest enhancer within the α-globin superenhancer (R2), is unable to recruit high levels of transcriptional coactivators, interact with the α-globin promoters with high frequency and transcribe high levels of enhancer RNA. Moreover, without facilitators, R2 cannot activate α-globin transcription to the expected degree. Crucially, this work demonstrates that the α-globin superenhancer does manifest emergent properties.

Chapter two: Methods:

2.1 Generating genetic models:

2.1.1 BAC engineering:

BAC design and synthesis:

BACs were synthesised from a WT template construct containing the lox sites and hypoxanthine phosphoribosyltransferase (Hprt) split gene required for RMGR (Wallace et al., 2007). Dr Christian Babbs generated the WT template construct from a mouse BAC (RP23-46918; BACPAC Resources Centre, Children's Hospital Oakland Research Institute; (Osoegawa et al., 2000)), by sequential shearing and lox integration steps using λ-red-mediated recombinase (Y. Zhang et al., 1998). The Lox sites were integrated 85kb apart, flanking the α-globin locus. Exon 4 of the Rhbdf gene at the 5' of the locus was missing in the WT BAC template, and so this was introduced into the construct, using synthetic fragments. Similarly, integration of synthetic fragments into the construct was used to introduce enhancer deletions and insertions in the appropriate models.

Synthetic BAC generation:

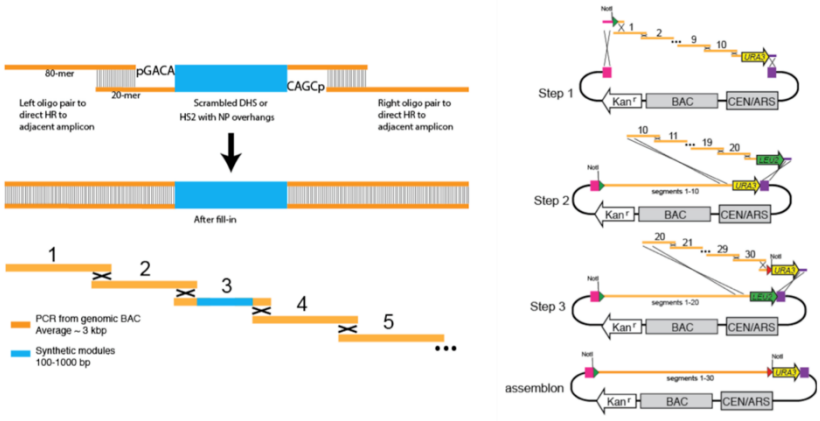
Two RMGR-ready versions of the α-globin locus, the first encoding the five enhancer elements and the second deleting all but the R2 element, were constructed. A previously constructed BAC spanning the α-globin locus plus RMGR parts (Wallace et al., 2007) (RP23-46918; BACPAC Resources Centre, Children's Hospital Oakland Research Institute; (Osoegawa et al., 2000)) was used as template to generate PCR amplicons with 50-200 base pairs of overlapping sequence for yeast homologous recombination. Gblocks (IDT) or fusion PCR products were used to provide homology with non-overlapping adjacent segments (e.g. enhancer deletions, vector-adjacent amplicons). A variant of the eSwAP-In method (Mitchell et al., 2021) was used to produce the two constructs, which were sequence verified using Illumina short read sequencing (**figure 2.1A**).

BAC transfection:

Helena Francis, a previous DPhil student, conducted all BAC transfection steps. Briefly, purified BAC DNA was co-transfected with a Pcaggs-Cre-IRES*puro* plasmid (A. J. H. Smith et al., 2002) into RMGR-competent mESCs by lipofection. Cells were selected for *Hprt* complementation over a one-two week period. The *Hprt* gene was later removed by transfection with a transient flippase (Flp)-expressing plasmid (Schaft et al., 2001), followed by screening, by selection with 6-thioguanine and PCR (Helena Francis, thesis).

The structural integrity of the R2-only BAC-integrated locus was screened using 10X linked-read sequencing (**figure 2.1B**). All BAC-integrated loci were further screened by PCR, Sanger sequencing, and ATAC-seq (Helena Francis, thesis).

A)



B)

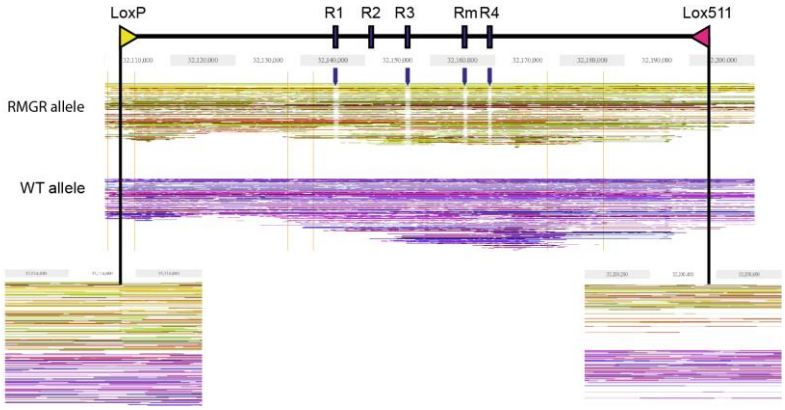


Figure 2.1 R2-only BAC generation and RMGR.

(A) Figure adapted from Mitchell et al., 2021. Short segments of DNA are generated by PCR from a target backbone or ordered as synthetic constructs with matching overhangs to allow for homologous recombination. Homologous recombination is performed in yeast with blocks of up to 10 segments in a single step. Alternating selection for the URA3 and LEU2 genes is used to perform subsequent assembly steps. The resulting BAC contains the assembled locus, modules necessary for replication in both yeast and E. Coli (BAC / CEN/ARS, respectively) and selectable markers including the kanamycin resistance gene (Kanr). (B) Figure adapted from Helena Francis thesis. Top: schematic of the RMGR region with enhancer elements and RMGR lox exchange sites annotated. Middle: phased linked-reads for the R2-only RMGR and WT alleles, visualised with proprietary Loupe software from 10X Genomics. Enhancer element deletions are clearly visible by sequencing dropout in the RMGR allele, as indicated by arrows. Bottom: junctions where lox sites remain in the R2-only allele are seen as mapped read ends in the RMGR allele, but not in the WT allele where there are no lox sites.

2.1.2 CRISPR-Cas9 editing:

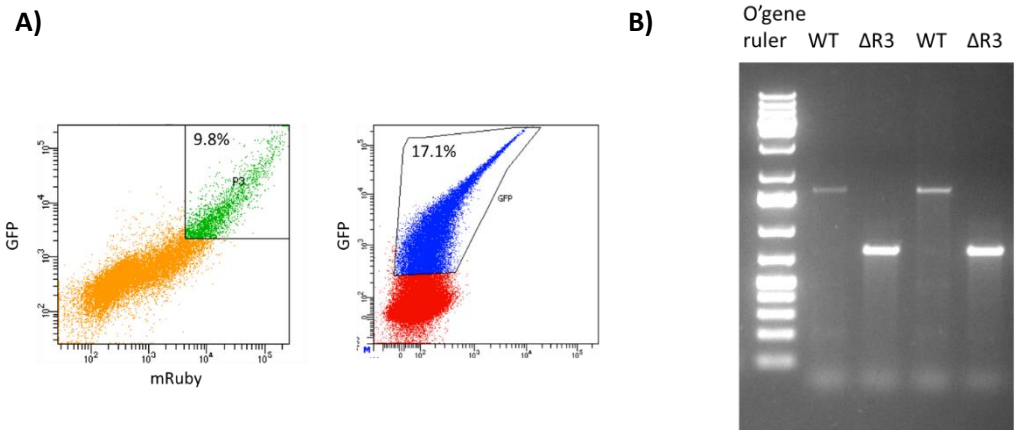
The initial steps of guide RNA design and cloning into vectors were performed by Philip Hublitz and the Weatherall Institute of Molecular Medicine (WIMM) genome engineering facility. Briefly, guide RNAs were designed using the CRISPOR and BreakingCas online gRNA design tools. Guides with the fewest predicted off-target effects were selected, and any guides with predicted genic off-targets were discarded. The three best candidate gRNAs were screened by surveyor assay (testing in vitro cleavage activity, according to the manufacturer, IDT), and the most efficient gRNAs were cloned into a pSpCas9 (BB)-2A-GFP (pX458) vector, a gift from Feng Zhang (Addgene plasmid: #48138), or a modified vector with the GFP tag exchanged for mRuby (pX458-ruby; (Kredel et al., 2009)).

To generate enhancer deletion models, gRNAs were designed to target the 5’ and 3’ of the enhancer targeted for deletion, matching as closely as possible the coordinates previously used to generate *in vivo* enhancer deletions (Hay et al., 2016). Each gRNA was cloned into a single pX458 vector; the 5’-targeting gRNA into a GFP-expressing plasmid, and the 3’-targeting gRNA into an mRuby-expressing plasmid (**table 2.1**).

Helena Francis previously generated an mESC line, in which one copy of the 86kb α-globin locus was deleted using a dual-guide CRISPR deletion strategy in RMGR-competent mESCs, derived from E14-TG2a.IV (E14) mESCs, as outlined above (creating a line hemizygous at the α-globin locus), and in which the second α-globin locus allele is exchanged by RMGR with a BAC-derived, WT α-globin locus (Helena Francis, thesis). This line was termed “hemizygous WT”. These hemizygous WT cells were co-

transfected, by lipofection, with the 5'-targeting vector (expressing GFP) and the 3'-targeting vector (expressing mRuby), and 24-36 hours later, GFP-mRuby co-fluorescent cells were FACS sorted into individual wells of a 96 well plate (**figure 2.2A**). Individual clones were grown in each well for 8-10 days without disruption. When colonies were visible in each well, cells were split into two plates: one for screening, and the other for analysis/freezing. Clones were screened for successful enhancer deletion by PCR; this entailed PCR amplification using primers (**table 2.2**) flanking each deleted element, such that a successfully deleted allele would produce a much smaller product than a WT allele (**figure 2.2B**). Onetaq 2X Master Mix with Standard Buffer (New England Biolabs) supplemented with 0.75% DMSO was used for PCR screening (**table 2.3**; **table 2.4**). Successful clones were further screened by Sanger sequencing, and later, differentiated clones were screened by ATAC-seq, to ensure the overall integrity of the edited α-globin locus. Entire process outlined in **figure 2.2C**.

At least three separate clones were generated for each single enhancer deletion model (ΔR3, ΔRm and ΔR4). Later, when generating double deletion models (ΔR3Rm, ΔR3R4 and ΔRmR4, or "R1R2R4", "R1R2Rm", "R1R2R3", respectively), the single deletion clones were re-targeted for deletion of a second enhancer element e.g. Rm was deleted in the ΔR3 background, producing ΔR3Rm/R1R2R4. Model generation and screening during the second round of targeting were performed identically to the first round.



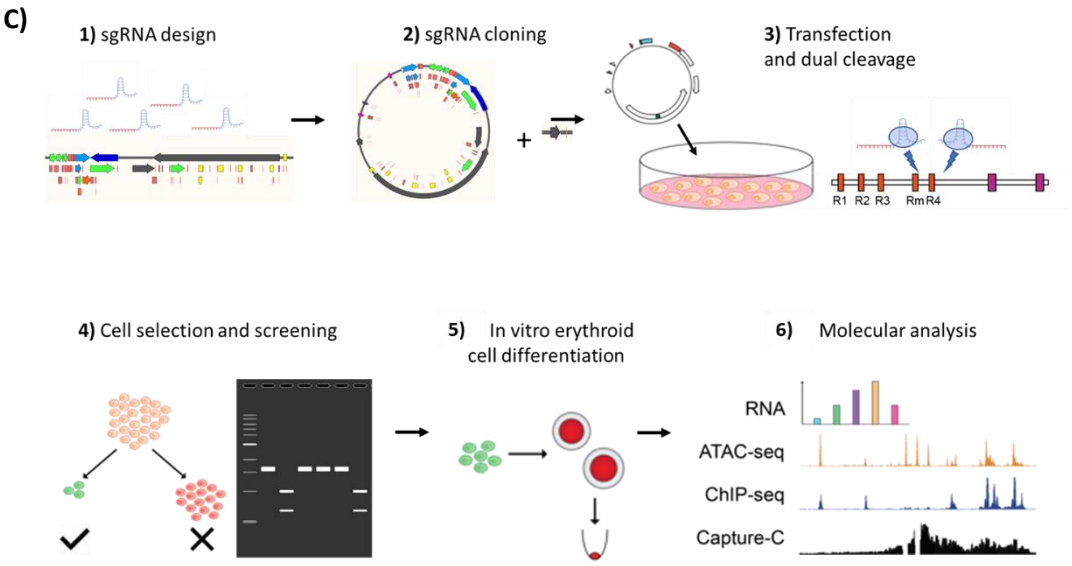


Figure 2.2 CRISPR editing to generate the enhancer titration series.

(A) Representative FACS plots from enhancer deletion targeting (transfected with mRuby and GFP-expressing plasmids (left)), and from HDR insertion targeting (transfected with GFP-expressing plasmid (right). (B) Representative genotyping gel. PCR conducted on hemizygous mESCs targeted for R3 deletion, with primers flanking the R3 element. PCR conducted using Onetaq 2X Master Mix with Standard Buffer (New England Biolabs). Left-right: O'generuler DNA ladder, WT PCR product (intact R3 element), R3 deletion product, WT PCR product (intact R3 element), R3 deletion product. (C) Schematic of genome-engineering workflow: 1) Design of sgRNAs flanking each element targeted for deletion (using CRISPOR and BreakingCas tools); 2) Cloning sgRNAs into pX458 vectors, which also express S. pyogenes Cas9, and GFP or Ruby; 3) Lipofection of hemizygous mESCs; 4) FACS sorting of GFP-Ruby co-fluorescent cells, followed by PCR genotyping and Sanger sequencing; in vitro differentiation of successfully edited clones, using EB differentiation and erythroid purification system (Francis et al., 2022); Molecular analysis of erythroid cells of each genotype.

To produce R2[R4] and R1R2R3[R4] mutants, existing SE-KO (in which all 5 regulatory elements have been removed, produced by Helena Francis) and R1R2-only (in which the R3, Rm and R4 elements have been removed, produced by Helena Francis) mutants were re-targeted. Each new model was generated using a single round of targeting – either through insertion of the R2 element at the position of R4 in SE-KO cells, or insertion of the R3 element in the position of R4 in R1R2-only cells (**figure 2.3A**; **figure 2.3B**). Homology directed repair (HDR) donors were designed encoding the R2 element, with 500bp homology arms, homologous to the native position of the R4 element. A Sal1 restriction enzyme recognition site was inserted at the 5' of the R2 enhancer in the HDR donor, and an Mlu1 site at the 3' of the element. This enabled efficient restriction-ligation exchange of the R2 element within the HDR donor with the R3 element. The HDR donor construct was ordered as a GeneART Gene synthesis custom design. The HDR donor was also designed to inactivate the protospacer

adjacent motif. R2[R4] and R1R2R3[R4] donors were screened using the online sasquatch and jaspar tools to ensure no novel accessibility sites or motifs were predicted (Fornes et al., 2020; Schwessinger et al., 2017), prior to synthesis and transfection.

Philip Hublitz and the WIMM genome engineering facility designed a gRNA targeting cleavage near the site for R2/R3 integration, just upstream of the position at which R4 was previously located (**table 2.1**). This was cloned into a GFP-expressing pX458 vector. Cells were co-transfected with the gRNA vector and the HDR-donor vector and GFP-expressing cells were FACS sorted into individual wells of a 96 well plate. Growth and screening was performed in the same manner as in enhancer deletion clones.

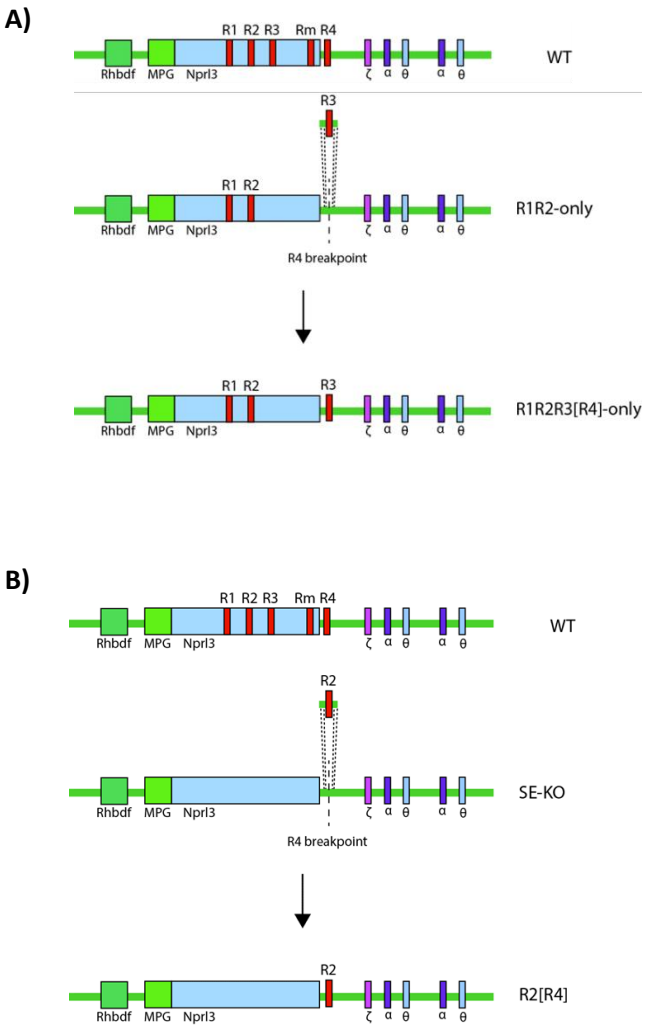


Figure 2.3 CRISPR HDR to generate enhancer rescue models.

(A) Schematic of generation of the R1R2R3[R4] model. Hemizygous R1R2-only cells lipofected with pX458 expressing an sgRNA targeting the R4 deletion breakpoint and an HDR vector consisting of the R3 element flanked by 500bp homology arms complementary to sequences flanking the R4 deletion breakpoint. Clones FACS sorted based on GFP-fluorescence and genotyped by PCR and Sanger sequencing. (B) Schematic of generation of the R2[R4] model. Hemizygous SE-KO cells lipofected with pX458 expressing an sgRNA targeting the R4 deletion breakpoint and an HDR vector consisting of the R2 element flanked by 500bp homology arms complementary to sequences flanking the R4 deletion breakpoint. Clones FACS sorted based on GFP-fluorescence and genotyped by PCR and Sanger sequencing.

Table 2.1 Guide RNA sequences used to engineer enhancer deletion clones and HDR insertion clones.

gRNA	Sequence
R3 5’	GGACTGAGGGAATTTTGTACA
R3 3’	GGACTGGCAGAAAGCTATGT
Rm 5’	GGACCAGCGTAGTCTAACTCC
Rm 3’	GGACTCAACCACATGACTCA
R4 5’	GGTGCATCACCTCAACTCCCA
R4 3’	GGCTAGCTGAATAATTTGCGG
R4 HDR	GGATCCCAGGTTAGTTAGTCC

2.1.3 Mouse model generation:

All mouse work was performed in accordance with UK Home office regulations, under the appropriate animal licenses. Mouse model generation and animal husbandry (including sacrificing mice for analysis) were conducted by the WIMM Mouse Transgenics Core Facility. R2-only BAC-integrated mESCs were karyotyped and microinjected into C57BL/6 blastocysts, before being implanted into pseudo-pregnant females. Chimeric males were back-crossed with WT females, and pups were screened by PCR for germline transmission.

Only one surviving homozygote was obtained from fifteen heterozygote crosses, which confined all further analysis of the R2-only homozygous phenotype to embryonic time points. Timed-heterozygote crosses were established, and pregnant mice were sacrificed at various stages of gestation: embryonic days 8.5, 9.5, 10.5, 12.5, 14.5 and 17.5. Embryos were dissected from the pregnant females, ordered based on their predicted genotypes, and photographed (**figure 2.4A, left**). Erythropoietic cells/compartments were then isolated for analysis. Whole E8.5-10.5 embryos were mechanically disaggregated in heparinised PBS, and erythroid-containing supernatant aspirated into fresh tubes for processing; remaining material was stored for genotyping by PCR (**figure 2.4A, right**). Foetal livers (the definitive erythroid

compartment) were isolated from E12.5-E17.5 embryos (**figure 2.4A, left, bottom**). Foetal livers were mechanically disaggregated to a single cell suspension in FACS buffer, and filtered through pre-separation filters (miltenyibiotec); brain tissue was stored for genotyping by PCR and gene expression analysis (RT-qPCR). Erythroid cells were processed for analysis by RT-qPCR/RNA-seq, ATAC-seq, ChIP/ChIPmentation and 3C-based methods on the day of harvest (below). FACS analysis, staining for the CD71 and Ter119 cell surface markers in E12.5 foetal liver cells, revealed that WT and R2-only foetal livers are composed of ~95% CD71+/Ter119+ erythroid cells, indicating that no further selection (beyond mechanical disaggregation, and filtration through pre-separation filters) was required (**figure 2.4B**).

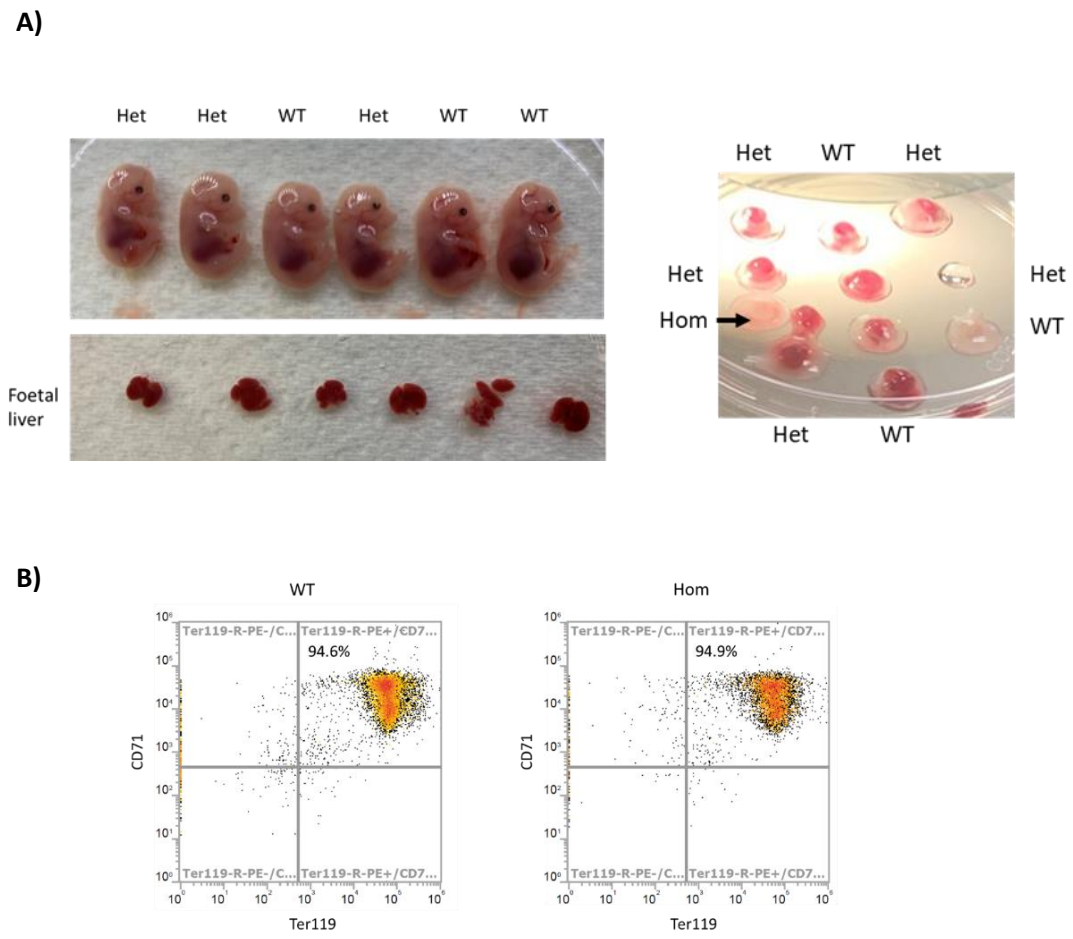


Figure 2.4 Harvesting embryonic material from R2-only heterozygote crosses.

(A) E12.5 embryos (left, top) and isolated foetal livers from the corresponding embryos (left, bottom); E10.5 embryos in heparinised PBS (right). (B) Representative FACS plots, staining for CD71 (y-axis) and Ter119 (x-axis) erythroid-associated cell surface markers, from E12.5 foetal liver erythroid populations. WT and R2-only homozygote populations contain ~95% CD71+/Ter119+ erythroid cells.

Pups and embryos were genotyped using two separate PCRs, performed on stored non-erythroid tissue. Ear notches from live pups, or brain tissue from embryos were lysed overnight at 55°C in DIRECTPCR-EAR PEQGOLD or DirectPCR Lysis-Reagent Cell, respectively, with 500ug/ml proteinase K. The following morning, proteinase K was inactivated at 95°C for 10 minutes, followed by spinning samples down and proceeding with PCR on the lysate. IMMOLASE DNA polymerase (Bioline) was utilised for genotyping (**table 2.2**; **table 2.3**). The first PCR utilised primers flanking the R4 element; PCR over the R2-only allele produced a ~1000bp product, whereas over WT alleles, the product was much larger and often too faint to see. The second PCR used primers complementary to regions within the R1 element; PCR over the R2-only allele did not produce a product, whereas over the WT allele, a ~1000bp product was produced (**table 2.4**).

Table 2.2 IMMOLASE DNA polymerase (Bioline) protocol.

Ingredient	**Volume for 20ul PCR (ul)**
10X ImmoBuffer	2
MgCl2 (50mM)	0.6
Combined F&R primer mix (10uM total concentration)	2
dNTPs (2mM each)	2
Immolase DNA polymerase	0.2
Betaine (5M)	4
Water	6.2

Table 2.3 IMMOLASE DNA polymerase (Bioline) cycling conditions.

Step	Temperature	Time
Initial denaturation	95°C	10 minutes
30 cycles	94°C 60°C 72°C	30 seconds 30 seconds 1 minute
Final extension	72°C	10 minutes
Hold	4°C	∞

Table 2.4 Genotyping primers used to genotype enhancer deletion clones, HDR insertion clones, and R2-only mice/embryos.

Target	**Sequence (F/R)**
R3 boundary PCR	TGTGAAATCACCAGAATTACAGG/ TGTGGGTCAGGCCTCTTAGAGGG
Rm boundary PCR	TAGGGAAAGGTTATGTGAACTGC/ TCAGGCTGGGCACTGGCTCTGCC
R4 boundary PCR	TTTGCCAGCAACCTTCACAGGAG/ ATTTTAAAGGCATTGCAGAGCCG
R2-only mutant PCR	AGCTCTGAAATGGACGTGGT/ GAGCCGTCCTGTGTATGTGA
R2-only WT PCR	GGGCACAGCAAAAGAGGAAA/ AATGGTCCTTGTCCTCCCAG

2.2 Cell culture and in vitro erythroid differentiation system:

Hemizygous WT mESCs, or genetic models derived from these cells, were cultured in gelatinised plates using standard methods (Nichols et al., 1990; A. G. Smith, 1991): cells grown in ES-complete medium, a GMEM-based medium supplemented with Leukemia inhibitory factor (LIF), an inhibitor of differentiation, and therefore used to maintain cell potency.

An *in vitro* embryoid body-based differentiation system was utilised to generate erythroid-like cells (Francis et al., 2022) for analysis. Briefly, 24 hours prior to differentiation, mESCs were passaged into "adaptation" medium (15% ΔFCS, 1000 U/ml Lif, in Iscove's modified Dulbecco's medium (IMDM)). Cells were transferred into 10cm petri dishes containing differentiation media, lacking LIF, and containing transferrin, to "encourage" cells down the erythroid lineage. Cells were cultured in differentiation medium for seven days, shaking the petri dishes every 1-2 days to avoid EBs sticking to the plasticware.

After seven days of differentiation, CD71+ erythroid cells were selected and isolated. EBs were disaggregated into single cell suspension, through incubation in 0.25% trypsin for ~3 minutes, and then quenched with FCS-containing media. Cells were incubated with anti-CD71 FITC-conjugated antibodies (Invitrogen, eBioscience 11-0711-85), followed by anti-FITC separation microbeads. CD71+ cells were then isolated by magnetic column separation (LS Column, Miltenyi), according to the manufacturer's protocol (**figure 2.5A**). CD71+ cells were processed for a range of molecular analyses, as outlined below.

A)

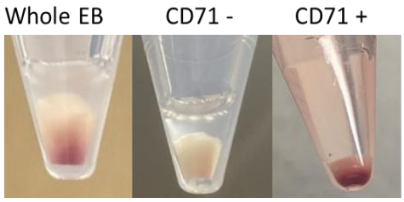


Figure 2.5 Harvesting day 7 erythroid cells from EB differentiations

EB differentiation: Total EB post-trypsinisation/disaggregation (left); CD71- post-column separation (middle); CD71+ post-column separation (right).

2.3 Gene expression analysis:

On the day of cell harvest, aliquots of $5x10^5$ cells (primary mouse cells, or CD71+ mESC-derived models) were lysed in trizol reagent, before being snap frozen and stored at -80°C. RNA was extracted using the Direct-zol MicroPrep kit (Zymo Research), according to the manufacturer's protocol (however, the 15 minute DNase treatment step was lengthened to 45 minutes). RNA quality was assessed by tape station, using RNA screentape (Agilent). Only samples with an RNA integrity score of at least 8 were taken forwards for subsequent analysis (**figure 2.6A**). The extracted RNA was reverse transcribed using Superscript III First-Strand Synthesis SuperMix (Life Technologies).

qPCR with taqman probes (**table 2.5**) was primarily utilised to analyse gene expression in each model (results were normalised to RPS18 or the relevant β-globin genes). Exon-spanning probes were used where possible, to detect mature mRNA; "no template" and "RT-negative" controls were conducted alongside each experiment. Analysis steps including all statistical tests (ANOVA) and graphical plotting were conducted in RStudio. The R package ggplot2 (Wickham, 2016) was used to generate each plot.

RNA-seq was also used to analyse gene expression in WT and R2-only E12.5 foetal liver cells (below)

A)

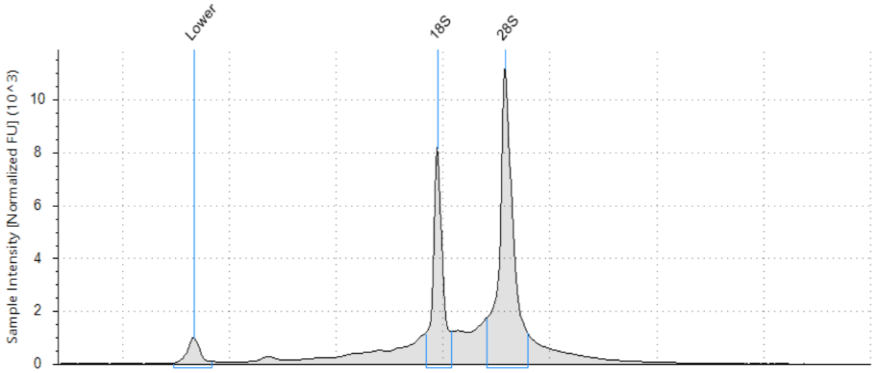


Figure 2.6 Assessing RNA quality by tape station.

(A) Example RNA screen tape; RIN: 9.0.

Table 2.5 Taqman assay IDs used in RT-qPCR gene expression analysis.

Target	Assay ID
Hba-a1/a2	Custom (IDT)
Hba-x	Mm00439255_m1
Hbb-bh1	Mm00433932_g1
Hbb-bh2	Mm01273444_g1
Hbb-bt/bs/b1/b2	Custom (IDT)
Hbb-y	Mm00433936_g1
Hbq1a	Mm00731011_s1
Hbq1b	Mm02747875_s1
Il9r	Mm00434313_m1
Mpg	Mm00447872_m1
Nprl3	Mm01193449_m1
Rhbdf1	Mm00711711_m1
Snrnp25	Mm00547218_m1

2.4 NGS assays:

2.4.1 ATAC-seq:

Assay for transposase-accessible chromatin (ATAC)-seq was performed on ~7x10^4 cells, using the Illumina Tagment DNA enzyme and buffer kit (illumina), as previously described (Buenrostro et al., 2015; Hay et al., 2016). Briefly, cells were lysed in ATAC lysis buffer (**table 2.6**) to maintain nuclear integrity, and resuspended in Tn5 buffer with illumina adaptor-loaded Tn5 enzyme. Cells were incubated for 30 minutes at 37°C, and then tagmented DNA was purified using XP Ampure beads (mybeckman), before indexing with Nextera indexing primers (**table 2.7**) and NEBNext 2X High fidelity

mastermix (New England Biolabs; **table 2.8**; **table 2.9**). Indexed ATAC samples were assessed by tape station, using a D1000 HS screen tape (Agilent) (**figure 2.7A**).

Table 2.6 ATAC lysis buffer

Component	Final conc.	Starting conc.	Volume
Tris-HCl, pH 7.4	10 mM	1 M	100 ul
NaCl	10 mM	5 M	20 ul
$MgCl_2$	3 mM	1 M	30 ul
NP40	0.1%	10%	100ul
H_2O	-	-	9.75 ml

Table 2.7 Nextera indexing primer sequences.

Index	Sequence
Ad1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT

Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

Table 2.8 NEBNext High fidelity PCR.

Component	**Volume per 50 ul RXN**
NEBNext High Fidelity 2x PCR Master Mix	25 µl
10 µM Forward Primer	2.5 µl
10 µM Reverse Primer	2.5 µl
Template DNA*	Variable
Nuclease-free water	To 50 µl

Table 2.9 NEBNext High Fidelity PCR cycling conditions

Step	**Temp**	**Time**	**Cycles**
Nick repair	72°C		
Initial Denaturation	98°C	30 seconds	1
Denaturation	98°C	10 seconds	9-11
Annealing	63°C	30 seconds	
Extension	72°C	1 minute	
Hold	4-10°C	∞	1

A)

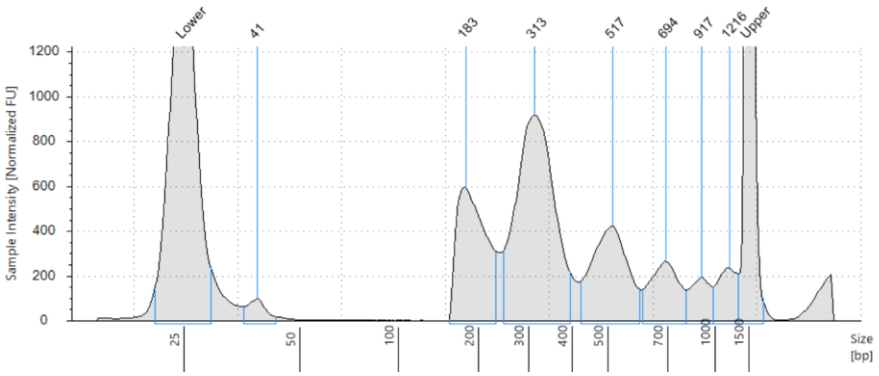


Figure 2.7 Assessing ATAC library quality by tape station.

A) Example tape station trace for ATAC sample (post-tagmentation and indexing).

2.4.2 ChIPmentation:

ChIPmentation experiments were performed as previously described (Schmidl et al., 2015), with modifications as described. Briefly, on the day of cell harvest, aliquots of $1x10^5$-$1x10^6$ cells (primary mouse cells, or CD71+ mESC-derived models) were either single-fixed with 1% formaldehyde for 10 minutes, followed by quenching with 125mM glycine, or double-fixed with 2mM disuccinimidyl glutarate (DSG) for 50 minutes, followed by 1% formaldehyde for 10 minutes, before quenching with 125mM Glycine. Single-fixed samples were ultimately used for ChIPmentation experiments assaying histone modifications, whereas double-fixed samples were used for experiments assaying TF occupancy.

Cells were spun down and washed with PBS, before being snap frozen. Fixed aliquots were stored at -80°C. Cell pellets were lysed in a 0.5% SDS lysis buffer and sonicated, using a covaris ME220 sonicator, to fragment DNA to an average fragment length of ~2-300bp. Sonicated chromatin was analysed by tape station, using a D1000 or D1000 HS screen tape (Agilent) (**figure 2.8A**). SDS in the lysis buffer was neutralised with 1% triton-X, and the sonicate was incubated overnight with a mix of protein A and G dynabeads (Thermofisher) and the appropriate antibody (**table 2.10**). The following morning, chromatin-bound beads were washed three times using a low salt, a high salt and a LiCl-containing wash buffer, followed by tagmentation of the

immunoprecipitated chromatin with sequencing adaptor-loaded tn5, before indexing, using Nextera indices (**table 2.7**) and NEBNext 2X High fidelity mastermix (New England Biolabs; **table 2.8**; **table 2.9**).

A)

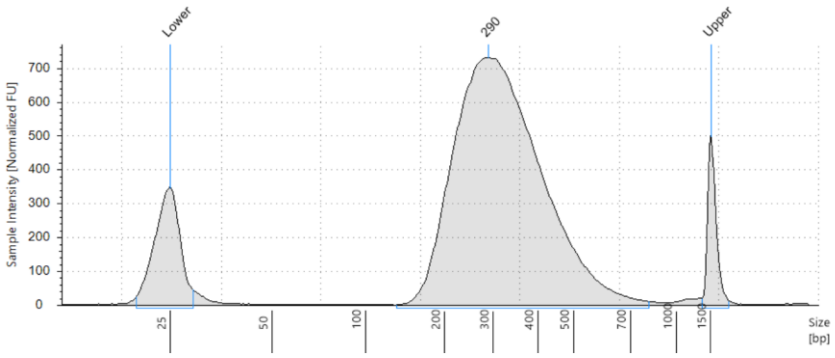


Figure 2.1 Assessing sonicated chromatin (for ChIPmentation) by tape station.
(A) Example tape station trace for sonicated chromatin (for ChIPmentation).

Table 10 Antibodies used for ChIPmentation.

Antibody	ID
H3K27ac	Abcam ab4729
H3K4me1	Abcam ab8895
H3K4me3	Abcam 8580
Gata1	Abcam ab11852
Nf-e2	Santa Cruz sc-22827
Med1	Bethyl IHC-00149
Brd4	Bethyl IHC-00396
Rad21	Abcam Ab992
CTCF	Cambridge Bioscience 61312
PolII	Santa Cruz sc-899

2.4.3 Tiled-C:

Tiled-C was conducted as previously described (Oudelaar et al., 2020). Briefly, on the day of harvest, aliquots of $5x10^5$ cells (primary mouse foetal liver cells) were fixed with 2% formaldehyde for 10 minutes, before quenching with 125mM Glycine. Cells were spun down, washed with PBS, and the pellet suspended in a mild NP-40-containing lysis buffer. Samples were then snap frozen and stored at -80°C. Cells in lysis buffer were thawed and spun down, before resuspension in restriction enzyme buffer mix. An appropriate volume of DpnII was added, and samples were incubated overnight at 37°C. Fresh aliquots of DpnII were added the following morning and afternoon. DpnII was heat inactivated, and proximal DpnII-digested "sticky ends" were ligated using T4 ligase. Digested-re-ligated DNA was extracted using XP Ampure beads (mybeckman)

and sonicated using a covaris ME220 sonicator. Sonicated chromatin was analysed by tape station, using a D1000 or D1000 HS screen tape (Agilent) (**figure 2.9A**). The resultant fragments were indexed using the NEBNext Ultra II library preparation kit (New England BioLabs). Fragments corresponding to the region of interest (chr11:29902951-33226736) were enriched using oligo capture with biotinylated oligos (Twist) complementary to every DpnII fragment within the tiled region, before streptavidin pulldown using Streptavidin dynabeads (Thermofisher).

A)

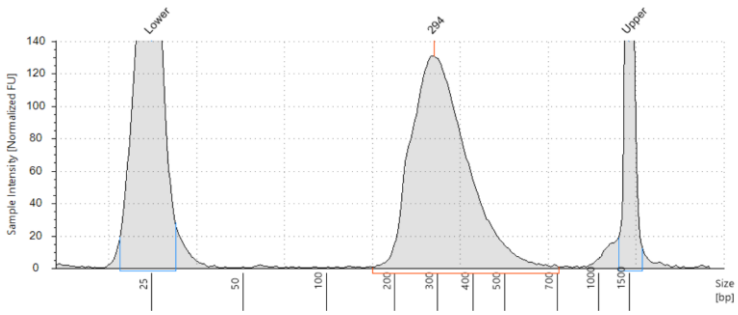


Figure 2.2 Assessing sonicated chromatin (for Tiled-C) by tape station.

(A) Example tape station trace for sonicated chromatin (for Tiled-C, post-digestion/re-ligation).

2.4.4 RNA-seq:

On the day of harvest, aliquots of $5x10^5$ cells (primary mouse foetal liver cells) were lysed in trizol reagent, before being snap frozen and stored at -80°C. RNA was extracted using the Direct-zol MicroPrep kit (Zymo Research), according to the manufacturer's protocol (however, the 15 minute DNase treatment step was lengthened to 45 minutes). RNA quality was assessed by tape station, using RNA screentape (Agilent). Only samples with an RNA integrity score of at least 8 were used (**figure 2.6A**).

Poly-A positive and negative RNA-seq was performed on $5x10^5$ cells, using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England BioLabs). Ribosomal RNA was depleted using the NEBNext rRNA Depletion Kit, and then the poly-A positive and negative fractions were separated using the NEBNext Poly(A) mRNA Magnetic Isolation Module.

2.5 Sequencing and bioinformatic analysis:

All NGS was performed using an illumina NextSeq 500 machine, with TG NSQ 500/550 Hi Output v2.5 (75 CYS) kits (illumina); these kits are paired-end sequencing kits which produce two 40 base pair reads, corresponding to the 5' and the 3' of the fragment being sequenced. Generally, ~25-40 million reads were desirable for each ATAC or ChIPmentation sample, ~10-20 million for each RNA-seq sample, and ~5-10 million for each Tiled-C sample, although actual sequencing depth was variable.

2.5.1 ATAC and ChIPmentation:

The quality of the FASTQ files from ATAC-seq and ChIPmentation were assessed using FASTQC (*Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data*, n.d.), and the reads aligned to the mm9 mouse genome, using bowtie2 (Langmead & Salzberg, 2012). Non-aligning reads were trimmed using Cutadapt trimgalore (Martin, 2011) and then realigned to the mm9 genome using bowtie2. All reads which still failed to align were extracted, and flashed using FLASh (Magoč et al., 2011), before realignment to the mm9 genome using bowtie2. All of the files containing successfully aligning reads were concatenated, and aligned to the mm9 genome together using bowtie2. Resultant SAM files were filtered, sorted, and PCR duplicates removed, using SAMtools (samtools view, sort, and rmdup, respectively) (H. Li et al., 2009). The resultant BAM file was indexed using SAMtools index, and converted to a bigwig file using deeptools bamcoverage (Ramírez et al., 2016). Merged bigwig files (averages of multiple biological replicates) were generated using deeptools bigWigCompare. Each bigwig was visualised using the University of California Santa Cruz (UCSC) genome browser (Kent et al., 2002), and traces corresponding to regions of interest were downloaded from there. Peaks were called in each sample using macs2 (Y. Zhang et al., 2008) with default parameters, and differential accessibility/binding analysis was conducted using Bioconductor DESeq2 in RStudio (Love et al., 2014). Generation of consensus peak files from multiple biological replicates was performed using bedtools intersect, and analysis of

overlapping peaks/peak distances was performed using bedtools intersect and bedtools closest (Quinlan & Hall, 2010). Motif analysis was performed using the MEME suite (meme-chip for de novo motif analysis and fimo for finding occurrences of known motifs) (Bailey et al., 2015), using HOCOMOCO mouse position weight matrices. Principal component analysis was performed on ATAC samples, using the DiffBind package in RStudio (Stark & Brown).

2.5.2 Tiled-C:

Tiled-C samples were analysed using the HiC-Pro pipeline (Servant et al., 2015), using the capture Hi-C workflow (aligning the data to the mm9 genome). To avoid interaction bias between regions within and outside of the tiled region, all data mapping to the tiled region was extracted and the remaining data discarded from subsequent analysis steps. Interaction matrices were ICE-normalised using HiC-Pro; this effectively makes all interactions "equally visible" by correcting for probability of interaction between two bins. Without normalization, matrices show extreme bias for local chromatin interactions, making it difficult to investigate long-range interactions. Interaction heatmaps were then generated for visualisation using ggplot2 in RStudio. Virtual capture plots were generated by extracting all entries from the tiled-C matrix in which specific viewpoints of interest participate, and interaction scores normalised by dividing interaction scores within each bin by the total number of interactions throughout the entire 3.3Mb tiled region. Virtual capture plots were produced for visualisation using ggplot2 in RStudio (Wickham, 2016). Loess (local regression) smoothing (span = 0.05) was used reduce noise in the virtual capture-C plots. This effectively runs multiple local regressions for each datapoint along the x-axis, with the span variable dictating the proportion of data taken into account when performing each regression. By re-plotting each datapoint based on the local regression prediction, therefore taking into consideration bins either side of the processed bin, I could reduce the noise in each plot. Because loess smoothing gives dramatically more weight to the values lying closest (along x) to the processed datapoint (weighting $\propto$

(1-(distance/maximum distance)^3)^3, it allows one to smooth the data with little information loss.

2.5.3 RNA-seq:

RNA-seq data was aligned to the mm9 genome, using star (Dobin et al., 2013). The resultant SAM files were then filtered and sorted using SAMtools (samtools view and sort, respectively). The resultant BAM files were indexed using SAMtools index, and directional, rpkm normalised bigwigs generated using deeptools bamcoverage, with the filteredRNAstrand flag enabled. Each sample bigwig was visualised using the University of California Santa Cruz (UCSC) genome browser, and traces corresponding to regions of interest downloaded from there. Read coverage over each gene in the mm9 genome was calculated using Rsubread featurecounts (Liao et al., 2019), and differential expression analysis performed using edgeR in RStudio (Y. Chen et al., n.d.). Plots were generated using ggplot2 in Rstudio. Principal component analysis was performed on RNA-seq samples, using the DiffBind, rgl and magick packages in RStudio. To compare enhancer RNA transcription in WT and R2-only cells, levels of poly-A negative RNA over the R1, R2, R3, Rm and R4 enhancers were visually assessed on the UCSC genome browser; however, this was only possible on the positive strand, as the Nprl3 gene, in which the R1, R2 and R3 enhancers are located, is transcribed on the negative strand. To compare R2 enhancer RNA transcription quantitatively, a virtual qPCR was performed, by normalizing the number of reads mapping to the R2 enhancer in each sample to the number of enhancer RNA reads mapping to the HS2 enhancer within the β-globin locus control region in the same sample. Levels of the normalised enhancer RNA transcription in WT and R2-only samples were then compared, and the results plotted, using ggplot2 in Rstudio.

Chapter three: Characterizing the R2-only mouse model: does R2 remain an active enhancer?

3.1 Introduction

Superenhancers, as defined by the ROSE algorithm, are dense clusters of enhancer elements enriched for particularly high levels of H3K27-acetylation and recruitment of coactivators, such as the mediator complex (Whyte et al., 2013) (**figure 1.3A**). Superenhancers often regulate cell identity genes, and their dysregulation is frequently associated with disease and complex traits (Dębek & Juszczyński, 2022; Harteveld & Higgs, 2010; Higgs et al., 2012; Tang et al., 2020; Yamagata et al., 2020). Over the past decade a significant amount of research has sought to understand how superenhancers up-regulate their target genes (Grosveld et al., 2021). The most pressing outstanding question is whether a superenhancer's constituents combine as independent enhancer elements, or whether they cooperate with one another such that their combination is greater than their sum (Blobel et al., 2021; Grosveld et al., 2021; Moorthy et al., 2017; Pott & Lieb, 2015; Wang et al., 2019) (**figure 1.3B**).

A number of studies have dissected superenhancers to explore the relationships between their constituents (Bender et al., 2012; Hay et al., 2016; Hnisz et al., 2015; Hörnblad et al., 2021; Huang et al., 2018; Shin et al., 2016; Thomas et al., 2021); however, all previous investigations have either restricted their analyses to individual and/or pairwise enhancer deletions, or used artificial reporter-based assays divorced from their biologically relevant chromatin contexts.

Genetic dissection of superenhancers is challenging, as many regulate genes which control complex transcriptional and epigenetic programmes. Disruption of such pathways makes it difficult to separate changes in superenhancer-regulated gene expression from associated changes in cell lineage and differentiation. Therefore, to investigate superenhancer cooperation in an interpretable manner, we required a model superenhancer which could be engineered *in situ* without causing changes to cell identity or wider transcriptional programmes. The mouse α-globin superenhancer, which up-regulates the two α-globin genes during erythropoiesis, is an ideal

candidate. The α-globin superenhancer is exclusively activated during terminal erythroid differentiation, and as such, its modification has no effect on development or cell identity (Harteveld & Higgs, 2010). Additionally, the α-globin superenhancer and genes are genomically insulated within a well-defined sub-TAD (Hanssen et al., 2017b), and α-globin has no downstream regulatory activity; any phenotypic changes can therefore be directly attributed to altered α-globin regulation (Oudelaar et al., 2021).

The mouse α-globin superenhancer is made up of five constituent elements: R1, R2, R3, R4 and Rm (a mouse-specific element). R1, R2, R3 and R4 show conservation in sequence and synteny over ~70 million years of evolution (Hughes et al., 2005), highlighting the α-globin superenhancer's importance within the locus and in activating α-globin gene expression. Our lab previously dissected the α-globin superenhancer *in vivo* by deleting each constituent *in situ,* individually and in informative pairs (Hay et al., 2016). Deleting R1 or R2 led to a 40% and 50% reduction in α-globin expression, respectively, marking these two elements as potent activators of gene expression. Meanwhile, deletion of R3 or Rm had no discernible effect on α-globin expression, and deleting R4 caused a small (~10%) but statistically significant reduction in gene expression (**figure 1.4C**). Summing the effects of all five deletions (ΔR1 = 40%; ΔR2 = 50%; ΔR3 = 0%; ΔRm = 0%; ΔR4 = 10%) results in 100%, suggesting that the five constituents combine additively, as independent elements, to drive α-globin expression (**figure 1.4C**). The finding that R1 and R2 are strong enhancers, while R3, Rm and R4 are weak/non-functional, was confirmed using lacZ reporter assays in mice (Hay et al., 2016).

Although these experiments shed light on the necessity of each individual element within the otherwise intact superenhancer, they did not report on each element's independent sufficiency to drive α-globin gene expression or the precise functional relationships between the five constituent elements. We postulated that a more rigorous test of each element's independence would be to assay their ability to up-regulate α-globin gene expression, in their native chromatin and developmental context, after all other elements have been removed from the locus. If a constituent

functions autonomously, its capacity to activate α-globin expression should be equal to the reduction in expression caused by its deletion.

Helena Francis, a previous member of the Higgs lab, generated a mouse model in which the R1, R3, Rm and R4 elements were removed from the native, otherwise intact, α-globin locus (R2-only) (Helena Francis, thesis; Blayney et al., 2022) (**figure 3.1A**). The fact that previous ΔR2 mice exhibited a ~50% reduction in gene expression and were otherwise healthy, led us to predict that R2-only mice would express ~50% α-globin and be free of pathology (**figure 3.1B**).

In this chapter, I rigorously characterise the R2-only mouse model. I begin by investigating its developmental phenotype and its ability to activate α-globin gene expression. I then proceed to characterise its histone modification profile, ability to recruit tissue-specific TFs, and its ability to recruit coactivators and components of the preinitiation complex. Finally, I investigate chromatin interactions throughout the α-globin locus, and R2's ability to transcribe bi-directional enhancer RNA. Cumulatively, I thoroughly explore R2's enhancer status and ability to activate its target gene, when separated from all other constituents of the α-globin superenhancer.

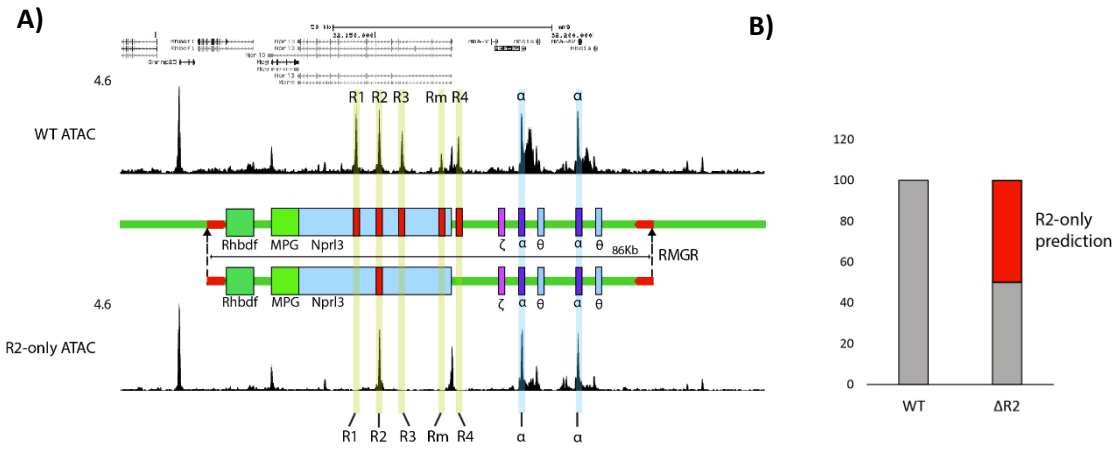


Figure 3.1 The R2-only mouse model.
(A) Design of the R2-only α-globin locus. R2-only locus synthesized, assembled into a bacterial artificial chromosome, and delivered into the WT α-globin locus through recombination-mediated genomic replacement. ATAC-seq trace from WT foetal liver cells (top) and homozygous R2-only foetal liver cells (bottom). (B) α-globin gene expression in WT and ΔR2 mouse models (from Hay et al., 2016), red = prediction of R2-only α-globin gene expression, calculated by subtracting ΔR2 α-globin expression from WT.

3.2 The R2-only model: Phenotype and viability

Helena Francis, a previous DPhil student in the Higgs lab, generated an R2-only mouse model, in which the native 86kb α-globin locus (containing the entire α-globin superenhancer, all α-like globin genes, and three additional genes lying upstream of the α-globin superenhancer: Nprl3, MPG and Rhbdf) is exchanged via recombinase-mediated genomic replacement (RMGR) (Wallace et al., 2007) with a synthetic locus (generated by Leslie Mitchell, using a recently developed protocol for de novo assembly of large DNA fragments (Mitchell et al., 2021)) in which the α-globin superenhancer is cropped to its strongest enhancer element, R2 (**figure 3.1A**). The R2-only locus lacks the R1, R3, Rm and R4 elements, but is otherwise WT in composition. The genome-integrated R2-only locus was screened by 10X linked-read sequencing. This confirmed that the overall locus integrity was intact, but it also identified four small mutations (**table 3.1**) within the BAC-derived locus. None of these mutations are predicted to have any effect on TF-recruitment, DNA accessibility, or gene expression, when screened using the *in silico* jaspar and sasquatch tools (Fornes et al., 2020; Schwessinger et al., 2017).

Table 3.1 Mutations present in genome-integrated R2-only locus.

Position in BAC	Type of mutation	Reference nucleotide	Mutated nucleotide	Region affected
6,145	Substitution	C	T	Rhbdf intron
44,425	Deletion	30nt	-	Nprl3 intron
46,911	Substitution	G	A	Nprl3 intron
58,887	Substitution	G	C	Intergenic

Our previous work showed that ΔR2 mice express ~50% α-globin compared to WT and suggested that the five α-globin superenhancer constituents are independent elements; we therefore predicted that R2-only mice would also express approximately 50% α-globin (**figure 3.1B**). Moreover, we predicted that R2-only mice, like ΔR2, would be viable and healthy.

We established heterozygote (R2-only/WT X R2-only/WT) crosses, but to our surprise, only obtained a single R2-only homozygous mouse from 15 such crosses. The single homozygous survivor presented with splenomegaly (**figure 3.2A**) and anaemia, and died prematurely at 7 weeks. The homozygote died in the afternoon/evening, and was not discovered until the following morning. Due to the elapsed time between the R2-

only homozygote's death and its discovery, it was impossible to investigate its phenotype further.

The postnatal lethality of R2-only homozygotes restricted investigation of the model's molecular phenotype to embryonic timepoints. We established timed heterozygote crosses, and sacrificed pregnant mice at various stages of gestation: embryonic days E9.5, E10.5, E12.5, E14.5 and E17.5. I dissected embryos from pregnant mothers and extracted erythroid cells from the appropriate tissues. In E9.5 and E10.5 embryos this entailed mechanical disaggregation in heparinised PBS. The erythroid cell-containing supernatant was then aspirated into tubes for processing, and remaining tissues were stored for genotyping. E12.5-E17.5 embryos were aligned according to their predicted genotypes and photographed. I then isolated the foetal liver (the definitive erythroid compartment throughout these timepoints) from each embryo, mechanically disaggregated them in FACS buffer, and collected the erythroid cells for processing. I stored brain tissue for genotyping, and also stored brain tissue from E17.5 embryos for gene expression analysis.

The genotypes of embryos comprising each litter (at every embryonic stage) approximated the expected Mendelian ratio (**table 3.2**); however, R2-only homozygotes were visibly smaller and paler than their WT and heterozygote littermates (**figure 3.2B**; **figure 3.2C**). R2-only postnatal lethality, and the visibly anaemic and developmentally stunted phenotype of homozygous embryos demonstrated that the phenotype of R2-only mice was much more severe than that of ΔR2.

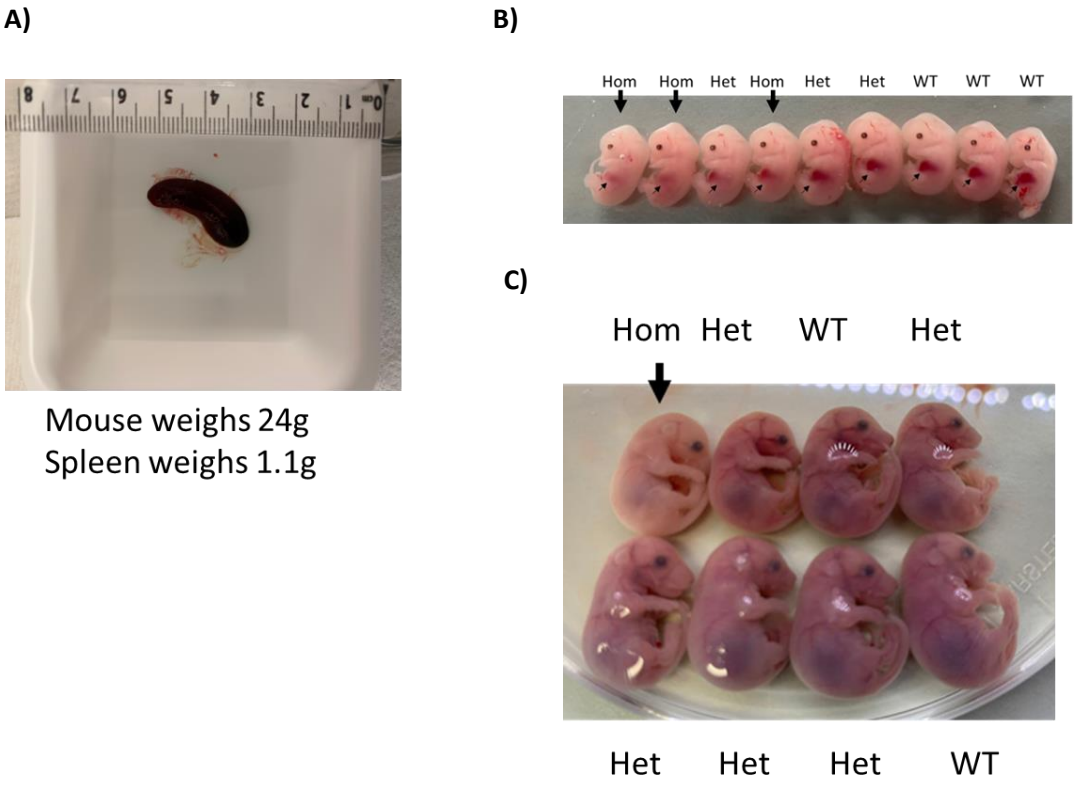


Figure 3.2 R2-only homozygous mice are non-viable.

(A) Spleen dissected from the only R2-only survivor (which died prematurely at 7 weeks compared average of 9-12 months). Spleen = 1.1g (4.6% of total body weight, compared to average of 0.35%), indicating severe splenomegaly. (B) Representative image of R2-only Het X Het litter. Pregnant female sacrificed at embryonic day E12.5, and foetuses extracted. Genotypes indicated above and foetal livers indicated by black arrows. (C) Representative image of R2-only Het X Het litter. Pregnant female sacrificed at embryonic day E17.5, and foetuses extracted. Genotypes indicated above/below.

Table 3.2 Genotypes of embryos obtained from Het X Het crosses.

Pregnant females sacrificed at various timepoints post-gestation, embryos extracted and genotyped. All timepoints contain genotypes at roughly the expected Mendelian ratio.

Stage	Hom	Het	WT
E9.5	1	5	2
E10.5	2	12	6
E12.5	5	7	5
E14.5	4	6	3
E17.5	3	7	6
cumulative	**15**	**37**	**22**
%	**20**	**50**	**30**

3.3 R2-only gene expression and locus transcription status (R2 transcription)

To investigate the severe R2-only phenotype further, I isolated erythroid cells from R2-only homozygous, heterozygous and WT embryos at each timepoint (E9.5-E17.5),

extracted RNA, and performed RT-qPCR. I normalized all α-globin expression data to β-globin expression, and normalized expression of all other genes to RPS18. All expression data is displayed relative to the mean average of WT littermates. This revealed that, rather than expressing 50% α-globin as would be expected in an additive model of superenhancer cooperation, R2-only homozygotes only expressed 10-15% (**figure 3.3A**). This observation was consistent across all timepoints studied (**figure 3.3B**).

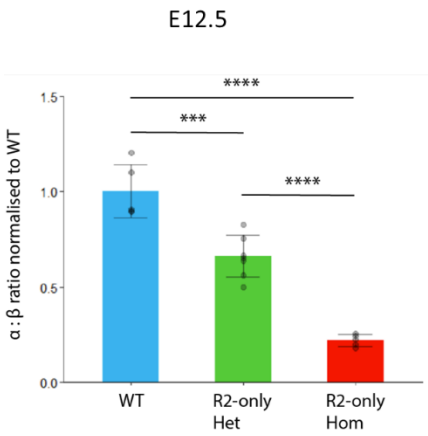
To explore the transcriptome of R2-only mice further, I performed poly-A minus and poly-A positive RNA-seq on foetal liver cells from three E12.5 R2-only homozygotes and two WT littermates. These data agreed well with the RT-qPCR experiments, demonstrating an ~85% reduction in α-globin expression in the R2-only erythroid cells (**figure 3.3C**; **figure 3.3D**). The poly-A minus RNA-seq data also revealed significant reductions in the expression of two genes lying upstream of the α-globin superenhancer: Snrnp25 and MPG (**figure 3.3C**). The reduction in Snrnp25, but not MPG, was detected as significant in the poly-A positive dataset (**figure 3.3D)**. The poly-A minus data is a more accurate representation of RNA abundance as a result of transcription, whereas the poly-A positive data is a representation of transcription and RNA half-life (primarily driven by degradation rate). Therefore, the fact MPG was only found to be significantly reduced in the poly-A minus dataset likely indicates that the rate of MPG transcription is reduced, but that the transcript remains fairly stable in the cell. The changes in Snrnp25 and MPG expression were unexpected, due to the fact that both of these genes lie upstream of the CTCF sites demarcating the 5' boundary of the α-globin sub-TAD (**figure 3.3E**) (Hanssen et al., 2017); therefore, we previously thought that Snrnp25, MPG and other upstream genes were insulated from the effects of the α-globin superenhancer. Despite the fact four out of the five α-globin superenhancer constituents fall within introns of the Nprl3 gene, there was no change in Nprl3 expression (**figure 3.3C**; **figure 3.3D**).

I performed RT-qPCR, examining expression of Nprl3, Snrnp25 and MPG in erythroid cells; this confirmed the RNA-seq data, showing a reduction in Snrnp25 and MPG expression, but no change in Nprl3 transcription (**figure 3.3F**). To confirm that the changes in Snrnp25 and MPG expression are due to reduced superenhancer activity,

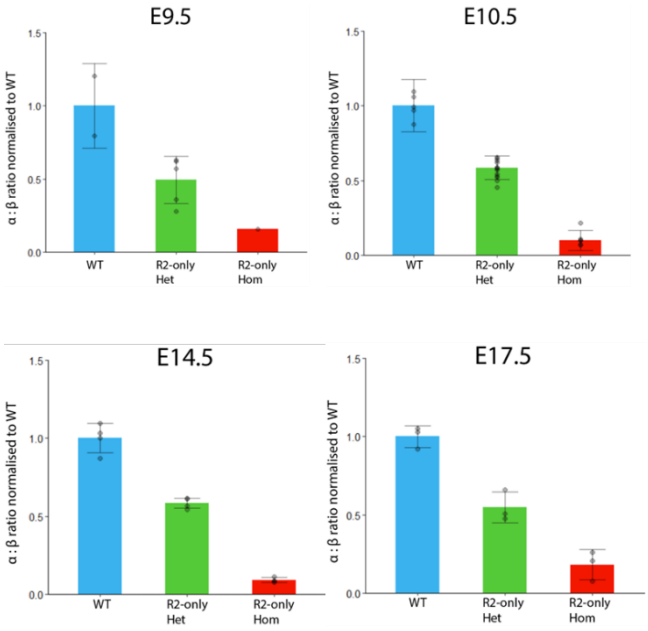
and not a result of the genome engineering process itself, I analysed Nprl3, Snrnp25 and MPG expression in brain tissue (in which all three genes are expressed, but the α-globin superenhancer is inactive) by RT-qPCR. All three genes were expressed to similar levels in R2-only and WT littermate brain samples (**figure 3.3F**), indicating that the expression changes detected in erythroid cells are specific to the erythroid compartment, in which the α-globin superenhancer is active.

The RNA-seq experiments showed no changes in the expression of a number of erythroid developmental markers, such as the β-like globin genes and transferrin receptor (Tfrc) (**figure 3.3C**; **figure 3.3D**). This suggested that, although α-globin expression was severely attenuated in the R2-only model, R2-only erythroid cells still move successfully through erythropoiesis.

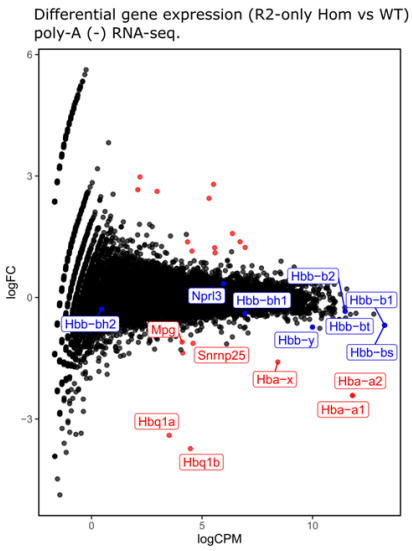
A)



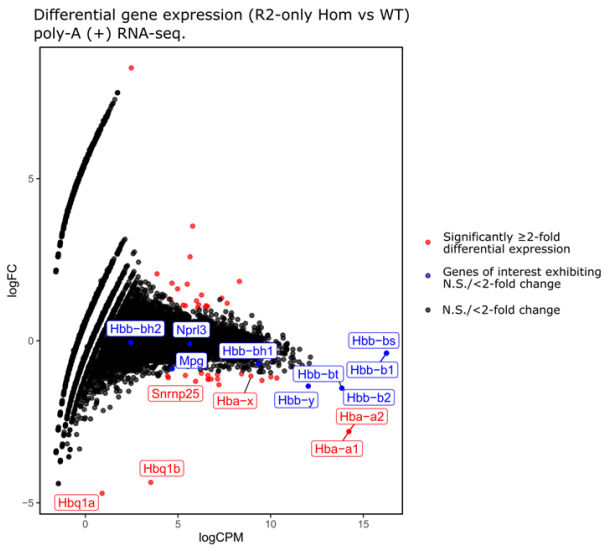
B)



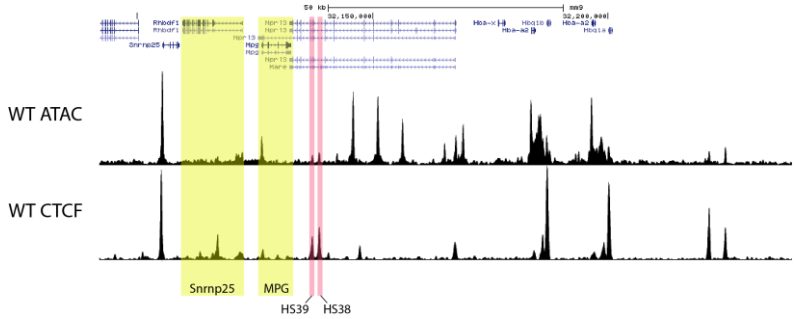
C)


Differential gene expression (R2-only Hom vs WT)
poly-A (-) RNA-seq.
logFC
logCPM
Hbb-b2
Nprl3
Hbb-bh1
Hbb-b1
Hbb-bh2
Mpg
Hbb-bt
Hbb-y
Snrnp25
Hbb-bs
Hba-x
Hba-a2
Hba-a1
Hbq1a
Hbq1b

D)


Differential gene expression (R2-only Hom vs WT)
poly-A (+) RNA-seq.
logFC
logCPM
Hbb-bh2
Nprl3
Hbb-bs
Mpg
Hbb-bh1
Hbb-bt
Hbb-b1
Snrnp25
Hba-x
Hbb-y
Hbb-b2
Hba-a2
Hba-a1
Hbq1b
Hbq1a
Significantly ≥2-fold differential expression
Genes of interest exhibiting N.S./<2-fold change
N.S./<2-fold change

E)


WT ATAC
WT CTCF
Snrnp25
MPG
HS39
HS38

F)

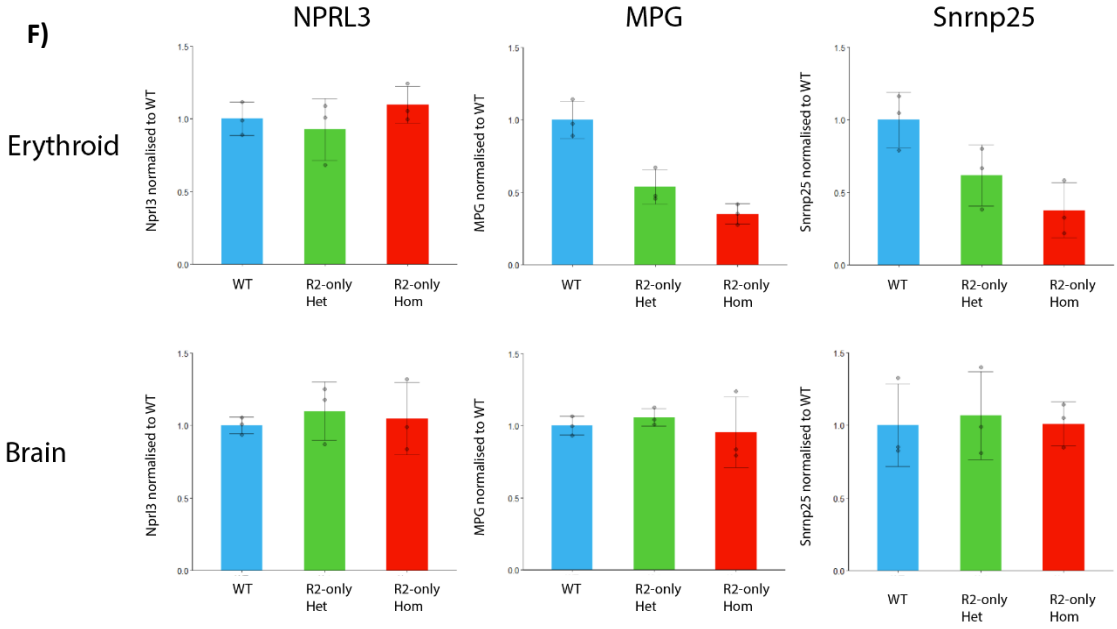

NPRL3
MPG
Snrnp25
Erythroid
Brain
Nprl3 normalised to WT
MPG normalised to WT
Snrnp25 normalised to WT
WT
R2-only Het
R2-only Hom

Figure 3.3 R2-only embryos express far lower α-globin than predicted.

(A) RT-qPCR comparing α-globin expression in foetal liver erythroid cells from WT, R2-only heterozygous and R2-only homozygous E12.5 littermates. Expression normalized to β-globin and displayed as a proportion of mean WT expression level. Dots = biological replicates; error bars = SE. (B) RT-qPCR comparing α-globin expression in foetal liver erythroid cells from WT, R2-only heterozygous and R2-only homozygous E9.5. E10.5, E14.5 and E17.5 littermates. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. (C) Poly-A minus RNA-sequencing comparing gene expression in foetal liver erythroid cells from WT (n=2) and R2-only homozygous (n=3) littermates. RNA was rRNA depleted, and poly-A minus transcripts isolated by removing poly-A positive fraction prior to library preparation. Red = statistically differentially expressed, blue = informative erythroid genes (non-differentially expressed). (D) Poly-A positive RNA-sequencing comparing gene expression in foetal liver erythroid cells from WT (n=2) and R2-only homozygous (n=3) littermates. RNA was rRNA depleted, and poly-A positive transcripts isolated using the NEBNext® Poly(A) mRNA Magnetic Isolation Module prior to library preparation. Red = statistically differentially expressed, blue = informative erythroid genes (non-differentially expressed). (E) ATAC-seq track from WT foetal liver cells (top); CTCF ChIP-seq track (bottom). HS38 and HS39 CTCF binding sites, which mark the 5' boundary of the α-globin sub-TAD, indicated. Snrnp25 and MPG genes, which both lie upstream of the 5' boundary of the α-globin sub-TAD, indicated in yellow. (F) RT-qPCR comparing Nprl3, MPG and Snrnp25 expression in WT, R2-only heterozygous and R2-only homozygous littermates. Expression normalized to RPS18 and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. Upper = E17.5 erythroid cells; lower = matched E17.5 Brain tissue.

3.4 R2 accessibility and epigenetic profile

Gene expression analysis demonstrated that R2 alone is insufficient to activate high levels of α-globin transcription. This led us to ask whether the R2 element maintains its active enhancer status in the R2-only context.

Chromatin accessibility, enhancer-associated histone modifications (including H3K4Me1 and H3K27Ac), and tissue-specific TF recruitment are three typical markers of active enhancers (Long et al., 2016). In order to assay R2's enhancer status, I sought to profile each of these characteristics in the R2-only model.

I performed ATAC-seq in three WT foetal liver samples and three R2-only homozygotes. This revealed that the R2 enhancer becomes accessible to a similar degree in WT and homozygous R2-only erythroid cells (**figure 3.4A**). There was a reduction in DNA accessibility at the two α-globin genes in R2-only homozygotes, consistent with lower levels of α-globin transcription. DNA accessibility throughout the rest of the locus, including CTCF sites and the upstream genes was unaltered in R2-only cells (**figure 3.4A**), demonstrating that the overall integrity of the α-globin locus was unaffected.

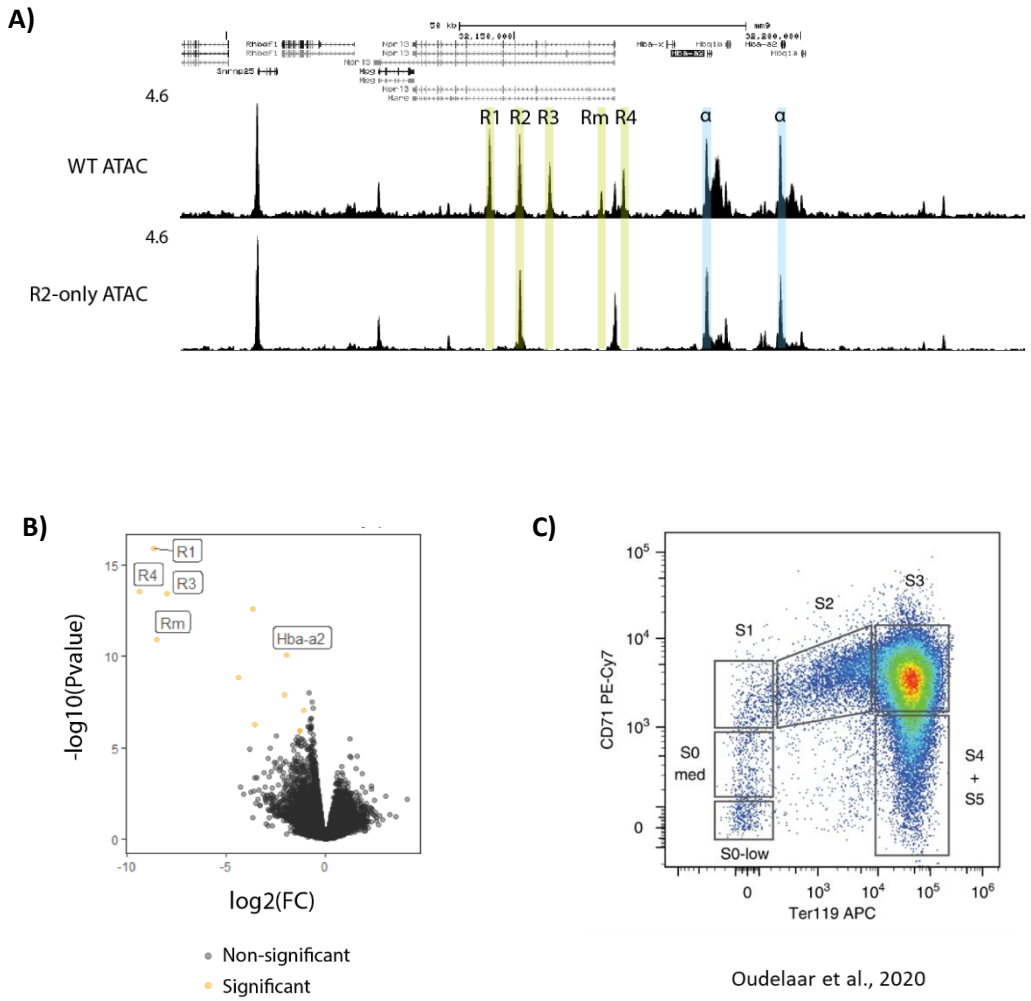
To verify that global chromatin accessibility was unperturbed in the R2-only model, I used macs2, a peak-calling tool, to identify genome-wide peaks of DNA accessibility in the WT and homozygous R2-only ATAC-seq datasets. I then performed differential accessibility analysis, which confirmed that R1, R3, Rm and R4 were by far the most differentially accessible regions, genome-wide (**figure 3.4B**). The differential

accessibility analysis also revealed significantly lower accessibility at the α-globin promoters in the R2-only model, as well as at seven other sites. Visual inspection of these seven loci showed that they were either regions of very low read coverage, or regions on the boundaries of blacklist coordinates, suggesting that these changes are most likely artefacts.

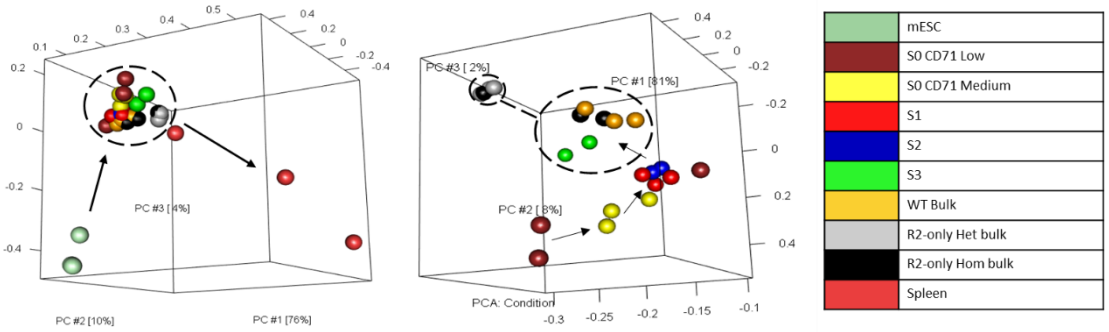
Dr Robert Beagrie, a previous post-doctoral researcher in the Higgs lab, used ATAC-seq to assay DNA accessibility in WT foetal liver cells as they progress through erythroid differentiation (Oudelaar et al., 2020). He used erythroid cell surface markers: CD71 and Ter119, to FACS sort cells into five categories: S0 low, S0 medium, S1, S2 and S3 (**figure 3.4C**) (Koulnis et al., 2011; Oudelaar et al., 2020), according to their developmental stages from haematopoietic stem cells (S0 low) to terminally differentiating erythroid cells (S3). S3 cells (CD71+/Ter119+) make up the majority (~95%) of the WT foetal liver population. To test whether R2-only homozygous erythroid cells progress normally through erythroid differentiation, I re-analysed Dr Beagrie's ATAC-seq data from developmentally-staged erythroid populations alongside my own R2-only homozygous, R2-only heterozygous, and WT bulk ATAC-seq data, using macs2 to identify DNA accessibility peaks throughout the genome. I then used principal component analysis to cluster samples based on their genome-wide accessibility profiles (**figure 3.4D**). To test whether the clusters formed by the principal component analysis could be used as a faithful proxy for developmental state, I conducted factor loading analysis to extract the variables with the highest loadings in each of the three most dominant eigenvectors (essentially identifying peaks contributing the most to principle components one, two and three). Peaks contributing to principal components one, two and three explained 81%, 8% and 2% of the overall variance between samples, respectively. I extracted the peaks with the highest loadings in each of these components and performed gene ontology (GO) analysis. This revealed that the most enriched GO terms were related to developmental processes (**table 3.3**). This strongly indicated that the majority of the variation between samples' (S0-S3) genome-wide accessibility profiles was related to their developmental stages, and that principal component analysis could therefore be used to cluster samples based on this. R2-only erythroid cells clustered very closely

with WT cells, and in particular WT S3 cells (which make up 95% of the WT FOETAL LIVERpopulation) (**figure 3.4D**), suggesting that R2-only cells can progress normally through erythroid differentiation.

FACS analysis, staining for CD71 and Ter119, supported the principal component analysis. ~95% of homozygous R2-only (n=3) and WT (n=2) foetal liver populations were CD71+/Ter119+ (**figure 3.4E**).



D)



E)

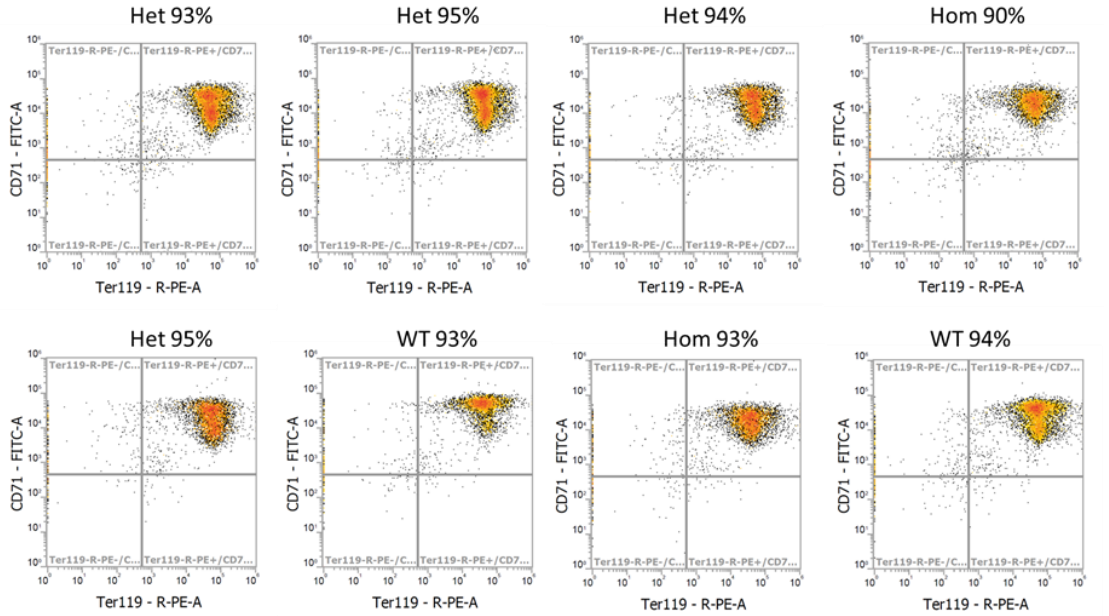


Figure 3.4 Genome-wide chromatin accessibility is unaffected in R2-only embryos.

(A) ATAC-seq in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells. Tracks = merged biological replicates. (Coordinates = chr11: 32,090,000-32,235,000). (B) Genome-wide differential accessibility assessed through ATAC-seq in foetal liver erythroid cells from WT (n=3) and R2-only homozygous (n=3) littermates. Yellow = significantly differentially accessible; black = non-significant. (C) Figure from Oudelaar et al., 2020 (work performed by Dr Robert Beagrie in the Higgs lab). E12.5 Foetal liver cells isolated and a proportion of Ter119+ cells removed to allow visualisation of early erythroid progenitors (>95% of total population is Ter119+). Remaining cells FACS sorted based on staining for CD71 (y-axis) and Ter119 (x-axis) erythroid-associated cell surface markers. Cells gated based on developmental stage from S0 low (haematopoietic stem cells) to S4 & S5 (terminal erythroid cells. S3 cells (CD71+/Ter119+). (D) Left: principal component analysis comparing genome-wide ATAC-seq peaks in WT, R2-only homozygous and R2-only heterozygous foetal liver erythroid cells, alongside mESC, developmentally-staged foetal liver erythroid cell, and adult spleen erythroid cell ATAC-seq peaks. Right: same analysis with mESC and spleen cells excluded to allow clearer separation of developmentally staged foetal liver cells. (E) FACS plots staining for CD71 (y-axis) and Ter119 (x-axis) erythroid-associated cell surface markers, from E12.5 foetal liver erythroid populations. WT and R2-only homozygote littermates contain ~95% CD71+/Ter119+ erythroid cells.

Table 3.3 Factor loading performed on principal component analysis.

ATAC-seq peaks contributing most to principal components 1, 2 and 3 isolated. Peaks then annotated using Homer and enriched gene ontology (GO) terms extracted. The majority of GO terms relate to developmental processes, and in particular, erythroid differentiation.

PC1 (81% variance explained)	PC2 (8% variance explained)	PC3 (2% variance explained)
multicellular organismal development	single-organism developmental process	negative regulation of cellular process
definitive erythrocyte differentiation	red blood cell differentiation	neurogenesis
developmental process	system development	multicellular organismal development
system development	multicellular organismal development	positive regulation of biological process
anatomical structure development	hematopoietic stem cell homeostasis	developmental process
erythrocyte cell differentiation	cell differentiation	single-organism developmental process
organ development	single-organism developmental process	positive regulation of cellular process
cell differentiation	developmental process	system development
regulation of cell communication	anatomical structure development	negative regulation of biological process
cellular developmental process	regulation of biological process	anatomical structure development

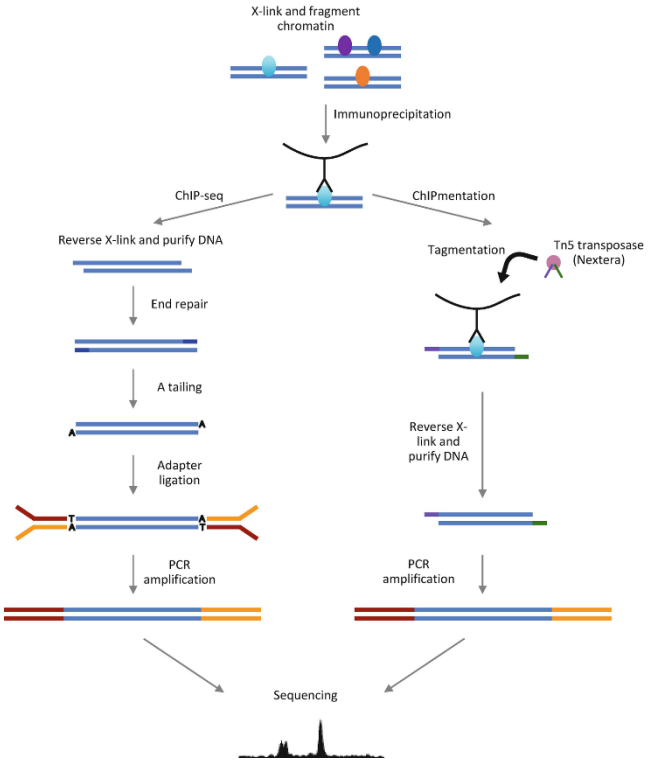
E12.5 foetal liver harvests generally yield ~10 million erythroid cells, and traditional ChIP-seq methods use 5-10 million cells per replicate. To increase the number of experiments I could perform on individual foetal liver samples (thereby minimising the number of mice we would need to sacrifice, in accordance with the "3 Rs" (Hubrecht & Carter, 2019)), I sought a lower input method for assaying genome-wide histone modifications and TF recruitment. I introduced the ChIPmentation method into the lab, optimizing the existing protocol (Schmidl et al., 2015), for use with mouse FOETAL LIVERerythroid cells. ChIPmentation is a chromatin immunoprecipitation-based technique, in which samples are fixed, lysed and sonicated, followed by immunoprecipitation with an antibody against the factor of interest. Rather than proceeding with typical library preparation, the ChIPmentation method uses sequencing adaptor-loaded tn5 enzyme to "tagment" the immunoprecipitated chromatin, before indexing PCR (**figure 3.5A**). This one-step library preparation workflow minimises sample losses, and allowed me to assay histone modifications in as few as 50,000 erythroid cells, and TF recruitment in as few as 100,000 erythroid cells.

I performed ChIPmentation experiments in foetal liver samples from three R2-only homozygotes and three WT littermates, using antibodies against H3K4Me1, H3K27Ac, and H3K4Me3: marks associated with active enhancers, active enhancers and promoters, and active promoters, respectively. H3K4Me1 and H3K27Ac were still present over the R2 enhancer in R2-only erythroid cells, although levels of both modifications were visibly reduced when compared to WT (**figure 3.5B**).

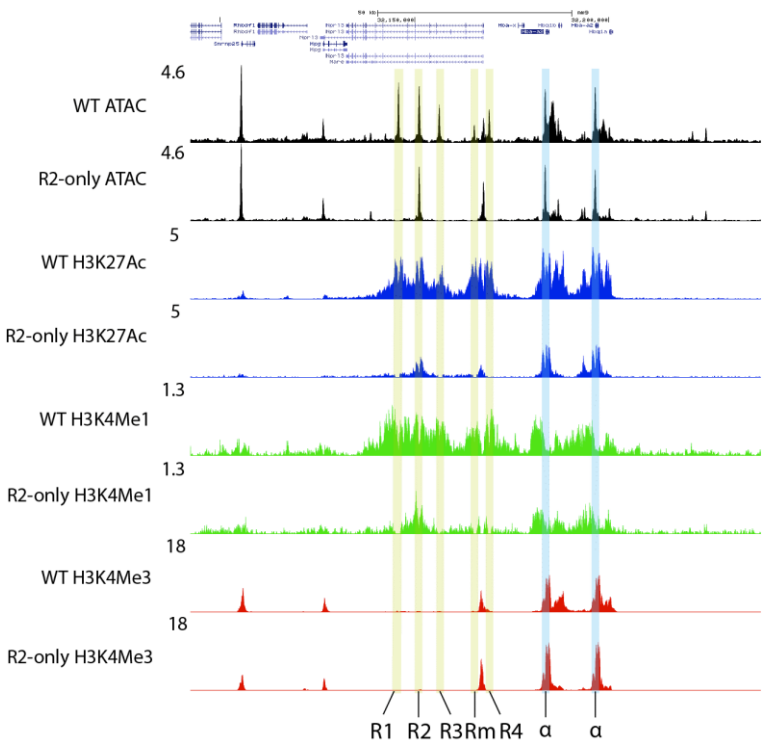
Levels of H3K27Ac were also reduced over the α-globin promoters in R2-only cells, as was the breadth of the H3K4Me3 signal (**figure 3.5B**). H3K4Me3 and H3K27Ac levels over the Snrnp25 promoter were also slightly reduced. These findings are consistent with the changes in α-globin and Snrnp25 gene expression, though histone modifications over MPG were unaffected. Histone modifications over other genes within and surrounding the locus were unperturbed, as were levels of H3K4Me1, H3K27Ac, and H3K4Me3 at the β-globin locus, suggesting that all changes at the α-globin locus were specific.

In conclusion, R2's accessibility and histone modification profile is consistent with it remaining an active enhancer in R2-only erythroid cells.

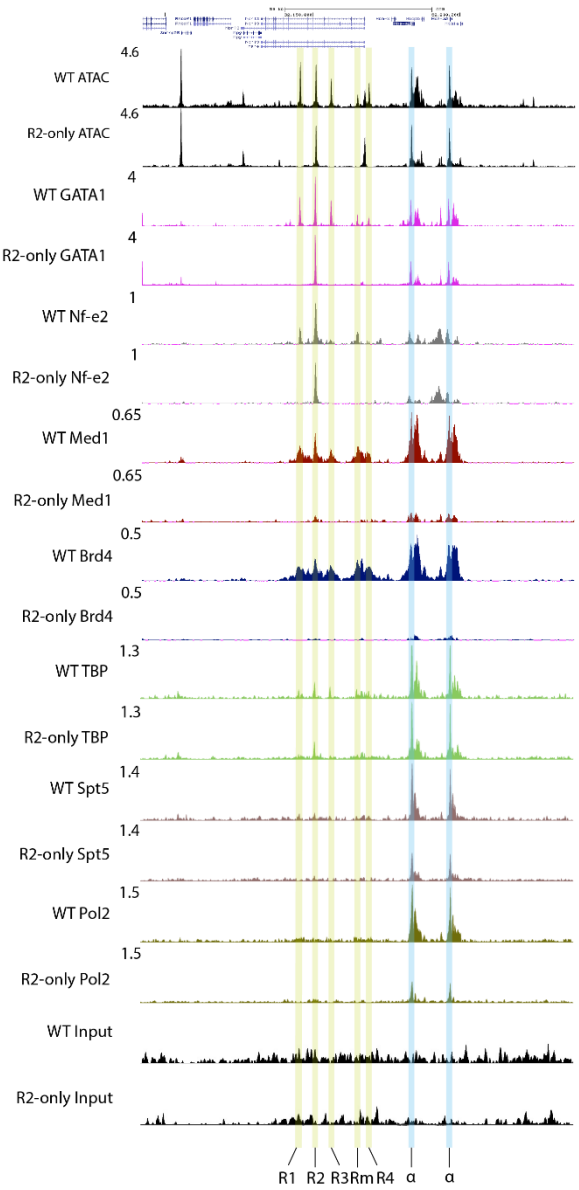
A)


X-link and fragment chromatin
Immunoprecipitation
ChIP-seq
ChIPmentation
Reverse X-link and purify DNA
Tagmentation
Tn5 transposase (Nextera)
End repair
A tailing
Adapter ligation
Reverse X-link and purify DNA
PCR amplification
PCR amplification
Sequencing

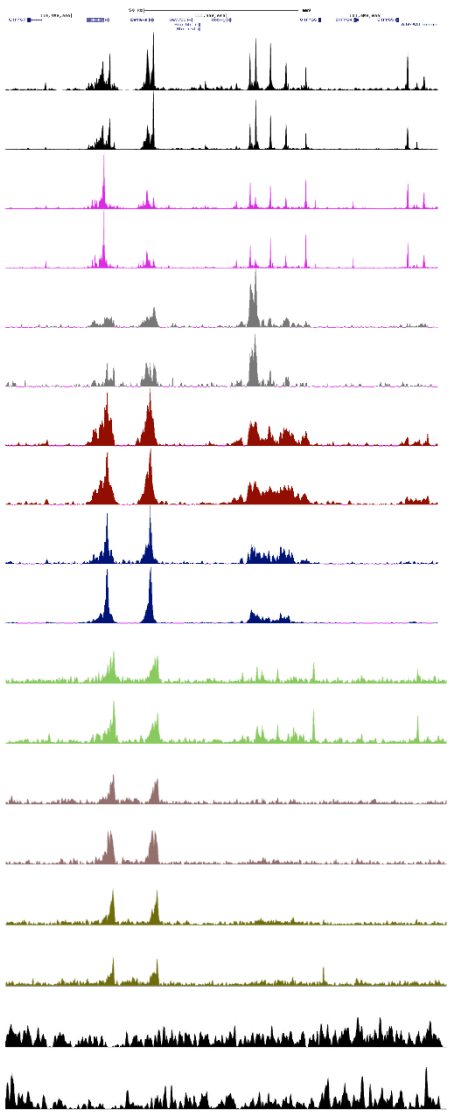
B)


4.6
WT ATAC
4.6
R2-only ATAC
5
WT H3K27Ac
5
R2-only H3K27Ac
1.3
WT H3K4Me1
1.3
R2-only H3K4Me1
18
WT H3K4Me3
18
R2-only H3K4Me3
R1 R2 R3 Rm R4 α α

C)


WT ATAC
R2-only ATAC
WT GATA1
R2-only GATA1
WT Nf-e2
R2-only Nf-e2
WT Med1
R2-only Med1
WT Brd4
R2-only Brd4
WT TBP
R2-only TBP
WT Spt5
R2-only Spt5
WT Pol2
R2-only Pol2
WT Input
R2-only Input
R1 R2 R3 Rm R4 α α

D)



E)

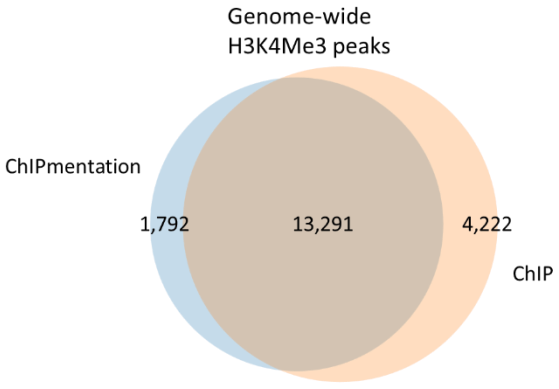

Genome-wide
H3K4Me3 peaks
ChIPmentation
1,792
13,291
4,222
ChIP

Figure 3.5 Histone modifications, TF recruitment and coactivator recruitment in WT and R2-only embryos.

(A) Schematic of ChIP-seq library preparation protocol (left) and ChIPmentation library preparation protocol (right). (B) ATAC-seq in WT (n=3) and R2-only homozygous foetal liver erythroid cells (black); H3K27Ac ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells (blue); H3K4Me1 ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells (green); H3K4Me3 ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells (red). Tracks = merged biological replicates. (Coordinates = chr11: 32,090,000-32,235,000). All tracks cpm normalised. (C) Top-bottom: ATAC-seq in WT (n=3) and R2-only homozygous foetal liver erythroid cells (black); Gata1 ChIPmentation in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells; Nf-e2 ChIPmentation in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells; Med1 ChIPmentation in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells; Brd4 ChIPmentation in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells; TBP ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells; Spt5 ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells; Pol2 ChIPmentation in WT (n=3) and R2-only homozygous (n=3) foetal liver erythroid cells; Input ChIPmentation in WT (n=2) and R2-only homozygous (n=2) foetal liver erythroid cells. Tracks = merged biological replicates. (Coordinates = chr11: 32,090,000-32,235,000). All tracks cpm normalised. (D) Same as C, but coordinates = chr7:110,936,000-111,070,000. (E) Genome-wide H3K4Me3 peaks from three ChIPmentation and three ChIP-seq experiments with the same antibody against H3K4Me3. The majority of ChIPmentation peaks overlap with those identified in ChIP-seq.

3.5 R2 TF recruitment

Active enhancers recruit high levels of tissue-specific TFs (Long et al., 2016; Spitz & Furlong, 2012; Zinzen et al., 2009). Previous studies have demonstrated that R2 is occupied by high levels of erythroid master regulators, including Gata1, Nf-e2, Klf1 and SCL/Tal1 (Hay et al., 2016; Oudelaar et al., 2021). To test whether R2 can still recruit TFs in R2-only erythroid cells, I performed ChIPmentation experiments with antibodies against Gata1 and Nf-e2, in erythroid cells from three WT and three R2-only homozygote foetal livers. This revealed that Gata1 and Nf-e2 recruitment were unperturbed at the R2 element in R2-only cells (**figure 3.5C**). Similarly, there was no detectable change in Gata1 or Nf-e2 recruitment at the α-globin promoters, or throughout the rest of the locus. Neither were there any changes throughout the β-globin locus (**figure 3.5D**).

As a superenhancer constituent, R2 is also occupied by high levels of transcriptional coactivators, such as the mediator complex and Brd4 (Hay et al., 2016). To test recruitment of these factors in R2-only erythroid cells, I performed ChIPmentation with antibodies against Med1 (a member of the mediator complex) and Brd4 (a transcriptional and epigenetic regulator) in FOETAL LIVERcells from three WT and three R2-only homozygotes. WT erythroid cells showed high levels of Med1 and Brd4 recruitment to all five constituents of the α-globin superenhancer, and especially high

levels of recruitment at the two α-globin promoters (**figure 3.5C**). There was a very striking reduction in recruitment of both coactivators throughout the α-globin locus in R2-only erythroid cells (**figure 3.5C**). Not only did levels of Med1 and Brd4 fall dramatically at the R2 enhancer, but there was an even greater decrease in occupancy at the α-globin promoters. Med1 and Brd4 recruitment were unperturbed at the β-globin locus in R2-only erythroid cells (**figure 3.5D**).

Mediator plays a crucial role in recruiting, stabilizing and activating Pol2 at promoters (Soutourina, 2017); I therefore asked what effect reduced Med1 and Brd4 recruitment has on formation of the preinitiation complex. I performed ChIPmentation on WT and R2-only homozygous FOETAL LIVERcells, with antibodies against TATA-binding protein (TBP – an essential member of the TFIID complex) and Pol2. Recruitment of TBP to the α-globin promoters was unaffected in R2-only erythroid cells; however, there was a stark reduction in Pol2 occupancy (**figure 3.5C**). There was also a reduction in recruitment of the Spt5 elongation factor, which is known to have a role in Pol2 pause-release (Fitz et al., 2018) (**figure 3.5C**). This suggests that both transcription initiation and pause-release are likely affected in the R2-only model, as previously reported at other loci (Gorbovytska et al., 2022; Larke et al., 2021).

ChIPmentation uses tn5 transposase to "tagment" immunoprecipitated chromatin; we therefore wondered whether it exhibits any bias for accessible chromatin, which could lead to over-representation of accessible regions in resultant bigwig tracks. I tagmented sonicated material prior to immunoprecipitation to generate "input" bigwig tracks. Input tracks for WT and R2-only homozygous FLs were equivalent, and there was only a small increase in signal over accessible regions (**figure 3.5C**; **figure 3.5D**). Furthermore, the fact that each ChIPmentation track is clearly distinguishable from ATAC, inputs, and ChIPmentation performed with different antibodies, demonstrates that use of tn5 has no major influence on ChIPmentation signal. To further validate my ChIPmentation data, I used macs2 to call genome-wide peaks from three H3K4Me3 ChIPmentation tracks (from 100K WT FOETAL LIVERcells), and used bedtools intersect to specifically extract the peaks which were present in all three files, thereby generating an H3K4Me3 ChIPmentation consensus file. I also performed peak-calling on three previously generated ChIP tracks using the same antibody (from 10

million WT FOETAL LIVERcells), and performed the same intersection to produce an H3K4Me3 ChIP consensus file. Next, I used bedtools intersect to separate ChIPmentation consensus peaks into: those which overlap with ChIP consensus peaks (overlapping), and those which do not overlap (non-overlapping). I found that H3K4Me3 ChIPmentation closely resembled the profile as determined by ChIP, with few novel, non-overlapping peaks; however, ChIP did identify ~28% more peaks than ChIPmentation (**figure 3.5E**). This suggests that ChIPmentation is comparable to, but potentially less sensitive than, ChIP.

3.6 R2 chromatin interactions

Previous chromosome conformation capture (3C) experiments have demonstrated that the α-globin superenhancer constituents, in particular the R1 and R2 enhancers, interact with the α-globin promoters with high frequency (Hanssen et al., 2017; Hay et al., 2016; Hua et al., 2021; Hughes et al., 2014; King et al., 2021; Oudelaar et al., 2018, 2019). Many studies have presented evidence that enhancer-promoter interactions are important for effective enhancer activity and target gene upregulation (Allahyar et al., 2018; Beagrie et al., 2017; Grosveld et al., 2021; Ing-Simmons et al., 2015; Y. Li et al., 2018; Oudelaar & Higgs, 2021). To investigate chromatin interactions throughout the α-globin locus in FOETAL LIVERcells, I sought a low-input 3C-based method, ideally requiring one million cells or fewer. This would allow me to assay chromatin accessibility, histone modifications, TF recruitment, coactivator recruitment, genome-wide transcription, and 3D chromatin interactions within the same samples. Yet, I also required a high-resolution method, as the entire α-globin locus is only 86kb in length, and the distance between R2 and the superenhancer-proximal α-globin promoter is only ~30kb. Tiled-C, a recently developed 3C-based technique, would allow me to assay all-vs-all pairwise chromatin interactions throughout the α-globin locus; moreover, it would allow me to do this in 500,000 cells, whilst maintaining a resolution of as little as 750bp (Oudelaar et al., 2020).

Tiled-C is a typical 3C proximity-ligation method, which begins with sample fixation and lysis, followed by DNA digestion with a restriction endonuclease (DpnII in my

experiments). Digested fragments are then re-ligated such that those fragments closest to one another in 3D space are ligated (this can be fragments adjacent to one another along the linear genome, or fragments brought into close apposition as a result of DNA looping (or other processes)) (**figure 3.6A**). Re-ligated fragments are then sonicated to a length at which the average fragment contains a single digestion-re-ligation junction. The library is then prepared for sequencing (adaptor ligation and indexing), followed by oligonucleotide capture with tiled DNA probes specific to every restriction fragment throughout a region of interest (3.3Mb surrounding the α-globin locus in my experiments). This allows analysis of all-vs-all pairwise interactions throughout the captured tiled locus (see methods for additional information).

I performed tiled-C in FOETAL LIVERcells from three WT and three R2-only homozygotes. Heatmaps displaying all-vs-all pairwise chromatin interactions demonstrated a lower intensity of interactions throughout the 65kb α-globin sub-TAD in R2-only cells (**figure 3.6B**), although this was unsurprising, given the loss of four out of the five α-globin superenhancer constituents. Next, I extracted all rows from the tiled-C matrices in which individual, informative viewpoints of interest participate. This allowed me to generate "virtual capture" profiles, examining all pairwise interactions throughout the tiled locus, involving these viewpoints of interest (**figure 3.6C**). As well as performing this analysis in the WT and R2-only foetal liver cells, I downloaded and reanalysed a previously published mESC tiled-C dataset (Oudelaar et al., 2020). This mESC data served as a non-erythroid baseline, against which I could compare WT and R2-only erythroid chromatin interactions.

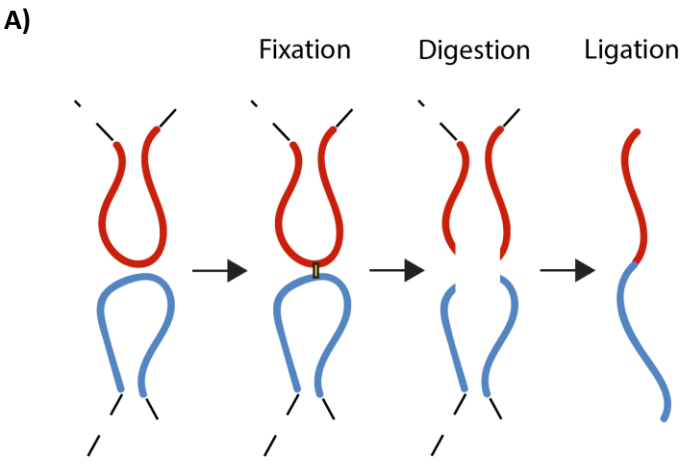
I created virtual capture profiles displaying pairwise interactions between three CTCF sites (two flanking the α-globin sub-TAD, and one situated between the R1 and R2 enhancers) and the rest of the locus. I also created profiles using the R2 enhancer and the α-globin promoters as viewpoints. The two α-globin promoters are identical in sequence save for a single SNP, therefore I considered both as viewpoints simultaneously.

Chromatin interactions between the three CTCF sites and the rest of the tiled locus were unperturbed in the R2-only erythroid cells, demonstrating that the 65kb α-globin sub-TAD still forms in the absence of the R1, R3, Rm and R4 elements (**figure 3.6C**).

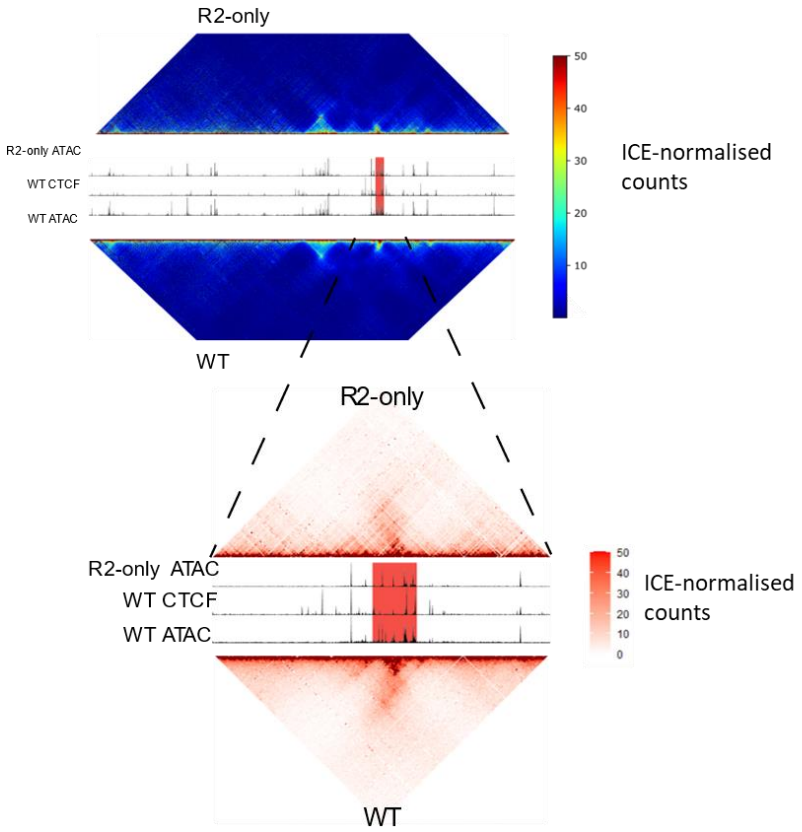
Interestingly, pairwise interaction frequency between each of the CTCF sites and the rest of the tiled locus were largely conserved in mESCs, apart from those throughout the erythroid-specific α-globin sub-TAD and involving the two erythroid-specific θ-globin CTCF sites, located immediately downstream of each α-globin gene.

Interrogation of R2's chromatin interaction profile demonstrated a striking reduction in interaction frequency between R2 and the α-globin promoters, which was corroborated by reciprocal virtual capture from the α-globin promoters (**figure 3.6C**). Although the frequency of interactions between R2 and the α-globin promoters is reduced in R2-only FLs, it is still higher than that in the WT mESC baseline, indicating that R2 can still interact with the α-globin promoters more frequently than expected by chance.

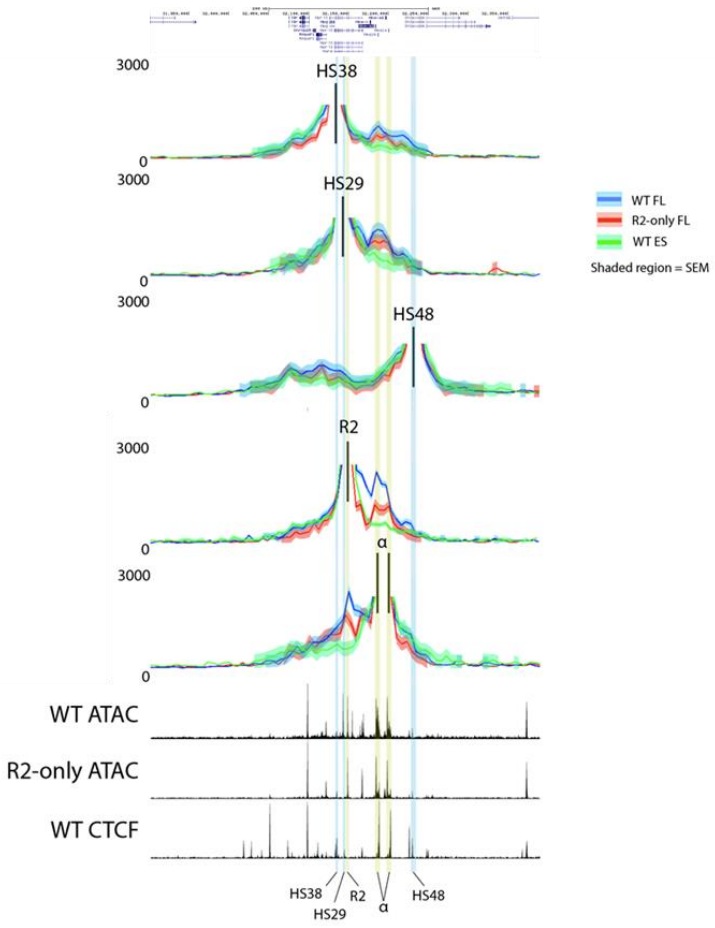
To reduce noise in the virtual capture-C plots, I used loess smoothing (see methods for details). To ensure the reduced interaction frequency between R2 and the α-globin promoters wasn't a smoothing artefact, produced by low signal in the adjacent R1, R3, Rm and R4 bins, I re-plotted R2's interaction profile without loess smoothing. Although noisier, this showed the same pattern of reduced interaction frequency between R2 and the α-globin promoters in R2-only cells (**figure 3.6D**).



B)


R2-only
R2-only ATAC
WT CTCF
WT ATAC
WT
50
40
30
20
10
ICE-normalised counts
R2-only
R2-only ATAC
WT CTCF
WT ATAC
WT
50
40
30
20
10
0
ICE-normalised counts

C)


HS38
HS29
HS48
R2
α
3000
0
WT FL
R2-only FL
WT ES
Shaded region = SEM
WT ATAC
R2-only ATAC
WT CTCF
HS38
HS29
R2
α
HS48

D)

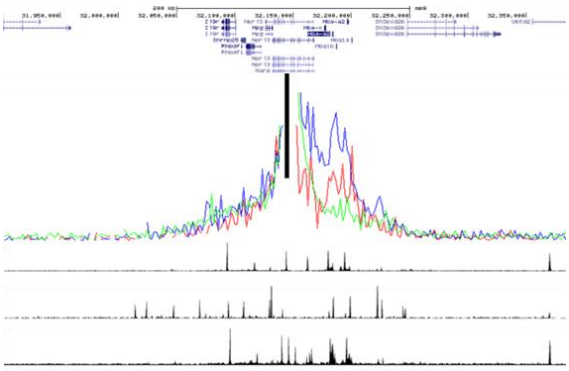


Figure 3.6 Chromatin interactions throughout the α-globin locus in WT and R2-only embryos.

(A) Schematic of typical chromosome conformation capture methodology: DNA fixed using 2% formaldehyde; fixed DNA digested using a specific restriction endonuclease (e.g. DpnII); digested DNA re-ligated using T4 ligase. (B) Tiled-C heatmaps comparing all-vs-all interaction frequency throughout the α-globin locus in R2-only homozygous (n=3) and WT (n=2) foetal liver erythroid cells. Upper: coordinates = chr11:29,900,000–33,230,000. Lower: zoomed heatmap over the α-locus; coordinates = chr11:31900000-32400000. Heatmaps = merged biological replicates. Bins = 2,000bp; tracks = R2-only homozygous ATAC-seq; WT CTCF ChIP-seq; WT ATAC-seq (top-bottom); red = α-locus; colour scale: ICE-normalised counts. (C) Virtual capture plots: pairwise interactions throughout the zoomed tiled locus (chr11:31900000-32400000) in which viewpoints participate. Viewpoints (top-bottom): HS38 CTCF site, HS29 CTCF site, HS48 CTCF site, R2 enhancer, and α-promoters (considered together due to similarity in sequence, see methods). Blue = WT foetal liver erythroid cells (n=2), red = R2-only homozygous foetal liver erythroid cells (n=3), green = mESCs (n=3). Tracks smoothed using loess local regression algorithm (span = 0.05). (D) same as C, but without loess smoothing. Viewpoint = R2 enhancer.

3.7 R2 eRNA transcription

Studies have reported that enhancer (e)RNA transcription (bi-directional transcripts originating from an enhancer) correlates with enhancer activity (P. R. Arnold et al., 2020; Gorbovytska et al., 2022; Sartorelli & Lauberth, 2020). Whether eRNA has a functional role in upregulating target genes, or whether it is simply an indicator of enhancer activity, remains unclear. I re-analysed my poly-A minus RNA-seq data from three R2-only homozygotes and two WT littermates, and visually inspected rpkm-normalised transcripts originating from the R1, R2, R3, Rm and R4 enhancers, using the UCSC genome browser (**figure 3.7A**). R1, R2, R3 and Rm lie within introns of the Nprl3 gene, which is active in erythroid cells and transcribed on the negative strand. Therefore, it was only possible to assay eRNA transcription on the positive strand. There was a striking reduction in eRNA levels over the R2 enhancer in R2-only erythroid cells, and of course a complete loss of eRNA originating from the other four α-globin superenhancer constituents (**figure 3.7A**). To measure the differences in eRNA transcription quantitatively, I performed a virtual qPCR, using Rsubread featurecounts to extract the number of poly-A minus RNA-seq reads mapping to the R2 enhancer

and the HS2 enhancer from the β-globin LCR. In each FOETAL LIVERsample, I normalised R2 eRNA signal to HS2 eRNA. This revealed a ~3-fold reduction in eRNA transcription over the R2 enhancer in R2-only erythroid cells compared to WT (**figure 3.7B**).

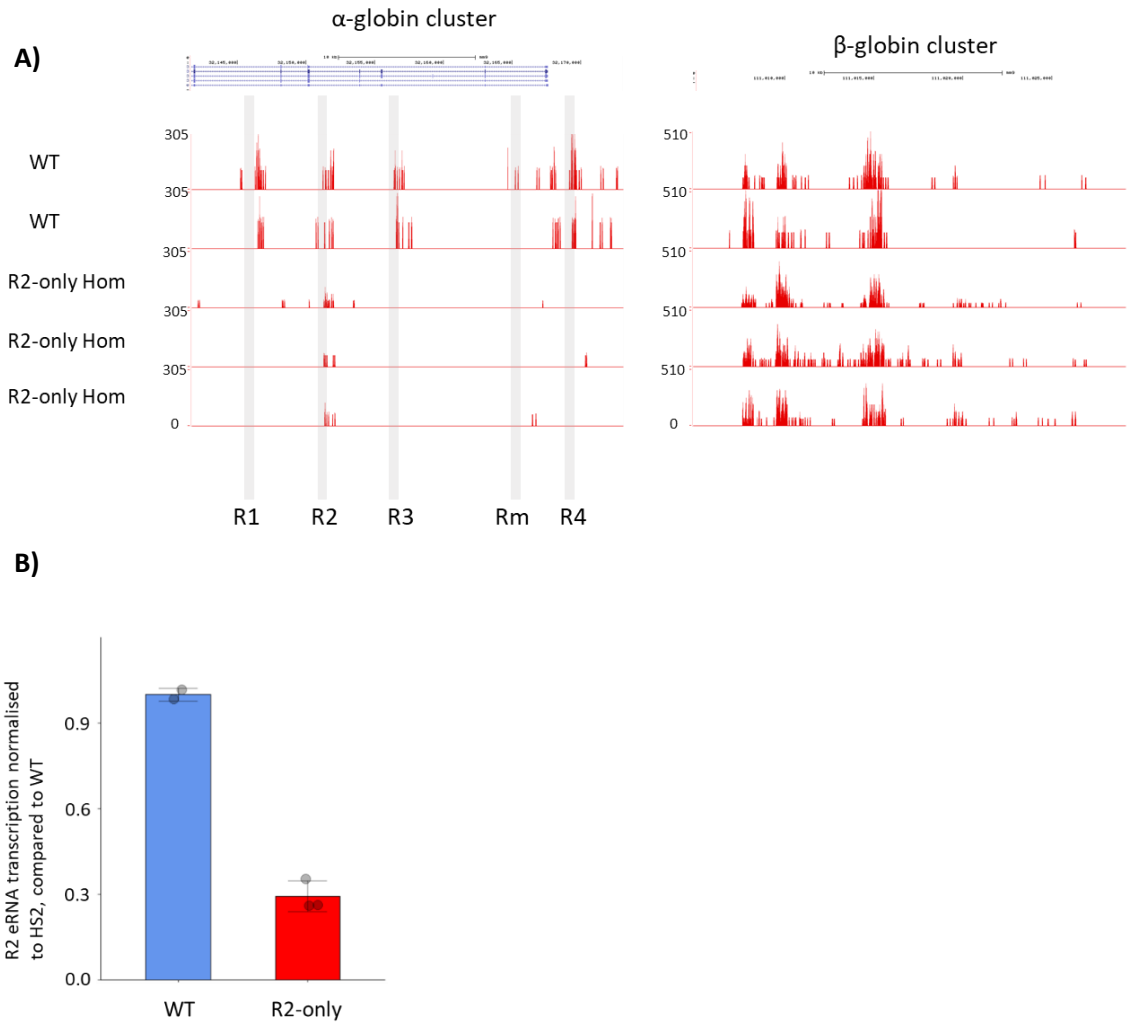


Figure 3.7 Enhancer RNA expression in WT and R2-only embryos.

(A) Poly A minus RNA-seq tracks from WT (n=2) and R2-only homozygous (n=3) foetal livers. Left = α-globin locus; right = β-globin locus. All tracks rpkm normalised. (B) Virtual qPCR comparing R2 enhancer RNA transcription in WT and R2-only foetal liver erythroid cells. R subread featurecounts used to count number of sequencing reads mapping to R2 at the α-globin locus and HS2 at the β-globin locus. R2 eRNA transcription normalized to HS2 eRNA transcription and displayed normalized to WT. Dots = biological replicates; error bars = SE.

3.8 Discussion

When separated from the rest of the α-globin superenhancer, R2 retains the hallmarks of an active enhancer: it remains accessible, it is still decorated by active enhancer-associated histone modifications (albeit to a lesser extent than in WT), and it continues to recruit high levels of erythroid master regulators of transcription. Despite this, R2 alone cannot activate high levels of α-globin transcription.

A number of studies have reported "priming elements", a specific class of superenhancer constituent which play pioneering roles in licensing their clusters for chromatin opening and activation (Long et al., 2016; Maekawa et al., 1989; Shin et al., 2016; Xu et al., 2012). R2 still becomes accessible in R2-only erythroid cells, suggesting that the mutant phenotype is not due to the removal of such a priming element. Instead, R2's accessibility is consistent with it retaining its active enhancer status.

Interpreting the reduced H3K27Ac and H3K4Me1 signal over the R2 enhancer and α-globin promoters is more complicated. Because ChIPmentation is a population-level assay, it is unclear whether the lower levels of H3K27 acetylation and H3K4 mono-methylation correspond to a complete loss of these modifications throughout the entire R2 enhancer in a subset of cells (which would indicate that the R2-only phenotype is the result of a mixed population of cells in which R2 is active, and cells in which it is not), or whether it corresponds to fewer histones being modified throughout the locus in all cells (which would suggest that R2 is active but functionally weaker in most cells – a rheostat-like model). Without single-cell, high resolution assays it will remain difficult to conclude whether reduced histone modifications represent a binary or rheostat phenomenon.

Although H3K27Ac has been reported to be dispensable for enhancer function (Bonn et al., 2012; Catarino & Stark, 2018; Pengelly et al., 2013; Pradeepa et al., 2016), its deposition is strongly correlated with enhancer activity (Bonn et al., 2012; Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2010; Zentner et al., 2011). ChIPmentation is not strictly a quantitative method, but the lower H3K27Ac signal over R2 in the R2-only model does appear to correlate with its diminished ability to up-regulate gene expression.

R2 alone is incapable of recruiting high levels of transcriptional coactivators, Med1 and Brd4, to itself and to its target gene promoters. This alone is a strong indicator of interdependency/cooperativity between the α-globin superenhancer constituents. Several studies have presented evidence that superenhancers establish transcriptional hubs, containing a high density of TFs, coactivators and Pol2, at their target genes (Boija et al., 2018; Gurumurthy et al., 2019; Hnisz et al., 2017; Oudelaar et al., 2018; Sabari et al., 2018). The biochemical processes leading to formation of such structures is debated, but two prominent theories include liquid-liquid phase separation (Boija et al., 2018; Gurumurthy et al., 2019; Hnisz et al., 2017; Sabari et al., 2018) and some form of TF trapping (Sigova et al., 2015; Thomas et al., 2021). Nevertheless, superenhancers contain a high density of TFBSs, facilitating high levels of tissue-specific TF recruitment. These tissue-specific TFs are then thought to cooperatively recruit coactivators, such as the mediator complex. Deleting R1, R3, Rm and R4 from the α-globin locus effectively removes 70% of the total α-globin superenhancer's TFBSs. Therefore, while tissue-specific TFs can still bind at R2, it is unsurprising that the total amount of coactivator recruitment throughout the locus is diminished. Furthermore, the reduction in mediator recruitment to R2 shows that the other four elements make a positive contribution to coactivator recruitment at R2.

TBP recruitment was unaffected in the R2-only model, consistent with its ability to bind DNA at the TATA-box in an enhancer-independent fashion (Cormack & Struhl, 1992; Mishal & Luna-Arias, 2022). Conversely, Pol2 occupancy at the α-globin promoters was clearly reduced, reflecting Pol2's reliance on the mediator complex for effective recruitment and stabilisation at promoters (Soutourina, 2017). Previous publications have reported that enhancers primarily influence transcription at the stages of Pol2 recruitment and transcription initiation (Larke et al., 2021), which is supported by our data. Enhancers have also been implicated in pause-release, the process via which paused Pol2 is released from the promoter, facilitating productive elongation (Gorbovytska et al., 2022). The reduction in Spt5 occupancy, an elongation factor involved in pause-release, could support this theory, although it could also simply be a knock-on effect of reduced Pol2 recruitment and transcription initiation.

Numerous publications have demonstrated high frequency interactions between superenhancers and their cognate target genes (Allahyar et al., 2018; Beagrie et al., 2017; Grosveld et al., 2021; Ing-Simmons et al., 2015; Y. Li et al., 2018; Oudelaar & Higgs, 2021). These interactions appear crucial for effective target gene up-regulation, although the mechanism(s) facilitating superenhancer-target gene interaction, and the spatiotemporal relationship between interaction and activation, remain controversial (Benabdallah et al., 2019; Oudelaar & Higgs, 2021). The frequency of interactions between R2 and the α-globin promoters was reduced in R2-only erythroid cells, but remained higher than that in mESCs. This demonstrates that in erythroid cells, R2 is only partially dependent on R1, R3, Rm and R4 to interact with its target genes; however, that these additional constituents boost interaction frequency significantly.

The cohesin complex includes four proteins: Rad21, SA1/2, SMC1 and SMC3, which together form a ring structure, topologically encircling DNA (Peters et al., 2008). Through a process of bi-directional loop extrusion, cohesin can translocate along DNA, bringing distal loci into close proximity as they are extruded through the cohesin ring. This process is important for the formation and maintenance of TADs and sub-TADs. Recently, studies have reported cohesin loading at active enhancers (Hua et al., 2021; Rinzema et al., 2021). This suggests that enhancers may be important for the formation and/or maintenance of TADs and sub-TADs. Our tiled-C data shows that, in the absence of four out of the five α-globin superenhancer constituents, the 65kb α-globin sub-TAD still forms. This suggests that either: R2 recruits sufficient cohesin to form the sub-TAD; cohesin is loading elsewhere; or mechanisms other than cohesin-mediated loop extrusion are responsible for the formation of the α-globin sub-TAD.

In summary, R2 fulfils the criteria of an active enhancer independently of the rest of the α-globin superenhancer: it becomes accessible, it is decorated with active enhancer-associated histone modifications, it continues to recruit tissue-specific TFs, and it can up-regulate α-globin transcription to some degree. However, once detached from its superenhancer context, R2's enhancer potential is greatly attenuated: coactivator recruitment, promoter interactions, and eRNA transcription all fall sharply, and R2's ability to up-regulate α-globin expression is 5-fold lower than predicted. Physiologically, this represents the difference between a healthy mouse devoid of

pathology, and post-natal lethality. In an additive model of superenhancer cooperation, the α-globin expression outputs of the ΔR2 and R2-only models are expected to sum to 100% of WT α-globin expression; instead, they sum to ~60%. It is clear that the other four α-globin superenhancer constituents, or a subset thereof, are essential for full R2 activity and effective α-globin gene activation. To dissect the contributions of each element and the cooperation manifested between them, I set about rebuilding the α-globin superenhancer from an enhancer-less baseline in multiple informative configurations.

Chapter four: An enhancer titration: How do the five α-globin superenhancer elements cooperate?

4.1 Introduction

In the previous chapter I showed that R2, the strongest enhancer within the α-globin superenhancer, is insufficient to activate high levels of α-globin gene expression. In the absence of the R1, R3, Rm and R4 elements, enhancer-associated histone modifications and coactivator recruitment were reduced over the R2 enhancer, and R2 was less able to interact with the α-globin promoters and to transcribe enhancer RNAs. All of this suggests that other α-globin superenhancer constituents cooperate with R2 to facilitate its full activity.

Several publications have reported non-additive/synergistic cooperation within superenhancer clusters (Brosh et al., 2022; Choi et al., 2021; Hnisz et al., 2015; Hörnblad et al., 2021; Huang et al., 2018; Shin et al., 2016; Thomas et al., 2021). This has been shown through assaying gene expression in various element deletion models, and also by showing that certain superenhancer constituents can boost TF recruitment and H3K27-acetylation at other elements within the same cluster (Hnisz et al., 2015; Huang et al., 2018; Thomas et al., 2021). Huang et al and Thomas et al both described distinct classes of superenhancer constituent: hub enhancers (Huang et al., 2018) and amplifiers (Thomas et al., 2021); both of these element types appear

to potentiate and regulate other elements within their cognate clusters. We speculated that R1, R3, Rm and R4 might have a similar effect on R2, boosting H3K27-acetylation and potentiating its enhancer activity.

We posited that the best way to disentangle the effect each element has within its cognate superenhancer would be to rebuild said cluster from an enhancer-less baseline, in multiple configurations, at the native locus. In this chapter I use a combination of genome-engineering approaches to generate an "enhancer titration" series, rebuilding the native α-globin superenhancer from a locus in which all enhancer-like elements have been removed, in all informative configurations. I then characterise these models using an *in vitro* erythroid differentiation system to unravel the activity of, and functional relationships between, each of the α-globin superenhancer constituent elements.

4.2 Model generation

To investigate the cause of the severe R2-only phenotype, and to probe the functional contributions of the R1, R3, Rm and R4 elements, we generated a series of mutants rebuilding the native α-globin superenhancer from an enhancer-less baseline. This series comprised numerous genetic models – too many to reasonably study *in vivo* – and we therefore combined a streamlined mESC genome-engineering pipeline with an *in vitro* differentiation and erythroid purification system, recently developed in our lab (Francis et al., 2022). This allowed rapid production of genetically engineered mouse erythroid cells, in which we could study the activity of the α-globin superenhancer, and also served as an orthogonal system to verify our existing *in vivo* results.

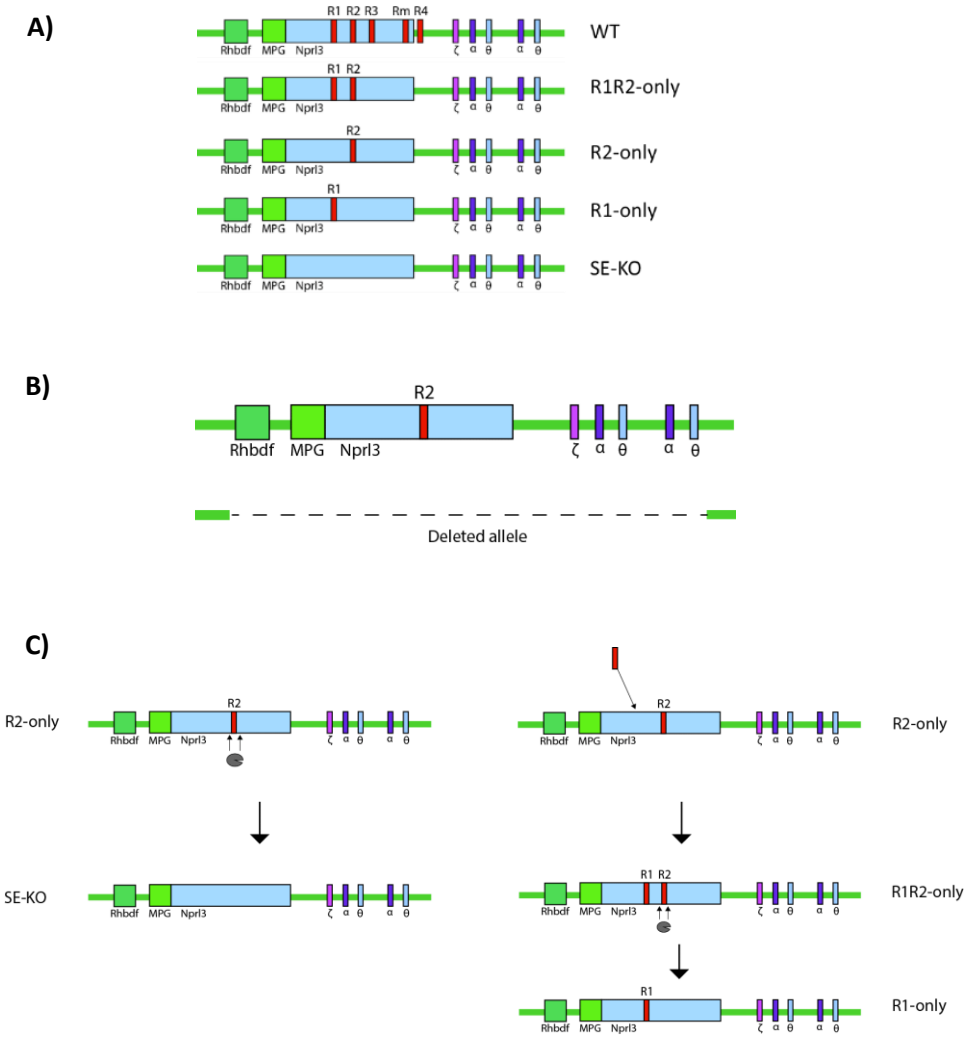
Helena Francis had previously generated several informative models: a model in which all five α-globin superenhancer constituents have been removed from the α-globin locus (SE-KO), R1-only, R2-only, and R1R2-only (**figure 4.1A**). These models were generated in hemizygous R2-only mESCs, in which one allele of the α-globin locus had been removed, using a dual-guide CRISPR deletion strategy (Helena Francis, thesis)

(**figure 4.1B**). Removing one copy of the α-globin locus made all subsequent genome editing steps much simpler, both in terms of editing efficiency and screening. To produce the R2-only mESC model, hemizygous RMGR-competent mESCs were targeted, and the remaining native α-globin locus exchanged (via RMGR) with a BAC-derived R2-only locus (the same process and BAC as used to generate the *in vivo* R2-only model (**figure 3.2A**)). The SE-KO, R1-only and R1R2-only models were all generated from this hemizygous R2-only background, through employment of additional CRISPR-based editing steps. The R2 enhancer was deleted from R2-only cells using dual-guide CRISPR deletion, to generate SE-KO cells (**figure 4.1C, left**); the R1 enhancer was inserted, via CRISPR HDR, at its native position in R2-only cells, to generate R1R2-only cells (**figure 4.1C, right, top**); the R2 enhancer was deleted in R1R2-only cells, to generate R1-only cells (**figure 4.1C, right, bottom**). All of this was performed by Helena Francis. She also generated a WT line, in which the native α-globin locus, in hemizygous mESCs, was replaced with a BAC-derived WT α-globin locus. This allowed direct comparison between each new model and a similarly targeted hemizygous BAC-derived WT model.

To enable like-for-like comparisons between our new models rebuilding the α-globin superenhancer and previous single element deletions, we generated ΔR1, ΔR2, ΔR3, ΔRm, ΔR4 and ΔR1R2 (R3RmR4-only) models in hemizygous mESCs (Helena Francis generated ΔR1, ΔR2 and ΔR1R2; I generated ΔR3, ΔRm and ΔR4) (**figure 4.1D**). To generate each model, I worked with the WIMM genome engineering facility to design CRISPR gRNAs targeting cleavage immediately upstream and downstream of each α-globin superenhancer constituent, using the same deletion coordinates as those used *in vivo* (Hay et al., 2016) (**figure 4.1E(1)**). These gRNAs were then cloned into pX458 vectors, downstream of a U6 promoter (**figure 4.1E(2)**). The pX458 vector also expresses *S. pyogenes* Cas9, and either GFP or mRuby. I co-transfected hemizygous WT mESCs with two pX458 plasmids (one driving cleavage upstream of the targeted element and one downstream) (**figure 4.1E(3)**). ~24 hours later, I FACS sorted GFP/mRuby co-fluorescent cells into individual wells of a 96 well plate (**figure 4.1E(4)**). After 10 days of incubation without disturbance, I split surviving colonies into separate 96 well plates for PCR screening and expansion. I conducted PCR with primers flanking

the targeted element, such that successful deletion clones would render a product ~1kb shorter than clones in which insertion was unsuccessful (**figure 4.1E(4)**). Clones in which deletion was successful were further screened by Sanger sequencing.

I obtained erythroid cells of each genotype using the *in vitro* differentiation (**figure 4.1E(5)**) system and conducted RT-qPCR to measure α-globin gene expression in each model (**figure 4.1E(6)**). Consistent with our previous *in vivo* results, deleting R1 or R2 led to a large reduction in α-globin expression, and deleting both simultaneously, leaving only the R3, Rm and R4 elements within the locus, levelled α-globin expression to ~2% of WT (**figure 4.2A**). Meanwhile, deletion of R3, Rm, or R4 had very little effect on gene expression. This strongly supported our previous conclusion that R1 and R2 are potent activators of transcription, whereas R3, Rm and R4 appear to have little/no conventional enhancer activity.



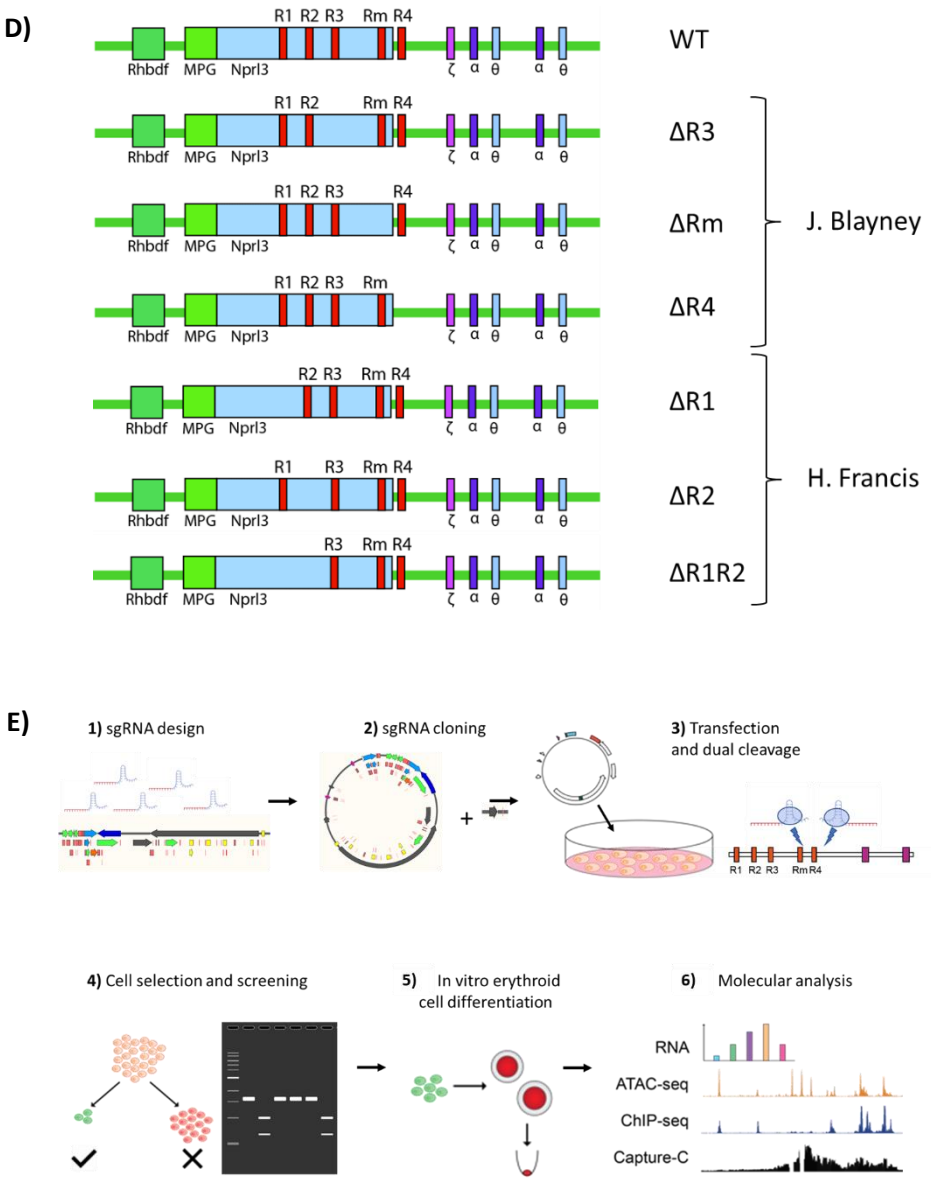


Figure 4.1 Engineering the α-globin locus in mESCs.

(A) Top-bottom: schematic of the WT, R1R2-only, R2-only, R1-only and SE-KO mESC models produced by Helena Francis. (B) Schematic of the hemizygous R2-only mESC model. (C) Schematic of SE-KO mESC model generation (left); schematic of R1R2-only and R1-only mESC models (right). (D) Top-bottom: schematic of the WT, ΔR3, ΔRm, ΔR4, ΔR1, ΔR2, ΔR1R2 mESC models. Models generated by me (J. Blayney) and H. Francis, as indicated. (E) Schematic of genome-engineering workflow, as described in methods.

If each enhancer functions independently of the others, our deletions (in primary mouse *and* EB-derived erythroid cells) suggest that R1-only should drive 40% α-globin, R2-only should activate 50%, and R1R2-only should activate >90% relative to WT (Hay et al., 2016) (**figure 1.4C**). However, none of these models were capable of driving high levels of gene expression. Instead, R1-only and R2-only each upregulated α-globin

expression to ~10% of WT, and both enhancers combined drove ~45% α-globin (**figure 4.2B**).

To detach enhancer-dependent transcription from the intrinsic activity of the α-globin promoters, I assayed α-globin expression in the SE-KO line. Hemizygous SE-KO erythroid cells expressed 0.1% of WT α-globin levels (**figure 4.2B**). Thus, despite being weaker than predicted, R1 and R2 are still potent activators of transcription, each upregulating α-globin 100-fold. R1 and R2 combine super-additively, up-regulating α-globin expression 450-fold.

R3, Rm and R4 have little/no conventional enhancer activity, evidenced by the dearth of α-globin expression in the ΔR1R2 model (**figure 4.2A**). Conversely, deletion of R3, Rm and R4 (in the R1R2-only model) led to a 55% drop in α-globin expression, showing that these three seemingly non-enhancer elements are vital for facilitating full superenhancer activity (**figure 4.2B**). We therefore termed R3, Rm and R4 as "facilitators", due to their ability to facilitate the enhancer activity of R1 and R2.

To dissect the roles of the three facilitator elements (R3, Rm and R4), I generated an "enhancer titration", a series of genetic models, rebuilding the α-globin superenhancer from R1R2-only back up to WT, creating all R3/Rm/R4 permutations between these states. Generating all cluster compositions in this manner allowed scrutiny of each element's contribution to the α-globin superenhancer in the presence/absence of each other constituent, and thereby enabled me to directly probe the cooperation between each of the five elements.

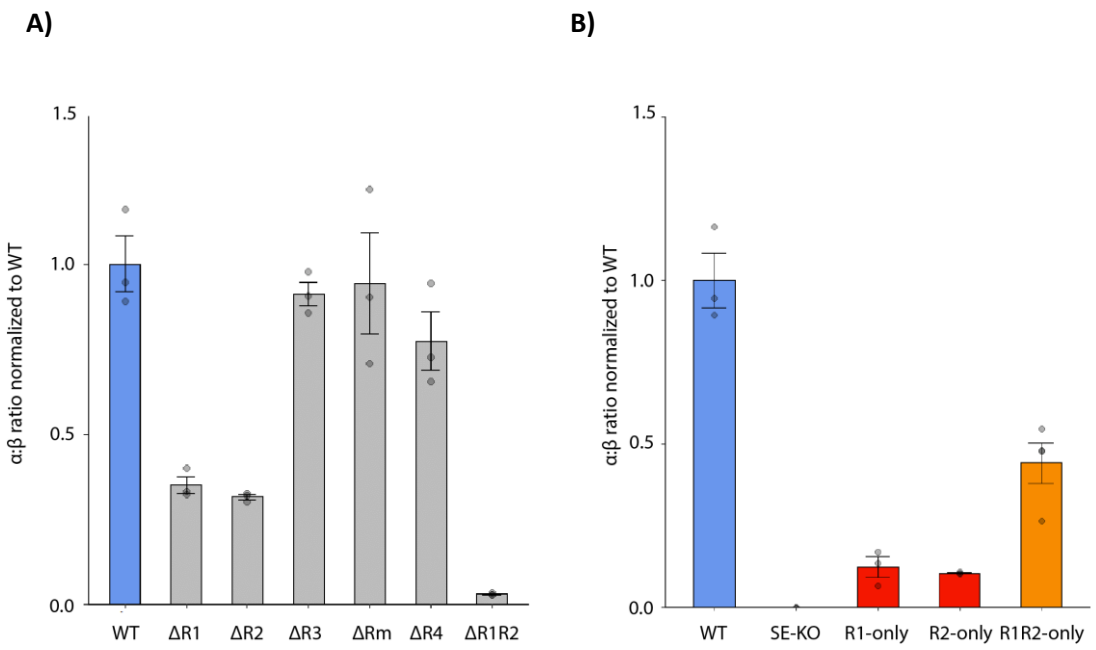


Figure 4.2 α-globin gene expression in EB-derived erythroid cells.

(A) α-globin gene expression in WT, ΔR1, ΔR2, ΔR3, ΔRm and ΔR4 EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. (B) α-globin gene expression in WT, superenhancer knockout (SE-KO), R1-only, R2-only and R1R2-only EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE.

To generate these models, I used a dual-guide CRISPR strategy, removing enhancers individually, or in combination, from hemizygous BAC-derived WT mESCs. Building the series in this background made editing efficient, and made my models comparable to the previously generated hemizygous SE-KO, R1-only, R2-only and R1R2-only models. To create the first round of mutants, I deleted either R3, Rm, or R4 from the WT locus, producing hemizygous ΔR3, ΔRm and ΔR4 models, respectively. I then re-targeted these models for deletion of a second element, deleting R4 in ΔRm cells, R3 in ΔR4 cells, and Rm in ΔR3 cells, to generate R1R2R3-only, R1R2Rm-only and R1R2R4-only models, respectively. I produced at least three separately targeted clones for each model, which were then handled as biological replicates. In total, I had ≥3 replicates for fourteen mESC models, rebuilding the α-globin superenhancer from an enhancer-less baseline (**figure 4.3A**).

4.3 Enhancer titration characterization:

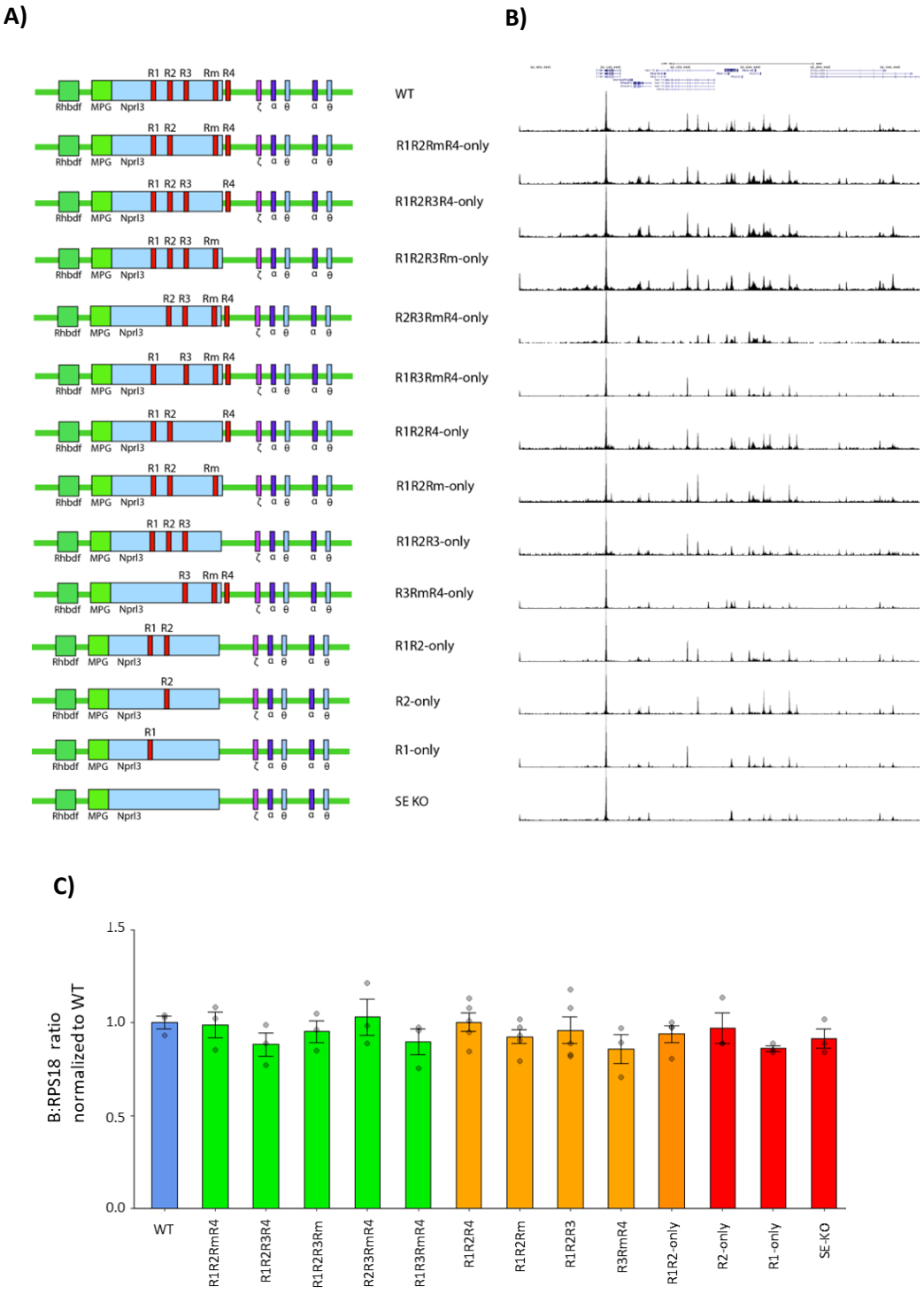
I screened each genetic model by PCR and Sanger sequencing, and performed ATAC-seq on differentiated erythroid cells of each genotype (**figure 4.3B**). This confirmed that the wider α-globin locus retained its integrity in each model, and that there were no gross genome-wide changes in DNA accessibility. The fact that in each model the appropriate α-globin superenhancer constituents become accessible suggests that each element can be activated and made accessible independently of the other α-globin superenhancer constituents.

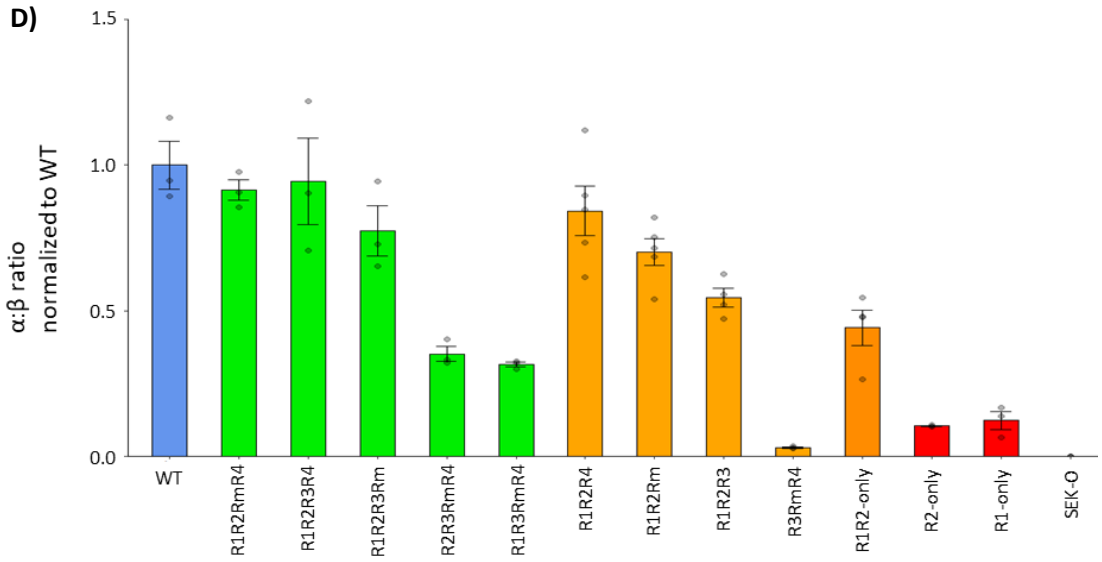
I conducted ≥2 separate EB differentiations with each clone, and performed RT-qPCR on CD71+ erythroid material, assessing α-globin expression. In each case, I normalized α-globin expression to β-globin. I also assayed expression of the housekeeper RPS18, as a non-erythroid control. Across all clones, the β-globin:RPS18 ratio remained approximately consistent (**figure 4.3C**), whereas the α-globin:β-globin ratio varied between models (**figure 4.3D**).

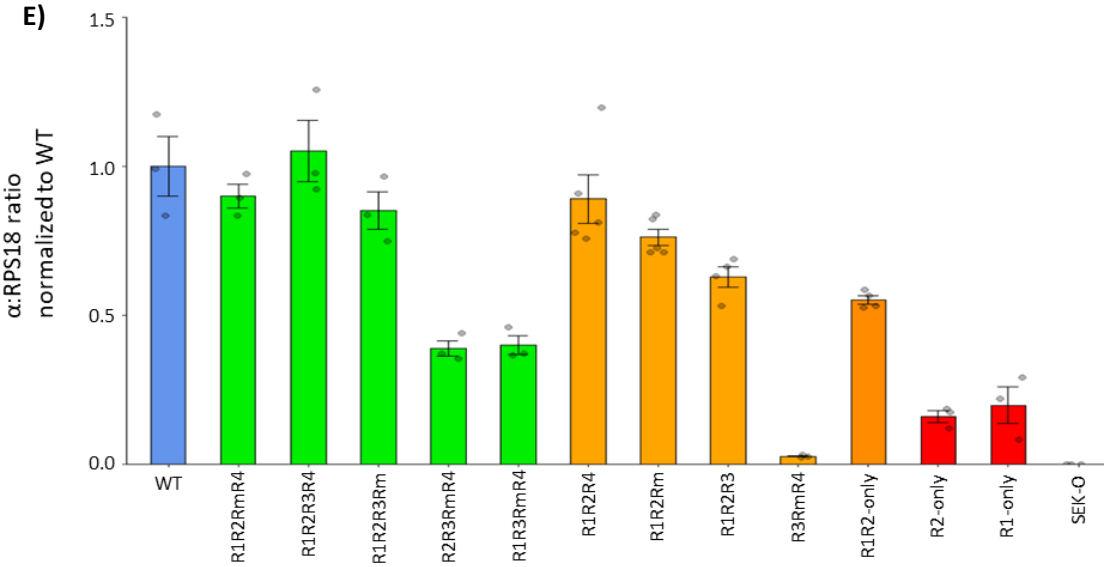
R1R2-only erythroid cells expressed ~45% α-globin compared to WT. Interestingly, reintroducing the R3 element (R1R2R3-only) had little effect, up-regulating expression to ~55% (R3: 10% rescue) (**figure 4.3D**). Reintroducing Rm or R4 had a much larger impact; R1R2Rm-only cells drove expression to ~70% (Rm: 25% rescue), and R1R2R4-only cells drove expression to ~85% of WT (R4: 40% rescue) (**figure 4.3D**). The same results were observed when normalising α-globin gene expression to expression of RPS18 (**figure 4.3E**).

Rebuilding the superenhancer further, by generating R1R2R3Rm, R1R2R3R4 and R1R2RmR4 models, revealed complex context-dependency within the superenhancer. In each model, re-introducing R3 had a small but consistent effect, always rescuing gene expression by ~10% (**figure 4.3D**; **figure 4.3F**). However, re-introducing Rm into a cluster already containing the R4 element (e.g. reintroducing Rm into R1R2R4 cells) attenuated its rescue capacity from ~25% to ~10% (**figure 4.3D**; **figure 4.3G**). Similarly, reintroducing R4 into a cluster containing Rm (e.g. reintroducing R4 into R1R2R3Rm cells) reduced its rescue potential from ~40% to ~20% (**figure 4.3D**; **figure 4.3H**). This

reveals partial redundancy between Rm and R4, compared to R3 which appears to rescue the R1R2-only model in a small, but consistent manner.




D)
α:β ratio normalized to WT
1.5
1.0
0.5
0.0
WT
R1R2RmR4
R1R2R3R4
R1R2R3Rm
R2R3RmR4
R1R3RmR4
R1R2R4
R1R2Rm
R1R2R3
R3RmR4
R1R2-only
R2-only
R1-only
SEK-O


E)
α:RPS18 ratio normalized to WT
1.5
1.0
0.5
0.0
WT
R1R2RmR4
R1R2R3R4
R1R2R3Rm
R2R3RmR4
R1R3RmR4
R1R2R4
R1R2Rm
R1R2R3
R3RmR4
R1R2-only
R2-only
R1-only
SEK-O

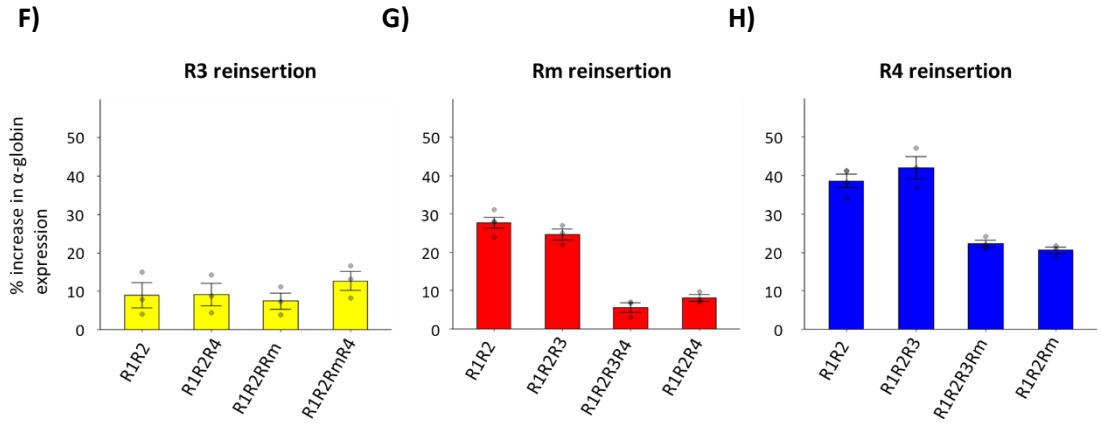

F)
G)
H)
R3 reinsertion
Rm reinsertion
R4 reinsertion
% increase in α-globin expression
50
40
30
20
10
0
R1R2
R1R2R4
R1R2RRm
R1R2RmR4
R1R2
R1R2R3
R1R2R3R4
R1R2R4
R1R2
R1R2R3
R1R2R3Rm
R1R2Rm

Figure 4.3 The enhancer titration: rebuilding the native α-globin locus from an enhancer-less baseline.

(A) Schematic of the enhancer titration: fourteen genetic models rebuilding the α-globin superenhancer in hemizygous mESCs. All models screened by PCR, Sanger sequencing and ATAC-seq. (B) ATAC-seq in EB-derived erythroid cells in the corresponding enhancer titration models (n≥3). Tracks = merged biological replicates. (C) β-globin gene expression in enhancer titration EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to RPS18 and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. (D) α-globin gene expression in enhancer titration EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. (E) α-globin gene expression in enhancer titration EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to RPS18 and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE. (F) Percentage increase in α-globin expression following R3-reinsertion in each corresponding genetic background, as assayed by RT-qPCR in EB-derived erythroid cells. (G) Percentage increase in α-globin expression following Rm-reinsertion in each corresponding genetic background, as assayed by RT-qPCR in EB-derived erythroid cells. (H) Percentage increase in α-globin expression following R4-reinsertion in each corresponding genetic background, as assayed by RT-qPCR in EB-derived erythroid cells.

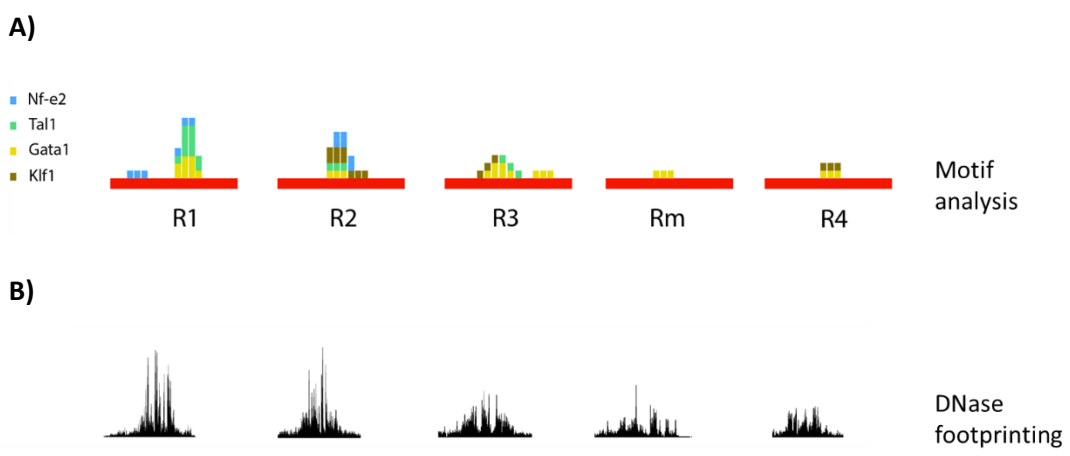
To investigate why Rm and R4 are capable of rescuing gene expression to a greater degree than R3, I performed motif analysis searching for erythroid TF motifs occurring throughout each of the five α-globin superenhancer constituents. This entailed downloading the HOCOMOCO position weight matrices for four erythroid master regulators: Gata1, Nf-e2, Klf1 and Tal1, and using the fimo tool (from the MEME suite) to call high confidence ($P < 1x10^4$) motif occurrences throughout the R1, R2, R3, Rm and R4 sequences. As predicted, R1 and R2 contained the most erythroid TF motifs, but out of the three remaining elements, R3 displayed the most motifs, both in terms of absolute number and TF diversity (**figure 4.4A**). This was counter-intuitive, as many studies have demonstrated a strong correlation between TF recruitment and enhancer strength, with the strongest enhancers tending to recruit diverse cohorts of TFs (Sahu et al., 2022; Singh et al., 2021; Spitz & Furlong, 2012; Zinzen et al., 2009). It is possible that Rm and R4 could recruit unknown factors; however, we lack the power to perform a convincing *de novo* motif search over each of the five elements.

To examine potential TF binding at each element in a more "unbiased" fashion, I re-analysed a DNase-seq dataset produced by Maria Suciu (a previous PhD student in the Higgs lab). The DNase footprints over the five α-globin superenhancer constituents were consistent with R1, R2 and R3 recruiting more TFs than Rm and R4 (**figure 4.4B**).

It is still plausible that R4 could recruit other factors that we are unaware of; however, our data suggests that R4's greater rescue potential, relative to R3 and Rm, is not exclusively encoded in its ability to recruit erythroid-specific TFs.

To test R4's effect on the activity of R1 and R2 directly (not simply through measuring α-globin gene expression), I performed ChIPmentation, with an antibody against H3K27Ac, in WT, R1R2-only and R1R2R4 EB-derived erythroid cells. H3K27Ac is a histone modification associated with active enhancers, and several groups have reported that levels of H3K27-acetylation can be used as an indicator of enhancer activity (Bonn et al., 2012; Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2010; Zentner et al., 2011). The level of H3K27Ac throughout the α-globin locus was much lower in R1R2-only cells compared to WT, with a striking difference over the two enhancers (**figure 4.4C**). Re-inserting R4 had a dramatic effect; this led to an increase in H3K27-acetylation over the position of R4, but a particularly large increase in acetylation at R1 and R2 (**figure 4.4C**). This strongly suggests that R4 has a major positive effect on the two strong enhancers: R1 and R2.

There is an inverse correlation between the ability of R3, Rm and R4 to rescue α-globin expression and their distances from the α-globin promoters (**figure 4.4D**). This led us to ask whether R4's position (as the element located closest to the α-globin promoters) underlies its potential to facilitate higher levels of gene expression.



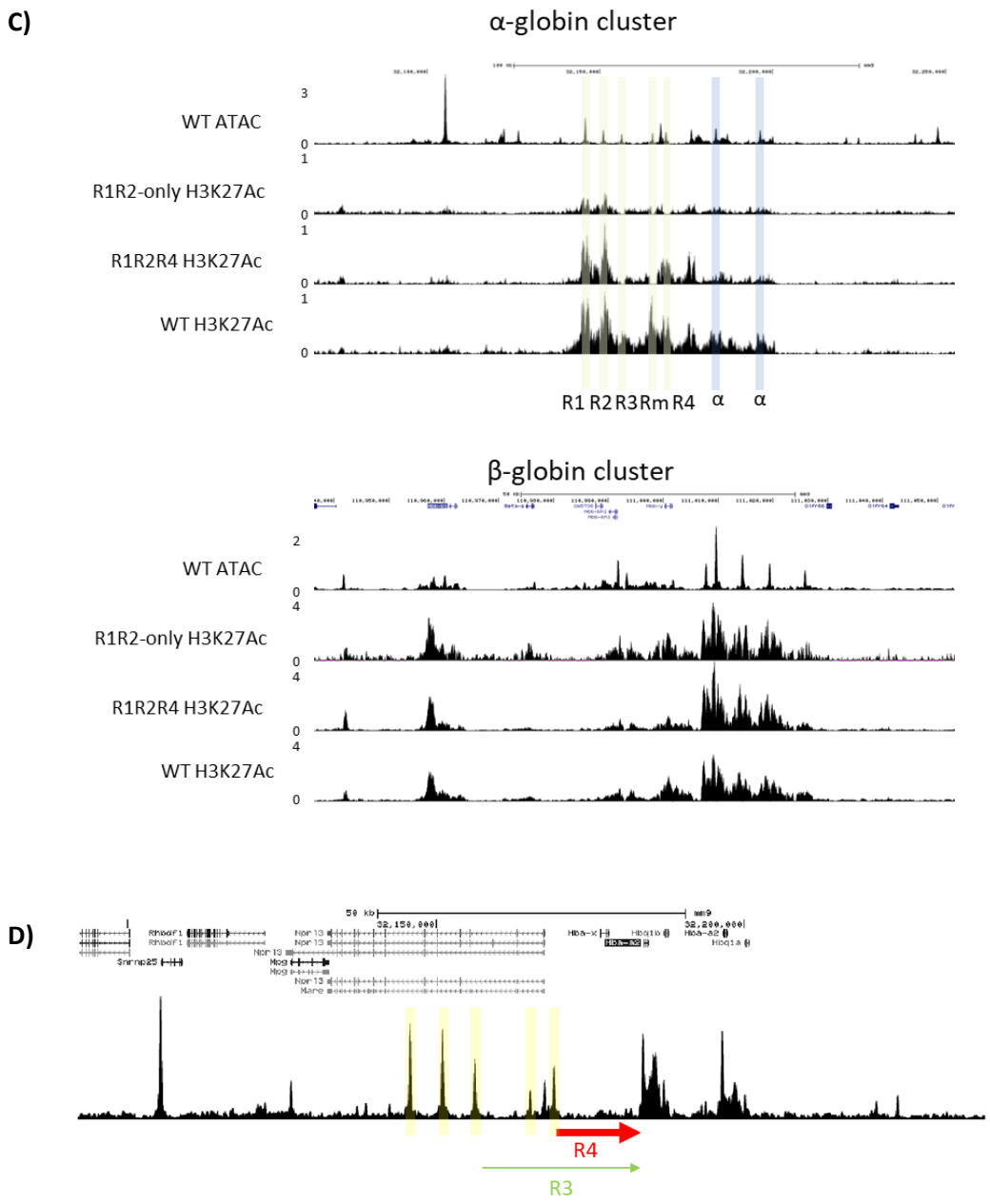


Figure 4.4 Investigating the hierarchy among facilitator elements.

(A) FIMO motif analysis conducted on each α-globin superenhancer constituent, searching for occurrences of Gata1, Nf-e2, Tal1 and Klf1 motifs. (B) DNase foot-printing over each of the α-globin superenhancer constituents. (C) Top-bottom: ATAC-seq in WT EB-derived erythroid cells (n=3); H3K27Ac ChIPmentation in R1R2-only (n=1), R1R2R4 (n=1) and WT (n=1) EB-derived erythroid cells. Tracks = merged biological replicates. Above: α-globin locus; below: β-globin locus. All tracks cpm normalised. (D) ATAC-seq in WT erythroid cells (n=3, merged). Green arrow indicates distance between R3 and the superenhancer proximal α-globin gene (~27kb); red arrow indicates distance between R4 and the superenhancer proximal α-globin gene (~14kb). Arrow thickness indicates ability of R3/R4 to rescue α-globin gene expression.

4.4 Discussion

R2-only EB-derived erythroid cells show the same reduction in α-globin expression as seen in R2-only mice. α-globin expression in EB-derived ΔR1, ΔR2, ΔR3, ΔRm, ΔR4 and ΔR1R2 erythroid cells also recapitulated previous *in vivo* results (Hay et al., 2016).

Deleting R1 or R2 individually caused a 50-60% drop in α-globin transcription, respectively; however, R1-only and R2-only models could only drive ~10% α-globin compared to WT erythroid cells. Deleting R1 and R2 in concert caused a 98% loss of transcription, and yet R1R2-only erythroid cells expressed only 43% α-globin compared to WT. If the five constituents of the α-globin superenhancer functioned independently of one another, then summing α-globin expression in the R1R2-only and ΔR1R2 models should equal the expression levels observed in WT erythroid cells; in actual fact, this only sums to 45% of total WT α-globin expression. Deleting R1 and R2 shows unequivocally that R3, Rm and R4 have very little/no intrinsic conventional enhancer activity, and this has also been demonstrated using various reporter assays (Hay et al., 2016). Conversely, deleting R3, Rm and R4 in tandem shows that they are still essential for full superenhancer activity.

In 2021, Thomas et al described the PE "amplifier" element at the Fgf5 locus (Thomas et al., 2021). Similar to R2, PE deletion caused a large reduction in target gene expression; however, PE's ability to drive gene expression was low in the absence of the other four Fgf5 enhancers. Thomas et al suggested that PE is a non-canonical enhancer and that it mainly functions to "amplify" the effects of the other four canonical enhancer elements. They observed that the level of Pol2 found over each Fgf5 superenhancer constituent inversely correlated with the distance between that element and the Fgf5 promoter, and suggested that PE, the Fgf5-proximal element, may act to "trap" Pol2 which has been released from the promoter and to recycle it to the canonical enhancers and the Fgf5 promoter. Though interesting, it is difficult to fully interpret the results at the Fgf5 cluster, due to the pleiotropic effects of altering Fgf5 gene expression, and the fact that it is unclear whether PE acts as an alternative promoter rather than an enhancer-like element. We see very little Pol2 accumulation over any of the α-globin superenhancer constituents (**figure 3.6C**); however, we do

see an inverse correlation between facilitator strength and distance to the α-globin promoters (**figure 4.4D**). The results presented at the Fgf5 locus support our conclusion that superenhancers can be comprised of multiple functionally distinct element types. A number of studies have presented evidence of superenhancer synergy and elements which have roles beyond simply enhancing their target gene's expression (Brosh et al., 2022; Choi et al., 2021; Hnisz et al., 2015; Hörnblad et al., 2021; Huang et al., 2018; Shin et al., 2016; Thomas et al., 2021), for example hub enhancers which are thought to activate and/or coordinate other elements within their cognate superenhancers (Huang et al., 2018).

In *Drosophila,* the Levine group have described "tethering elements": elements which show some of the hallmarks of enhancers (chromatin accessibility and TF-binding), but no intrinsic enhancer activity in reporter assays (Calhoun et al., 2002; Levo et al., 2022). Tethering elements are usually located between enhancers and their target genes, in a promoter-proximal position, such that they can physically tether canonical enhancers to their target genes and thereby promote gene activation (Batut et al., 2022; Levo et al., 2022). The R3, Rm and R4 facilitator elements display some resemblance to tethering elements; however, the mechanism(s) of facilitator function remain unclear. Nevertheless, tetherers demonstrate the existence and importance of elements other than promoters, enhancers and boundary elements, for establishing and/or maintaining appropriate gene expression.

To explore facilitator function further, we wanted to address R4's greater facilitator function relative to R3. Our first question was whether R4's superior facilitator potential is derived from its sequence or its position.

Chapter five: Enhancer rescue: position or sequence?

5.1 Introduction

R3, Rm and R4 have very little intrinsic enhancer activity when assayed *in vivo*, *in vitro* and in transgenic reporter assays (**figure 4.3D**; Hay et al., 2016). In the previous chapter, I demonstrated that these three elements can rescue gene expression in

R1R2-only erythroid cells, to varying degrees (**figure 4.3E**; **figure 4.3F**; **figure 4.3G**). They appear to facilitate the activities of the strong enhancers, R1 and R2. Of these elements, R4 appears to be the most effective facilitator, causing a ~40% increase in α-globin expression when inserted into R1R2-only cells.

There is a positive correlation between TFBS number and enhancer activity, in canonical enhancers (Sahu et al., 2022; Singh et al., 2021; Spitz & Furlong, 2012; Zinzen et al., 2009). R4 shows little canonical enhancer activity and also contains fewer TFBSs than the relatively weaker R3 and Rm elements (**figure 4.4A**). Both of these observations support our conclusion that R3, Rm and R4 are not classical enhancers, but "facilitators" of enhancer activity. The rescue potential of R3, Rm and R4 inversely correlates with their distances from the α-globin promoters: R4 is located closest to the α-globin promoters and is the most potent facilitator; R3 is located furthest from the promoters and is the weakest facilitator.

A number of studies have reported a relationship between the genomic distance separating regulatory elements and their target genes, and target gene expression levels. For instance, Zuin et al (2022) recently demonstrated that moving an enhancer closer to its cognate promoter increases target gene transcription (Zuin et al., 2022), and Thomas et al (2021) showed that Pol2 occupancy over regulatory elements (enhancers and so-called "amplifiers") is inversely correlated with distance to the closest active promoter (Thomas et al., 2021). Conversely, other studies have shown that enhancers can activate target genes located over 1Mb away (Lettice et al., 2003; Long et al., 2016). This shows that the linear genomic distance between a regulatory element and promoter *can* be important for establishing high levels of gene activation, but that enhancers must be examined on a case-by-case basis.

R4's main activity appears to be in boosting the activities of R1 and R2. This is illustrated by the effect R4 reinsertion has on H3K27-acetylation. Reinserting R4 into R1R2-only cells caused a moderate increase in acetylation over the R4 element itself, but a much greater increase over the R1 and R2 elements (**figure 4.4C**).

The combined observations that R4's activity correlates with its position, and R4 predominantly acts to boost R1 and R2 activity, led us to speculate that R4 might have

a role in positioning R1 and R2, and potentially increasing the frequency of interactions between R1 and R2, and the α-globin promoters. In this chapter, I test whether R4's ability to rescue α-globin expression to a higher degree than R3 and Rm is mostly determined by its sequence or its position, and then investigate whether artificially increasing the frequency of interactions between R2 and the α-globin promoters is sufficient to overcome the R2-only phenotype.

5.2 Testing R3 in the position of R4

To distinguish whether R4's sequence or its position predominantly underlies its rescue potential, I generated a model in which R1 and R2 are maintained in their native positions and R3 is inserted into the position of R4 (R1R2R3[R4]) (**figure 5.1A**), with all other elements (Rm and R4) removed from the locus. I reasoned that if R4's sequence determines its rescue potential, for example, if it recruits an unknown TF to the locus, then placing R3 in its position shouldn't rescue gene expression beyond the R1R2R3-only level; however, if R4's position, rather than its sequence, primarily underlies its capacity to rescue gene expression, then inserting R3 in its place might rescue gene expression to a higher degree than R1R2R3-only.

I used a CRISPR-HDR strategy to generate hemizygous R1R2R3[R4] mESCs. The WIMM genome engineering facility helped design a gRNA targeting Cas9 cleavage at the breakpoint where R4 had previously been located, in R1R2-only cells. This gRNA was then cloned into pX458 plasmids, which also express *S. pyogenes* Cas9, and GFP. We also designed an HDR donor, encoding the R3 element flanked by 500bp homology arms corresponding to the regions either side of the R4 deletion breakpoint. This allowed insertion of the R3 element directly into R4's native position (**figure 5.1A**). We inserted a Sal1 restriction site immediately 5' of the R3 sequence in the HDR donor, and an Mlu1 restriction site directly 3'. This enabled efficient exchange of the R3 element for other sequences (e.g. the R2 enhancer) within the HDR donor vector. Finally, we inserted a single base pair substitution within the HDR donor to destroy the gRNA protospacer adjacent motif (PAM) upon successful insertion. I used the SASQUATCH and JASPAR *in silico* tools to screen all newly generated sequences: the

junctions between R3 and the flanking homology arms, the sequences encoding the Sal1 and Mlu1 restriction sites, and the PAM destroying substitution. This confirmed that there were no novel peaks of accessibility or TF motifs predicted to arise at these positions. I also verified the gRNA-containing pX458 vector and the R3 HDR vector sequences by restriction digest and Sanger sequencing.

To produce the R1R2R3[R4] model, I followed the same protocol as used when generating deletion mutants (**figure 4.1E**); however, rather than transfecting R1R2-only with two pX458 plasmids, I co-transfected them with the R4-targeting pX458 plasmid and the R3 HDR donor vector, described above. ~24 hours later, I FACS sorted GFP-positive cells into individual wells of a 96 well plate. After 10 days of incubation without disturbance, I split surviving colonies into separate 96 well plates for PCR screening and expansion. I conducted PCR with primers flanking the R3 construct, such that successful insertion clones would render a product ~1kb longer than clones in which insertion was unsuccessful. Clones in which insertion was successful were further screened by Sanger sequencing. I obtained five R1R2R3[R4] clones which were then handled as biological replicates.

I performed ATAC-seq in EB-derived R1R2R3[R4] erythroid cells, which demonstrated that R3 had been successfully transplanted to the position of R4, and that there were no changes in accessibility surrounding the α-globin locus (**figure 5.1B, upper**). Next, I performed RT-qPCR on the R1R2R3[R4] erythroid cells. R3 in its native position had previously been shown to rescue R1R2-only expression by ~10%; however, transplanting R3 to the position of R4 had a dramatic effect, leading to a rescue of ~50% (a 120% increase in expression relative to R1R2-only) (**figure 5.1C**). This boosted the R1R2-only expression level of 43% compared to WT to ~95% - a near complete rescue of gene expression. This strongly indicates that having a facilitator at/near R4's position – close to the α-globin promoters – is important for rescuing R1R2-only gene expression, whereas R4's sequence is dispensable.

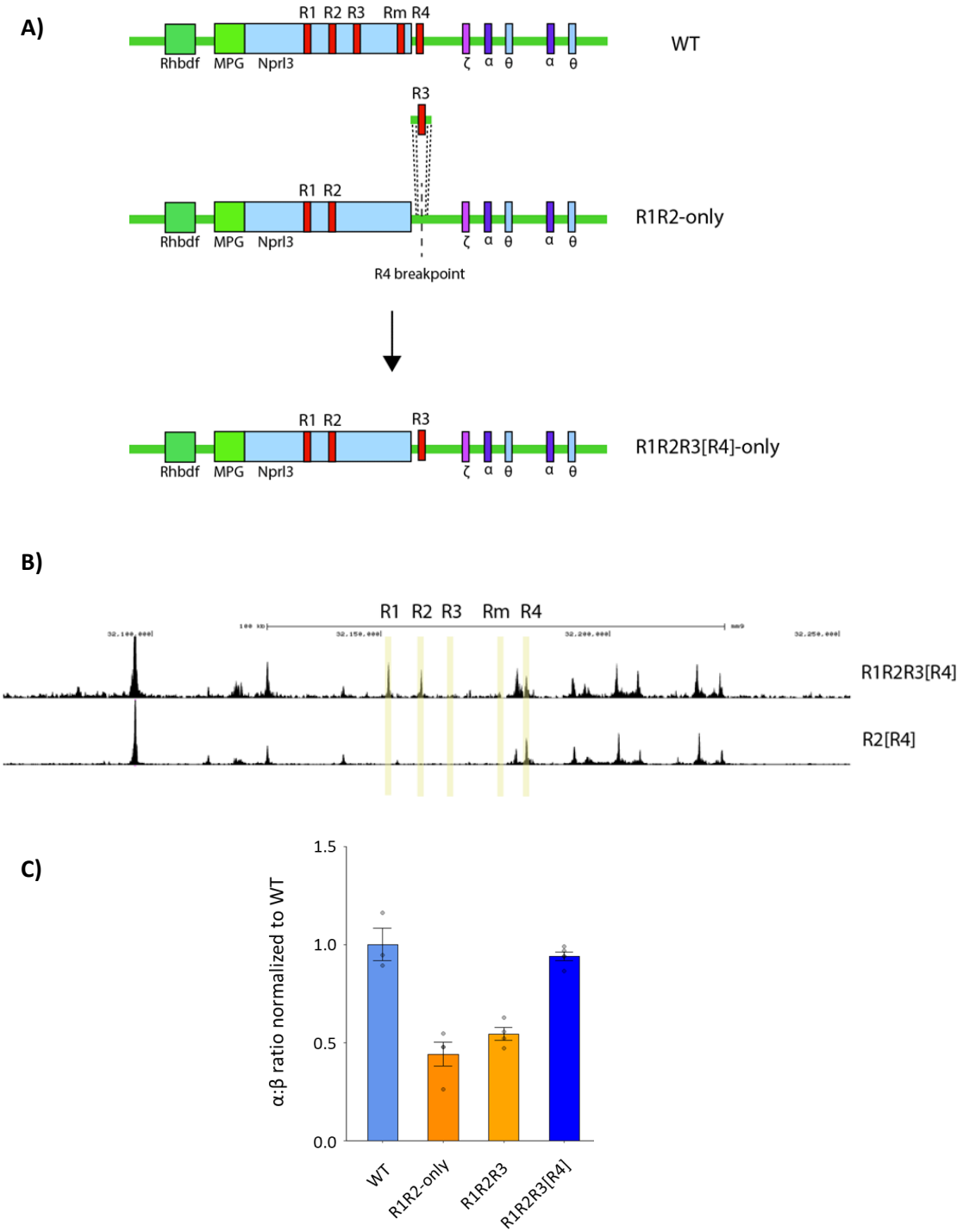


Figure 5.1 Investigating the importance of R3's sequence and position in determining its facilitator strength.

(A) Schematic of generation of the R1R2R3[R4] model. Hemizygous R1R2-only cells lipofected with pX458 expressing an sgRNA targeting the R4 deletion breakpoint and an HDR vector consisting of the R3 element flanked by 500bp homology arms complementary to sequences flanking the R4 deletion breakpoint. Clones FACS sorted based on GFP-fluorescence and genotyped by PCR and Sanger sequencing. (B) ATAC-seq conducted on R1R2R3[R4] EB-derived erythroid cells (top) (n=3) and R2[R4] EB-derived erythroid cells (bottom) (n=2). Tracks = merged biological replicates. (C) α-globin gene expression in WT, R1R2-only, R1R2R3 and R1R2R3[R4] EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE.

5.3 Moving R4 to the position of R1

Helena Francis had previously generated another model testing position versus sequence. She modified hemizygous R2-only mESCs, by inserting the R4 element at the position of R1 (the element located furthest from the α-globin promoters) (R2R4[R1]) (Helena Francis, thesis) (**figure 5.2A**). I repeated RT-qPCR on EB-derived R2-only and R2R4[R1] erythroid cells. This reproduced Helena's finding that re-introducing R4 at the position of R1 has no positive effect on α-globin expression (**figure 5.2B**). This shows that R4's sequence, when moved further from the α-globin promoters, is insufficient to rescue R2-only gene expression. This supports our conclusion that R4's greater rescue potential relative to R3 and Rm is based on its position rather than its sequence.

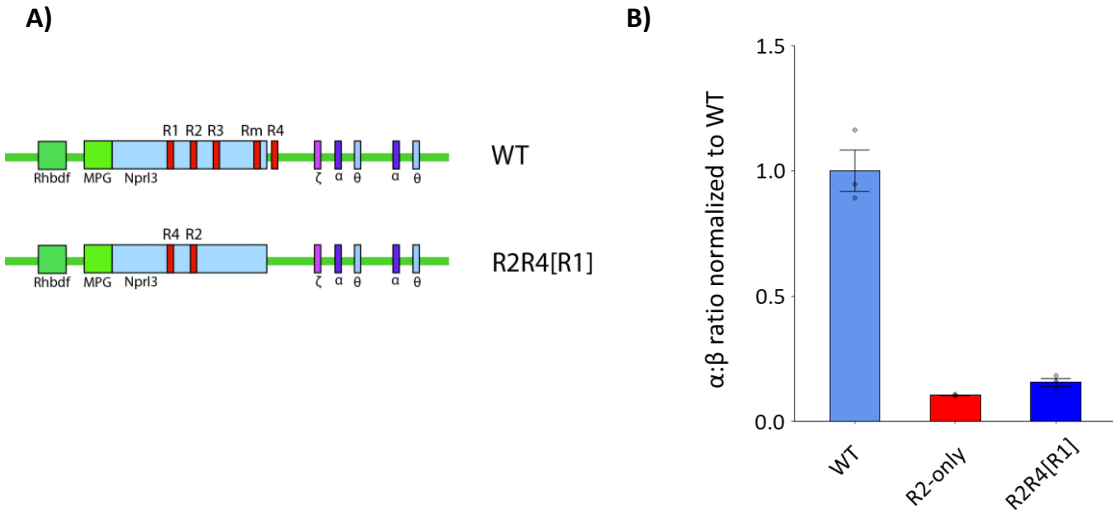


Figure 5.2 Investigating the importance of R4's sequence and position in determining its facilitator strength.
(A) Schematic of the R2R4[R1] model, generated by Helena Francis (H. Francis, thesis). (B) α-globin gene expression in WT, R2-only and R2R4[R1] EB-derived erythroid cells (n≥3) assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE.

5.4 Moving R2 closer to the α-globin promoters

Given that facilitator position appears to influence overall superenhancer activity, I hypothesised that the facilitators might act to bring the activators (R1 and R2) into closer proximity with the α-globin promoters. To explore this hypothesis, I generated

another model in which R2 is inserted into the position of R4, in the absence of the R1, R3, Rm and R4 elements (R2[R4]) (**figure 5.3A**). I reasoned that this would artificially increase interaction frequency between R2 and the α-globin promoters, through diffusion (as previously demonstrated by Zuin et al (2022), with the Sox2 enhancer and promoter (Zuin et al., 2022)).

To generate this model I digested the R3 HDR vector, which I had previously used to generate the R1R2R3[R4] model, using Sal1 and Mlu1. I then cloned the R2 sequence into the HDR vector. The new HDR donor therefore consisted of R2 flanked by homology arms targeting recombination into the position of R4 (**figure 5.3A**). I verified the new R2 HDR donor sequence by Sanger sequencing. Next, I co-transfected a hemizygous SE-KO clone with the R4-targeting pX458 plasmid (the same gRNA sequence as that used for generating the R1R2R3[R4] model), and the R2 HDR donor. FACS sorting and screening was carried out in the same manner as used in creating other models (**figure 4.1E**). I obtained two successful R2[R4] clones, which I treated as biological replicates.

I differentiated both clones and isolated CD71+ erythroid cells for ATAC-seq and RT-qPCR. ATAC-seq confirmed successful integration of R2 at the position of R4, and verified that the gross integrity of the locus remained intact (**figure 5.1B, lower**). I predicted that if the facilitator elements simply act to bring R2 into closer proximity with the α-globin promoters, then transplanting R2 to the position of R4 should have a similar effect. Surprisingly, α-globin expression in EB-derived R2-only and R2[R4] erythroid cells was equivalent (R2-only = 11%; R2[R4] = 12%) (**figure 5.3B**). This shows that simply increasing the likelihood of interaction between R2 and the α-globin promoters is insufficient to restore high levels of α-globin gene expression, and indicates that the other four α-globin superenhancer constituents may play roles beyond simply positioning the R2 element.

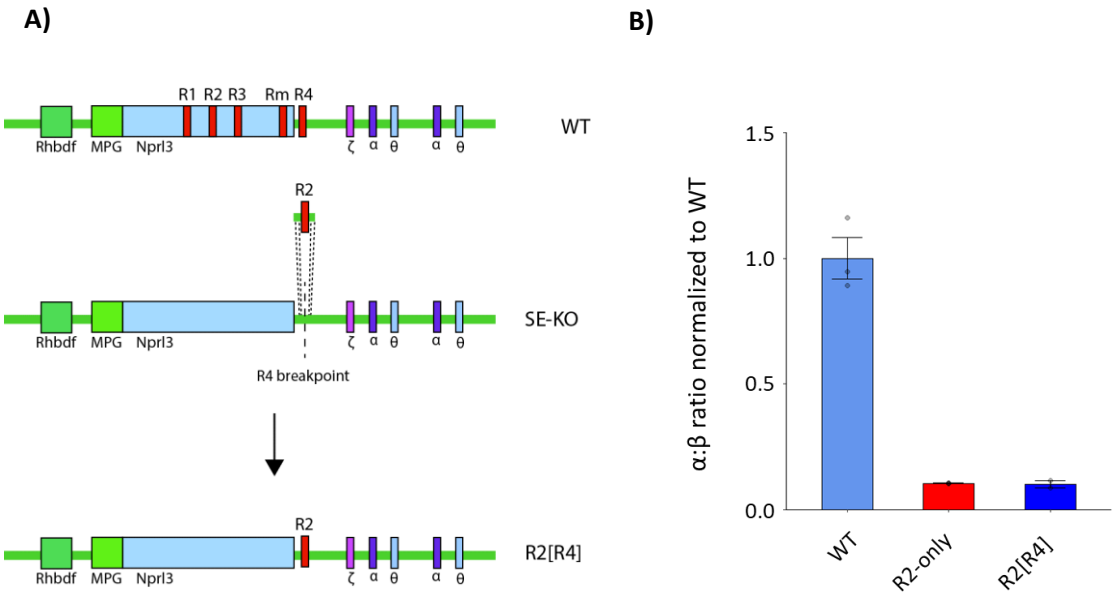


Figure 5.3 Reducing the distance between R2 and the α-globin promoters has no effect on gene expression.

(A) Schematic of generation of the R2[R4] model. Hemizygous SE-KO cells lipofected with pX458 expressing an sgRNA targeting the R4 deletion breakpoint and an HDR vector consisting of the R2 element flanked by 500bp homology arms complementary to sequences flanking the R4 deletion breakpoint. Clones FACS sorted based on GFP-fluorescence and genotyped by PCR and Sanger sequencing. (B) α-globin gene expression in WT (n=3), R2-only (n=3) and R2[R4] (n=2) EB-derived erythroid cells assayed by RT-qPCR. Expression normalized to β-globin and displayed as a proportion of WT expression. Dots = biological replicates; error bars = SE.

5.5 Discussion

Motif analysis and DNase footprinting suggest that R4, the most competent facilitator element in the WT α-globin superenhancer, recruits fewer TFs than R3, the least competent facilitator. To test whether R4's sequence or its position are most important for its relatively greater rescue potential, I made several models shifting the positions of various elements with respect to the α-globin promoters.

R4's lack of rescue potential when moved to the position of R1, in the R2R4[R1] model, indicates that simply having R4's sequence within the locus is not sufficient to boost α-globin expression. Meanwhile, moving R3, which usually has very little ability to rescue gene expression, to the position of R4 in the R1R2R3[R4] model led to a large increase in α-globin expression. Combined, these models strongly suggest that R4's position, rather than its sequence, underlies its greater rescue potential.

Although R4's position seems to be important for facilitating superenhancer activity, it remains unclear why this is. Recently, several groups have reported cohesin loading at active enhancer elements (Hua et al., 2021; Rinzema et al., 2021); it is tempting to speculate that facilitators such as R4 may also load cohesin and facilitate increased enhancer-promoter interactions through bi-directional loop extrusion. We also wondered whether R4 may act similarly to tethering elements, as described by the Levine group in *Drosophila*, potentially stabilizing interactions between the strong canonical enhancers (R1 and R2) and the α-globin promoters. To test whether R4's main role is to increase the frequency of interactions between R2 and the α-globin promoters (regardless of mechanism), I moved R2 to the position of R4 (close to the α-globin promoters), in the absence of all other regulatory elements, reasoning that the decreased linear distance between R2 and the promoters would lead to an increase in interaction frequency through diffusion. Unexpectedly, simply moving R2 closer to the α-globin promoters had no positive effect on α-globin gene expression. This was particularly surprising given recent work which has shown that an enhancer's ability to up-regulate its target gene is inversely correlated with the linear genomic distance between said enhancer and promoter (Rinzema et al., 2021; Zuin et al., 2022). However, it has been shown that the relationship between enhancer-promoter interaction frequency and gene activation is complex, with many unknown factors and processes determining how long an enhancer and promoter have to be in close proximity in order to stimulate gene expression (Zuin et al., 2022). The distance between R2 and the α-globin promoters is relatively short in WT cells (~30kb), and so the lack of increased activation potential when moved closer to the α-globin promoters (shortening the distance between R2 and the superenhancer-proximal α-globin promoter to ~14kb) could indicate that R2 is already acting at saturation in its native position.

R1-only and R2-only erythroid cells each express 10-11% α-globin compared to WT, meanwhile R1R2-only cells express 43%. This suggests a synergistic relationship between the two canonical enhancers. One explanation for this could be that combining R1' and R2's TFBSs satisfies some critical threshold, allowing recruitment and maintenance of significantly more TFs than summing those that R1 and R2 can

recruit individually. This would fit the R2[R4] phenotype. Perhaps, without the addition of more TFBSs, R2 can only drive 10% α-globin, regardless of its genomic position. It would be interesting to create a model in which both R1 *and* R2 are moved closer to the α-globin promoters to see whether, when combined, they could activate α-globin to a higher degree than when they are kept in their native positions.

More work is required to address the mechanism(s) via which facilitators boost the potential of activators. It would be interesting to conduct tiled-C and Med1 ChIPmentation in R1R2R3-only and R1R2R4-only erythroid cells. This could help address whether R4 increases coactivator recruitment and/or interaction frequency between R1 and R2, and the α-globin promoters. It may well be difficult to discriminate these mechanisms. It is feasible that R4 could facilitate more interactions between the activators and α-globin promoters, which in turn would increase the local density of TFBSs, potentially leading to greater coactivator recruitment. Alternatively, R4 could primarily boost mediator recruitment, which may then aid enhancer-promoter interactions through establishment of a transcriptional hub.

It would be interesting to test what the minimal requirements are to "build" a facilitator, and to investigate whether they have a specialised function or if they are merely clutches of TFBSs which don't independently satisfy the criteria to be considered enhancers. One could test whether a canonical enhancer can fill the role of a facilitator by inserting a second copy of R2 at the position of R4 in R1R2-only cells. This would help discriminate whether facilitators are *functionally distinct* regulatory elements. Similarly, it would be interesting to modify the R1R2-only model by recruiting a TF's DNA binding domain (separated from its transactivation domain) to the position of R4, to test whether opening chromatin at R4's position, without recruiting coactivators, is sufficient for facilitator activity.

Chapter six: Final conclusions:

6.1 Key findings:

The aim of this thesis was to investigate and disentangle the mode(s) of cooperation between the five putative enhancers comprising the α-globin superenhancer. To do this, I firstly tested the sufficiency of the strongest enhancer within the cluster, R2, to activate α-globin gene expression; this revealed that, independently of the other four constituents, R2 is unable to activate α-globin expression to the degree predicted. I then rebuilt the superenhancer in a combinatorial manner, uncovering synergistic, redundant and additive relationships between the five constituent elements. Rebuilding the cluster in multiple conformations also revealed that three of the five α-globin superenhancer constituents (R3, Rm and R4) have no canonical enhancer function, but instead serve to facilitate the activity of the two conventional enhancers (R1 and R2); hence, I named these elements facilitators. The three facilitators are inequivalent in their abilities to potentiate R1 and R2; therefore, I probed the relative importance of sequence versus position with respect to each facilitator's ability to boost superenhancer activity. This suggested that the hierarchy among the three facilitators is based more on their positions than their sequences. The key findings of my work are summarised in more depth, below:

1) **Characterizing the R2-only mouse model: does R2 remain an active enhancer?**

- R2's ability to activate α-globin gene expression is greatly attenuated in the absence of the other four α-globin superenhancer constituents.

- By itself, R2 retains the chromatin accessibility, histone modification profile and tissue-specific TF recruitment, expected of an active enhancer.

- R2 relies on R1, R3, Rm and R4 (or a subset thereof) to recruit high levels of transcriptional coactivators, to interact with the α-globin promoters effectively, and to transcribe hight levels of enhancer RNA.

2) An enhancer titration: How do the five α-globin superenhancer elements cooperate?

- The two strongest enhancers within the α-globin superenhancer, R1 and R2, are insufficient individually, and in combination, to activate high levels of α-globin gene expression.

- Even when combined, the R3, Rm and R4 elements show almost no canonical enhancer function or ability to activate α-globin gene expression.

- Reinserting R3, Rm or R4 into R1R2-only cells rescues α-globin gene expression to differing degrees.

- R3, Rm and R4 are facilitators; they have minimal enhancer activity, but facilitate the activities of the R1 and R2 enhancers.

3) Enhancer rescue: position or sequence?

- R3, in its native position, is the weakest facilitator, but contains more tissue-specific TF motifs than Rm and R4 combined.

- R4's ability to rescue α-globin gene expression is position-dependent. In its native promoter-proximal position, R4 acts as the strongest facilitator; however, transplanting R4 to the position of R1, the element furthest from the α-globin promoters, abolishes its ability to rescue α-globin gene expression.

- R3's potential to rescue α-globin gene expression is position-dependent; transplanting R3 to the position of R4, the element closest to the α-globin promoters, boosts its ability to rescue α-globin gene expression ~5-fold.

- Facilitators do not appear to act by simply increasing enhancer-promoter proximity; transplanting the R2 enhancer to the position of R4, the element closest to the α-globin promoters, is insufficient to restore gene expression to the degree predicted.

6.2 Contribution to the field of gene regulation:

Enhancer clusters have been studied for nearly four decades (Mercola et al., 1983). A wide variety of enhancer cluster sub-categories have been described over the years (Grosveld et al., 1987; Hong et al., 2008; Markenscoff-Papadimitriou et al., 2014; Montavon et al., 2011; Parker et al., 2013; Whyte et al., 2013); however, today the most commonly studied class is the superenhancer. Superenhancers are distinguished from regular enhancer clusters bioinformatically. Fundamentally, they are called based on the short distances separating their individual constituents, high levels of H3K27Ac, and unusually high recruitment of mediator complex components (Whyte et al., 2013). The fact that these distinctions are descriptive rather than functional has raised questions around the legitimacy and utility of defining superenhancers as a distinct species of regulatory element (Blobel et al., 2021; Moorthy et al., 2017). One key question is whether superenhancers manifest emergent properties beyond clusters of regular independently-functioning enhancers. A number of studies have suggested that superenhancer constituents cooperate synergistically/non-additively with one another (Brosh et al., 2022; Choi et al., 2021; Hnisz et al., 2015; Hörnblad et al., 2021; Huang et al., 2018; Shin et al., 2016; Thomas et al., 2021); others have indicated that they combine additively (Bender et al., 2012; Choi et al., 2021; Hay et al., 2016). It is quite possible that superenhancer cooperation varies on a case-by-case basis, as is true for TF cooperation within regular enhancers, evoking for example the proposed "enhanceosome" and "billboard" models (Kulkarni & Arnosti, 2003; Maniatis et al., 1998; Panne et al., 2007; Spitz & Furlong, 2012).

Recent studies have provided evidence that superenhancers can be composed of multiple functionally distinct element types, for example, hub and non-hub enhancers (Huang et al., 2018), amplifiers (Thomas et al., 2021), and chromatin-dependent and -independent enhancers (Sahu et al., 2022). Although these studies support our observations, each provides a limited and incomplete view. For example, in Huang et al (2018), the majority of conclusions are based on bioinformatic descriptions and CRISPR interference (dead Cas9 fused to the KRAB transcriptional repressor) targeted to individual superenhancer constituents (Huang et al., 2018). Similarly, the superenhancer studied in Thomas et al (2021), the Fgf5 superenhancer, activates a key regulator of cell identity and development; it is therefore difficult to distinguish direct and indirect effects of altering Fgf5 gene expression. Furthermore, the "amplifier" identified within the Fgf5 superenhancer is located directly upstream of the Fgf5 gene and it recruits very high levels of Pol2; it is therefore hard to discount the possibility that this element is actually an alternative promoter rather than a superenhancer constituent (Thomas et al., 2021). Lastly, in Sahu et al (2022), the classification of chromatin-dependent and -independent enhancers is based solely on reporter (STARR-seq) assays, divorced from their biologically relevant chromatin contexts (Sahu et al., 2022). Nonetheless, these observations suggest that superenhancers can be composed of elements with distinct and complementary functions. My work has provided clearer evidence and has shown, through careful dissection of the native α-globin superenhancer, that this cluster is composed of two functionally distinct element classes: conventional enhancer elements capable of activating target gene expression, and facilitators which have little/no inherent enhancer activity, but boost target gene expression by potentiating their associated enhancers. This clearly demonstrates that superenhancers *can* manifest emergent properties beyond clusters of regular independently-acting enhancers.

Our original dissection of the α-globin superenhancer (Hay et al., 2016), like most other superenhancer dissection studies, identified a mixture of active and inactive elements (Bender et al., 2012; Brosh et al., 2022; Choi et al., 2021; Hay et al., 2016; Hörnblad et al., 2021; Huang et al., 2018; Sahu et al., 2022). It is worth noting that dissecting the α-globin superenhancer by removing each element individually or in

selective pairs was insufficient to resolve the true function of the three facilitator elements (Hay et al., 2016). This is due to the redundancy displayed between the three elements, combined with the fact that the facilitators *only* have activity in the presence of the active enhancers, R1 and R2. Therefore, deleting R1 and R2 simultaneously almost completely abolished α-globin gene expression, suggesting that R3, Rm and R4 are inactive. This should serve as a cautionary tale. Other superenhancer dissection studies have limited their deletions to individual and pairwise elements, an approach that fails to unmask the activity of facilitator elements. What's more, facilitators will not score in conventional enhancer reporter assays, as they do not directly activate gene expression. Combined, this means that facilitator elements have likely been described as active enhancers if their deletion *in situ* caused disruption in expression, or disregarded as "inactive" elements if no effect was observed upon their deletion, when in actual fact they play a non-classical role in regulating optimal levels of target gene expression.

In the absence of facilitator elements, the strongest enhancer of the α-globin superenhancer, R2, is incapable of driving high levels of gene expression. Interestingly, R2 does not appear to require any additional elements in order to become accessible or to recruit high levels of tissue-specific TFs; however, R2 does need additional elements in order to recruit high levels of transcriptional coactivators such as the mediator complex and Brd4. R2 also requires additional constituents in order to recruit components of the preinitiation complex, to interact with its target gene promoters with high frequency, and to transcribe high levels of bi-directional enhancer RNAs. The R2-only model lacks the R1 enhancer as well as the three facilitator elements; it is therefore unclear which aspects of the R2-only phenotype, if any, are specifically attributable to the absence of facilitators. Regardless, removing the three facilitators from the α-globin locus (in the R1R2-only model) causes α-globin expression to drop by >50%, underscoring the significance of these elements in regulating gene expression.

The three α-globin facilitator elements are inequivalent in their abilities to potentiate enhancer activity. R4 is the element closest to the α-globin promoters and is the strongest facilitator; R3 is located furthest from the promoters and is the weakest

facilitator. Contrary to expectation, R3 – the weakest facilitator – contains more tissue-specific TF binding motifs than Rm and R4 combined. Many studies have shown a positive correlation between number and diversity of TF binding motifs and enhancer strength (Sahu et al., 2022; Singh et al., 2021; Spitz & Furlong, 2012; Zinzen et al., 2009); however, it is possible that this relationship does not apply as strictly to facilitators. R4's ability to rescue gene expression was abolished when R4 was transplanted to the position of R1, whereas inserting R3 into the position of R4 increased its ability to rescue α-globin expression ~5-fold. These observations both suggest that the hierarchy among the facilitators is more dependent on each element's positioning rather than its sequence. This observation is particularly interesting in light of other recent work exploring the α-globin superenhancer. Kassouf et al (2022) have shown that the α-globin superenhancer (as well as a large number of other multipartite erythroid superenhancers) activates its target genes in a directional manner. Inverting the orientation of the superenhancer re-directs its activation potential towards upstream genes and reduces α-globin gene expression (Kassouf et al., 2022). In this super-enhancer inversion model, the position of R1 and R2 is almost unchanged, whereas the three facilitators are re-oriented towards the upstream genes, suggesting that their re-positioning is likely the cause of the associated changes in gene expression. The α-globin superenhancer inversion study challenges one of the key tenets of enhancer biology, orientation-independent function, and reinforces the fact that there remains much to investigate and understand about superenhancers. Furthermore, similar to the work described in this thesis, it demonstrates that superenhancers do manifest emergent properties. It would be interesting to investigate, in more depth, whether the facilitator elements play a particular role in defining the direction in which the α-globin superenhancer predominantly acts. One way of testing this would be to edit R1R2-only cells by reinserting the R4 facilitator upstream of R1 – far from the α-globin genes, but close to the genes upstream of the superenhancer. This would allow us to test whether re-positioning a facilitator redirects the superenhancer's activity and phenocopies a complete superenhancer inversion.

Facilitators as structural modulators of enhancer-promoter interactions:

Because a facilitator's position appears to be important in determining its strength, I hypothesised that facilitators might play a role in positioning active enhancers. For example, R4 might play a role in increasing interaction frequency between R2 and the α-globin promoters. Many studies have shown that enhancer-promoter interaction frequency and/or proximity are important for target gene activation, although the mechanisms linking proximity and activation remain unclear (Benabdallah et al., 2019; Oudelaar & Higgs, 2021b; Zuin et al., 2022). This hypothesis would be consistent with the reduction in interactions between R2 and the promoters in R2-only cells and the associated reduction in α-globin expression. I speculated that if the R2-only phenotype was simply a result of reduced interactions between R2 and the α-globin promoters, then artificially increasing interaction frequency should be sufficient to restore higher levels of gene expression. Reducing the linear distance between R2 and the α-globin promoters would be expected to increase their interaction frequency through diffusion (and potentially other mechanisms too). Surprisingly, moving R2 to the position of R4 had no positive effect on gene expression when compared to cells in which R2 is located in its native position. This suggests that facilitators do not solely act to increase enhancer-promoter interaction frequency. One caveat to this conclusion is the fact that I did not explicitly show that moving R2 to the position of R4 had any effect on interaction frequency between R2 and the α-globin promoters, but simply assumed a relationship between linear and 3D proximity. Recently, other studies have reported a relationship between enhancer-promoter proximity and target gene activation (Rinzema et al., 2021; Zuin et al., 2022). The finding that moving R2 closer to the α-globin promoters has no effect on α-globin expression could be due to the fact that R2 is only moved 20kb, compared to the larger distances tested in other studies (up to 400kb). It is also possible that R2 is already working at saturation in its native position and has no capacity to increase gene expression when moved closer to the α-globin promoters without the addition of other elements/more TF binding motifs.

Facilitators as TF and cofactor recruiters:

Facilitators may primarily act through recruiting additional TFs at particular positions along the linear genome. Several publications have suggested that superenhancers up-regulate their target genes by recruiting and retaining particularly high concentrations of TFs and transcriptional apparatus, such as mediator and Pol2. This process could involve phase separation, it could involve TF trapping, or it could simply be a function of TF motif density and affinity. Regardless, TF concentration will be influenced by TF motif density, which is in turn dependent on two things: number of motifs and distance between motifs in 3D space. If R2 and R4 exist in close 3D proximity more frequently than R2 and R3, as is suggested by micro-Capture-C (a 3C-based technique which allows interrogation of chromatin interactions at extremely high resolution) (Hua et al., 2021), then motif density may be greater in an α-globin superenhancer composed of R1, R2 and R4 than one composed of R1, R2 and R3, despite the fact the latter contains a greater absolute number of TF motifs. This could explain both the importance of R4's genomic position, and the superenhancer's relative agnosticism to facilitator element sequence. It would be very interesting to test "synthetic facilitators" composed of varying numbers of TF motifs in the position of R4 to investigate this hypothesis further.

An alternative interpretation would be that R3, Rm and R4 are simply weak enhancers which have little enhancer activity in the absence of R1 and R2. The fact that reinserting R4 into R1R2-only cells causes little change in H3K27Ac over R4, but a large increase over R1 and R2, strongly suggests that R4's predominant role is in facilitating and augmenting the activity of R1 and R2, rather than acting as an enhancer itself.

Facilitators could play multiple simultaneous roles e.g. in enhancer positioning, TF recruitment, and/or enhancer RNA transcription. Alternatively, they may influence currently unknown processes. More work is required to determine the precise nature of facilitator elements and to disentangle their functions within superenhancers.

Facilitators beyond the α-globin superenhancer:

This work has focused on the α-globin superenhancer and it remains unclear whether facilitators exist in other superenhancers too. Anecdotal evidence from other studies

suggests that facilitators may be present at other loci. For example, dissection of both the Pri-miR-290-295 and β-globin superenhancers has identified elements resembling facilitators. E8 from the Pri-miR-290-295 cluster and HS1 from the β-globin cluster both lack enhancer activity in traditional reporter assays; however, deleting each of these elements from their native loci causes a modest-large reduction in gene expression (Hardison et al., 1997; Hnisz et al., 2015; Schübeler et al., 2001). Additionally, both of these elements contain fewer tissue-specific TF binding motifs than other constituents within their superenhancers, and both elements are the most promoter-proximal constituents within their respective clusters. Collectively, these characteristics bear resemblance to what we have seen at R4 within the α-globin superenhancer. Clearly these are just two examples and more rigorous work is required to investigate facilitators genome-wide.

Interestingly, elements reminiscent of facilitators have been described in Drosophila; tethering elements, which have no inherent activation potential, are thought to be one of the two distinct classes of regulatory element (alongside insulators) that regulate chromatin structure, which in turn influences gene regulation. These tethering elements are thought to mediate enhancer-promoter interactions from a promoter-proximal position, in order to ensure fast activation kinetics of target genes (Batut et al., 2022; Levo et al., 2022). The authors anticipate that such elements and associated mechanisms will prove important in vertebrate genomes, where large distances often separate genes from their regulatory sequences. Perhaps facilitators are the mammalian equivalent.

6.3 Next steps

There are two clear directions in which this work could be extended: 1) to investigate the mechanisms underlying facilitator function; 2) to search for a method of identifying facilitators genome-wide.

Mechanism(s) of action:

We have proposed two potential mechanisms via which facilitators may function; one is structural and is potentially implicated in directing the canonical enhancers towards their target promoter. The second is in recruiting more TFs and coactivators, creating a more permissive environment for transcription. In the R2-only model, we saw phenotypes indicative of both mechanisms: reductions in enhancer-promoter interaction frequency, and reductions in coactivator recruitment. It is likely that these phenotypes are linked with one another. It is therefore possible that reinsertion of a facilitator could rescue all phenotypes simultaneously, even if facilitators only directly affect one process. Nonetheless, it would be very interesting to probe each of these phenomena in R1R2R3-only and R1R2R4-only cells, to investigate whether the addition of a facilitator element rescues one or more of these phenotypes. It would be particularly interesting if one of these phenomena were rescued by R4 reinsertion and not R3 reinsertion, mirroring the effect reinsertion of each element has on α-globin gene expression and bringing us closer to one or the other mode of action.

Identifying facilitators genome-wide:

There are two approaches I would use to search for facilitators elsewhere in the genome. The first would focus on analysing existing publicly available data. Sahu et al (2022) recently used a STARR-seq protocol in HepG2 cells to assay enhancer activity in elements bearing the marks of active enhancers (chromatin accessibility, enhancer-associated histone modifications and tissue-specific TF recruitment), genome-wide (Sahu et al., 2022). They identified “chromatin-dependent” enhancers which display no functional enhancer activity and weaker enrichment for TF binding motifs when compared to active (“chromatin independent”) enhancers. It would be interesting to isolate all *superenhancer* constituent elements from this dataset and to subset these elements into those with enhancer activity and those devoid of enhancer activity. One could then download all histone modification and TF ChIP-seq files available in this tissue type to search for signatures which are preferentially or exclusively enriched over facilitators compared to active enhancers. If a facilitator-associated signature was identified, then this could be used to classify facilitators genome-wide, potentially in a variety of tissue types.

The second approach would be experimental. I would revisit the literature and identify all superenhancer constituents categorised as inactive in previous dissection studies and reporter assays, for example four of the five Myc superenhancer constituents (Sahu et al., 2022) and three of the six β-globin superenhancer elements (Bender et al., 2012). I would then use the position of R4, within the α-globin superenhancer, as a functional screening assay. I would insert each inactive element into R1R2-only cells at the position of R4, and assay the effects on α-globin expression. If gene expression increased, we could conclude that the element acts as a facilitator; this would obviate the need for exhaustive dissection of each superenhancer *in situ* as we have done at the α-globin locus. This strategy would also allow us to functionally test candidate facilitator sequences identified from a genome-wide bioinformatic or reporter-assay approach, as well as synthetic sequences designed with specific hypotheses in mind.

6.4 Closing remarks

The work presented in this thesis is the most thorough *in situ* dissection of a superenhancer to date. Without rebuilding the α-globin superenhancer in multiple configurations, we would not have been able to identify facilitators. This showcases the continued importance of careful genetic dissection at individual loci. It is likely that facilitator elements are not features of all superenhancers and it may also transpire that these are not elements discernible by a distinct signature. They may well be position- and context- dependent, serving to satisfy a critical threshold of TF recruitment or to provide the chromatin configuration necessary for effective enhancer-promoter interaction. In any case, my work shows that superenhancers can manifest emergent properties, and opens a new direction for probing enhancer clusters. It is my hope that this work will serve not only to enhance our understanding of α-globin gene expression, but to further our overall understanding of gene regulation.

References

Allahyar, A., Vermeulen, C., Bouwman, B. A. M., Krijger, P. H. L., Verstegen, M. J. A. M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J. H. I., Jalink, K., Teunissen, H., Renkens, I. J., Kloosterman, W. P., Rowland, B. D., de Wit, E., de Ridder, J., & de Laat, W. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. *Nature Genetics*, *50*(8), 1151–1160. https://doi.org/10.1038/S41588-018-0161-5

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455. https://doi.org/10.1038/NATURE12787

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/SUPPL_FILE/ARNOLD.SM.PDF

Arnold, P. R., Wells, A. D., & Li, X. C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in Cell and Developmental Biology*, *7*, 377. https://doi.org/10.3389/FCELL.2019.00377/BIBTEX

Arnosti, D. N., & Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry*, *94*(5), 890–898. https://doi.org/10.1002/JCB.20352

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). Retrieved December 10, 2021, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. *Nucleic Acids Research*, *43*(W1), W39–W49. https://doi.org/10.1093/NAR/GKV416

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

Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J., & Hadjur, S. (2019). Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nature Communications 2019 10:1*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-10725-9

Batut, P. J., Bing, X. Y., Sisco, Z., Raimundo, J., Levo, M., & Levine, M. S. (2022). Genome organization controls transcriptional dynamics during development. *Science (New York, N.Y.)*, *375*(6580). https://doi.org/10.1126/SCIENCE.ABI7178

Beagan, J. A., & Phillips-Cremins, J. E. (2020). On the existence and functionality of topologically associating domains. *Nature Genetics 2020 52:1*, *52*(1), 8–16. https://doi.org/10.1038/s41588-019-0561-1

Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., Barbieri, M., de Santiago, I., Lavitas, L. M., Branco, M. R., Fraser, J., Dostie, J., Game, L., Dillon,

N., Edwards, P. A. W., Nicodemi, M., & Pombo, A. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature 2017 543:7646*, *543*(7646), 519–524. https://doi.org/10.1038/nature21411

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

Bender, M. A., Ragoczy, T., Lee, J., Byron, R., Telling, A., Dean, A., & Groudine, M. (2012). The hypersensitive sites of the murine β-globin locus control region act independently to affect nuclear localization and transcriptional elongation. *Blood*, *119*(16). https://doi.org/10.1182/blood-2011-09-380485

Bernet, A., Sabatier, S., Picketts, D., Ouazana, R., Morle, F., Higgs, D., & Godet, J. (1995). Targeted Inactivation of the Major Positive Regulatory Element (HS-40) of the Human α-Globin Gene Locus. *Blood*, *86*(3), 1202–1211. https://doi.org/10.1182/BLOOD.V86.3.1202.1202

Blayney, J., Francis, H., Camellato, B., Mitchell, L., Stolper, R., Boeke, J., Higgs, D., & Kassouf, M. (2022). Super-enhancers require a combination of classical enhancers and novel facilitator elements to drive high levels of gene expression. *BioRxiv*, 2022.06.20.496856. https://doi.org/10.1101/2022.06.20.496856

Blobel, G. A., Higgs, D. R., Mitchell, J. A., Notani, D., & Young, R. A. (2021). Testing the super-enhancer concept. *Nature Reviews. Genetics*, *22*(12), 749–755. https://doi.org/10.1038/S41576-021-00398-W

Boija, A., Klein, I. A., Sabari, B. R., Dall'Agnese, A., Coffey, E. L., Zamudio, A. v., Li, C. H., Shrinivas, K., Manteiga, J. C., Hannett, N. M., Abraham, B. J., Afeyan, L. K., Guo, Y. E., Rimel, J. K., Fant, C. B., Schuijers, J., Lee, T. I., Taatjes, D. J., & 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

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

Brosh, R., Coelho, C., Ribeiro-dos-Santos, A. M., Hogan, M. S., Ashe, H. J., Ellis, G., Somogyi, N., Ordoñez, R., Luther, R. D., Huang, E., Boeke, J. D., & Maurano, M. T. (2022). Dissection of a complex enhancer cluster at the Sox2 locus. *BioRxiv*, 2022.06.20.495832. https://doi.org/10.1101/2022.06.20.495832

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *109*, 21.29.1. https://doi.org/10.1002/0471142727.MB2129S109

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

Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjana, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G. C., Zhang, F., Orkin, S. H., & Bauer, D. E. (2015). BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*, *527*(7577). https://doi.org/10.1038/nature15521

Catarino, R. R., & Stark, A. (2018). *Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation*. https://doi.org/10.1101/gad.310367

Chen, H., & Liang, H. (2020). A High-Resolution Map of Human Enhancer RNA Loci Characterizes Super-enhancer Activities in Cancer. *Cancer Cell*, *38*(5), 701-715.e5. https://doi.org/10.1016/J.CCELL.2020.08.020

Chen, Y., McCarthy, D., Ritchie, M., Robinson, M., Smyth, G., & Hall, E. (n.d.). *edgeR: differential analysis of sequence read count data User's Guide*.

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

Coelho, A., Picanço, I., Seuanes, F., Seixas, M. T., & Faustino, P. (2010). Novel large deletions in the human alpha-globin gene cluster: Clarifying the HS-40 long-range regulatory role in the native chromosome environment. *Blood Cells, Molecules & Diseases*, *45*(2), 147–153. https://doi.org/10.1016/J.BCMD.2010.05.010

Cormack, B. P., & Struhl, K. (1992). The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell*, *69*(4), 685–696. https://doi.org/10.1016/0092-8674(92)90232-2

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/S41586-019-1517-4

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & 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/SUPPL_FILE/ST01.XLSX

Crump, N. T., Ballabio, E., Godfrey, L., Thorne, R., Repapi, E., Kerry, J., Tapia, M., Hua, P., Lagerholm, C., Filippakopoulos, P., Davies, J. O. J., & Milne, T. A. (2021). BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-020-20400-z

Dębek, S., & Juszczyński, P. (2022). Super enhancers as master gene regulators in the pathogenesis of hematologic malignancies. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, *1877*(2), 188697. https://doi.org/10.1016/J.BBCAN.2022.188697

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

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, *29*(1), 15–21. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Doré, L. C., & Crispino, J. D. (2011). Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*, *118*(2), 231–239. https://doi.org/10.1182/BLOOD-2011-04-285981

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature 2012 489:7414*, *489*(7414), 57–74. https://doi.org/10.1038/nature11247

Ernst, J., & Kellis, M. (2017). Chromatin state discovery and genome annotation with ChromHMM. *Nature Protocols*, *12*(12), 2478. https://doi.org/10.1038/NPROT.2017.124

Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S., & Levine, M. S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6508–6513. https://doi.org/10.1073/PNAS.1605085113/SUPPL_FILE/PNAS.1605085113.SD01.XLSX

Fitz, J., Neumann, T., & Pavri, R. (2018). Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *The EMBO Journal*, *37*(8). https://doi.org/10.15252/EMBJ.201797965

Flemming, W. (1875). Studien in der Entwicklungsgeschichte der Najaden. *SBien*, 132. https://eurekamag.com/research/023/678/023678887.php

Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W., & Mathelier, A. (2020). JASPAR 2020: Update of the open-Access database of transcription factor binding profiles. *Nucleic Acids Research*, *48*(D1). https://doi.org/10.1093/nar/gkz1001

Francis, H. S., Harold, C. L., Beagrie, R. A., King, A. J., Gosden, M. E., Blayney, J. W., Jeziorska, D. M., Babbs, C., Higgs, D. R., & Kassouf, M. T. (2022). Scalable in vitro production of defined mouse erythroblasts. *PLOS ONE*, *17*(1), e0261950. https://doi.org/10.1371/JOURNAL.PONE.0261950

Geyer, P. K., Green, M. M., & Corces, V. G. (1990). Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in Drosophila. *The EMBO Journal*, *9*(7), 2247–2256. https://doi.org/10.1002/J.1460-2075.1990.TB07395.X

Gorbovytska, V., Kim, S. K., Kuybu, F., Götze, M., Um, D., Kang, K., Pittroff, A., Brennecke, T., Schneider, L. M., Leitner, A., Kim, T. K., & Kuhn, C. D. (2022). Enhancer RNAs stimulate

Pol II pause release by harnessing multivalent interactions to NELF. *Nature Communications 2022 13:1*, *13*(1), 1–22. https://doi.org/10.1038/s41467-022-29934-w

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

Grosveld, F., van Staalduinen, J., & Stadhouders, R. (2021). Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms. *Annual Review of Genomics and Human Genetics*, *22*, 127–146. https://doi.org/10.1146/ANNUREV-GENOM-122220-093818

Gurumurthy, A., Shen, Y., Gunn, E. M., & Bungert, J. (2019). Phase Separation and Transcription Regulation: Are Super-Enhancers and Locus Control Regions Primary Sites of Transcription Complex Assembly? *BioEssays*, *41*(1), 1800164. https://doi.org/10.1002/BIES.201800164

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

Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Biggs, D., Preece, C., Downes, D. J., Gosden, M., Sharpe, J. A., Sloane-Stanley, J. A., Hughes, J. R., Davies, B., & Higgs, D. R. (2017b). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology*, *19*(8), 952–961. https://doi.org/10.1038/NCB3573

Hardison, R., Slightom, J. L., Gumucio, D. L., Goodman, M., Stojanovic, N., & Miller, W. (1997). Locus control regions of mammalian β-globin gene clusters: combining phylogenetic analyses and experimental results to gain functional insights. *Gene*, *205*(1–2), 73–94. https://doi.org/10.1016/S0378-1119(97)00474-5

Harteveld, C. L., & Higgs, D. R. (2010). Alpha-thalassaemia. *Orphanet Journal of Rare Diseases*, *5*(1). https://doi.org/10.1186/1750-1172-5-13

Hay, D., Hughes, J. R., Babbs, C., Davies, J. O. J., Graham, B. J., Hanssen, L. L. P., Kassouf, M. T., Oudelaar, A. M., Sharpe, J. A., Suciu, M. C., Telenius, J., Williams, R., Rode, C., Li, P. S., Pennacchio, L. A., Sloane-Stanley, J. A., Ayyub, H., Butler, S., Sauka-Spengler, T., … Higgs, D. R. (2016). Genetic dissection of the α-globin super-enhancer in vivo. *Nature Genetics*, *48*(8). https://doi.org/10.1038/ng.3605

Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. v., Stewart, R., Thomson, J. A., Crawford, G. E., … Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature 2009 459:7243*, *459*(7243), 108–112. https://doi.org/10.1038/nature07829

Higgs, D. R., Engel, J. D., & Stamatoyannopoulos, G. (2012). Thalassaemia. *Lancet (London, England)*, *379*(9813), 373–383. https://doi.org/10.1016/S0140-6736(11)60283-3

Higgs, D. R., & Wood, W. G. (2008). Long-range regulation of alpha globin gene expression during erythropoiesis. *Current Opinion in Hematology*, *15*(3), 176–183. https://doi.org/10.1097/MOH.0B013E3282F734C4

Hnisz, D., Schuijers, J., Lin, C. Y., Weintraub, A. S., Abraham, B. J., Lee, T. I., Bradner, J. E., & Young, R. A. (2015). Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers. *Molecular Cell*, *58*(2). 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 predicts key features of transcriptional control. *Cell*, *169*(1), 13. https://doi.org/10.1016/J.CELL.2017.02.007

Hoffman, M. M., Buske, O. J., Wang, J., Weng, Z., Bilmes, J. A., & Noble, W. S. (2012). Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods*, *9*(5), 473. https://doi.org/10.1038/NMETH.1937

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

Hörnblad, A., Bastide, S., Langenfeld, K., Langa, F., & Spitz, F. (2021). Dissection of the Fgf8 regulatory landscape by in vivo CRISPR-editing reveals extensive intra- and inter-enhancer redundancy. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20714-y

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

Hua, P., Badat, M., Hanssen, L. L. P., Hentges, L. D., Crump, N., Downes, D. J., Jeziorska, D. M., Oudelaar, A. M., Schwessinger, R., Taylor, S., Milne, T. A., Hughes, J. R., Higgs, D. R., & Davies, J. O. J. (2021). Defining genome architecture at base-pair resolution. *Nature*, *595*(7865). https://doi.org/10.1038/s41586-021-03639-4

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

Hubrecht, R. C., & Carter, E. (2019). The 3Rs and Humane Experimental Technique: Implementing Change. *Animals : An Open Access Journal from MDPI*, *9*(10). https://doi.org/10.3390/ANI9100754

Hughes, J. R., Cheng, J. F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., de Gobbi, M., de Jong, P., Rubin, E., & Higgs, D. R. (2005a). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(28). https://doi.org/10.1073/pnas.0503401102

Hughes, J. R., Cheng, J. F., Ventress, N., Prabhakar, S., Clark, K., Anguita, E., de Gobbi, M., de Jong, P., Rubin, E., & Higgs, D. R. (2005b). Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved

sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(28), 9830–9835. https://doi.org/10.1073/PNAS.0503401102/SUPPL_FILE/03401FIG7.PDF

Hughes, J. R., Roberts, N., Mcgowan, S., Hay, D., Giannoulatou, E., Lynch, M., de Gobbi, M., Taylor, S., Gibbons, R., & Higgs, D. R. (2014). Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics*, *46*(2). https://doi.org/10.1038/ng.2871

Ibrahim, D. M., & Mundlos, S. (2020). Three-dimensional chromatin in disease: What holds us together and what drives us apart? *Current Opinion in Cell Biology*, *64*, 1–9. https://doi.org/10.1016/J.CEB.2020.01.003

Ing-Simmons, E., Seitan, V. C., Faure, A. J., Flicek, P., Carroll, T., Dekker, J., Fisher, A. G., Lenhard, B., & Merkenschlager, M. (2015). Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Research*, *25*(4), 504–513. https://doi.org/10.1101/GR.184986.114

Inoue, F., & Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics*, *106*(3), 159–164. https://doi.org/10.1016/J.YGENO.2015.06.005

Kassouf, M. T., Francis, H. S., Gosden, M., Suciu, M. C., Downes, D. J., Harrold, C., Larke, M., Oudelaar, M., Cornell, L., Blayney, J., Telenius, J., Xella, B., Shen, Y., Sousos, N., Sharpe, J. A., Sloane-Stanley, J., Smith, A., Babbs, C., Hughes, J. R., & Higgs, D. R. (2022). Multipartite super-enhancers function in an orientation-dependent manner. *BioRxiv*, 2022.07.14.499999. https://doi.org/10.1101/2022.07.14.499999

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. https://doi.org/10.1101/GR.229102

King, A. J., Songdej, D., Downes, D. J., Beagrie, R. A., Liu, S., Buckley, M., Hua, P., Suciu, M. C., Marieke Oudelaar, A., Hanssen, L. L. P., Jeziorska, D., Roberts, N., Carpenter, S. J., Francis, H., Telenius, J., Olijnik, A. A., Sharpe, J. A., Sloane-Stanley, J., Eglinton, J., … Babbs, C. (2021). Reactivation of a developmentally silenced embryonic globin gene. *Nature Communications 2021 12:1*, *12*(1), 1–15. https://doi.org/10.1038/s41467-021-24402-3

Koulnis, M., Pop, R., Porpiglia, E., Shearstone, J. R., Hidalgo, D., & Socolovsky, M. (2011). Identification and Analysis of Mouse Erythroid Progenitors using the CD71/TER119 Flow-cytometric Assay. *Journal of Visualized Experiments : JoVE*, *54*, 2809. https://doi.org/10.3791/2809

Kredel, S., Oswald, F., Nienhaus, K., Deuschle, K., Röcker, C., Wolff, M., Heilker, R., Nienhaus, G. U., & Wiedenmann, J. (2009). mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS ONE*, *4*(2). https://doi.org/10.1371/journal.pone.0004391

Krietenstein, N., Abraham, S., Venev, S. v., Abdennur, N., Gibcus, J., Hsieh, T. H. S., Parsi, K. M., Yang, L., Maehr, R., Mirny, L. A., Dekker, J., & Rando, O. J. (2020). Ultrastructural details of mammalian chromosome architecture. *Molecular Cell*, *78*(3), 554. https://doi.org/10.1016/J.MOLCEL.2020.03.003

Krivega, I., & Dean, A. (2017). LDB1-mediated enhancer looping can be established independent of mediator and cohesin. *Nucleic Acids Research*, *45*(14), 8255. https://doi.org/10.1093/NAR/GKX433

Kulkarni, M. M., & Arnosti, D. N. (2003). Information display by transcriptional enhancers. *Development*, *130*(26), 6569–6575. https://doi.org/10.1242/DEV.00890

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods 2012 9:4*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Larke, M. S. C., Schwessinger, R., Nojima, T., Proudfoot, N. J., Higgs, D. R., & Hughes Correspondence, J. R. (2021). Enhancers predominantly regulate gene expression during differentiation via transcription initiation. *Molecular Cell*, *81*, 983–997. https://doi.org/10.1016/j.molcel.2021.01.002

Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & 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

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/BIOINFORMATICS/BTP352

Li, X., & Noll, M. (1994). Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. *The EMBO Journal*, *13*(2), 400–406. https://doi.org/10.1002/J.1460-2075.1994.TB06274.X

Li, Y., Hu, M., & Shen, Y. (2018). Gene regulation in the 3D genome. *Human Molecular Genetics*, *27*(R2), R228–R233. https://doi.org/10.1093/HMG/DDY164

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*(8). https://doi.org/10.1093/nar/gkz114

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell*, *126*(2), 403–413. https://doi.org/10.1016/J.CELL.2006.06.035

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

Long, H. K., Prescott, S. L., & Wysocka, J. (2016b). Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, *167*(5), 1170. https://doi.org/10.1016/J.CELL.2016.09.018

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Maekawa, T., Imamoto, F., Merlino, G. T., Pastan, I., & Ishii, S. (1989). Cooperative Function of Two Separate Enhancers of the Human Epidermal Growth Factor Receptor Proto-oncogene. *Journal of Biological Chemistry*, *264*(10), 5488–5494. https://doi.org/10.1016/S0021-9258(18)83571-2

Magoč, T., Magoč, M., & Salzberg, S. L. (2011). *FLASH: fast length adjustment of short reads to improve genome assemblies*. *27*(21), 2957–2963. https://doi.org/10.1093/bioinformatics/btr507

Maniatis, T., Falvo, J. v., Kim, T. H., Kim, T. K., Lin, C. H., Parekh, B. S., & Wathelet, M. G. (1998). Structure and Function of the Interferon-β Enhanceosome. *Cold Spring Harbor Symposia on Quantitative Biology*, *63*, 609–620. https://doi.org/10.1101/SQB.1998.63.609

Markenscoff-Papadimitriou, E., Allen, W. E., Colquitt, B. M., Goh, T., Murphy, K. K., Monahan, K., Mosley, C. P., Ahituv, N., & Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. *Cell*, *159*(3), 543. https://doi.org/10.1016/J.CELL.2014.09.033

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

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

Mishal, R., & Luna-Arias, J. P. (2022). Role of the TATA-box binding protein (TBP) and associated family members in transcription regulation. *Gene*, *833*. https://doi.org/10.1016/J.GENE.2022.146581

Mitchell, L. A., McCulloch, L. H., Pinglay, S., Berger, H., Bosco, N., Brosh, R., Bulajić, M., Huang, E., Hogan, M. S., Martin, J. A., Mazzoni, E. O., Davoli, T., Maurano, M. T., & Boeke, J. D. (2021). De novo assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics*, *218*(1). https://doi.org/10.1093/GENETICS/IYAB038

Monfils, K., & Barakat, T. S. (2021). Models behind the mystery of establishing enhancer-promoter interactions. *European Journal of Cell Biology*, *100*(5–6), 151170. https://doi.org/10.1016/J.EJCB.2021.151170

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., de Laat, W., Spitz, F., & 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., Martchenko, A., So, V., Macpherson, N. N., & 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). https://doi.org/10.1101/gr.210930.116

Nichols, J., Evans, E. P., & Smith, A. G. (1990). Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development (Cambridge, England)*, *110*(4), 1341–1348. https://doi.org/10.1242/DEV.110.4.1341

Ong, C. T., & Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews. Genetics*, *12*(4), 283. https://doi.org/10.1038/NRG2957

Osoegawa, K., Tateno, M., Woon, P. Y., Frengen, E., Mammoser, A. G., Catanese, J. J., Hayashizaki, Y., & de Jong, P. J. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Research*, *10*(1). https://doi.org/10.1101/gr.10.1.116

Oudelaar, A. M., Beagrie, R. A., Gosden, M., de Ornellas, S., Georgiades, E., Kerry, J., Hidalgo, D., Carrelha, J., Shivalingam, A., El-Sagheer, A. H., Telenius, J. M., Brown, T., Buckle, V. J., Socolovsky, M., Higgs, D. R., & Hughes, J. R. (2020). Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16598-7

Oudelaar, A. M., Beagrie, R. A., Kassouf, M. T., & Higgs, D. R. (2021). The mouse alpha-globin cluster: a paradigm for studying genome regulation and organization. *Current Opinion in Genetics & Development*, *67*, 18–24. https://doi.org/10.1016/J.GDE.2020.10.003

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

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

Oudelaar, A. M., & Higgs, D. R. (2021a). The relationship between genome structure and function. *Nature Reviews. Genetics*, *22*(3), 154–168. https://doi.org/10.1038/S41576-020-00303-X

Oudelaar, A. M., & Higgs, D. R. (2021b). The relationship between genome structure and function. *Nature Reviews Genetics*, *22*(3), 154–168. https://doi.org/10.1038/S41576-020-00303-X

Panne, D., Maniatis, T., & Harrison, S. C. (2007). An Atomic Model of the Interferon-β Enhanceosome. *Cell*, *129*(6), 1111–1123. https://doi.org/10.1016/j.cell.2007.05.019

Parker, S. C. J., Stitzel, M. L., Taylor, D. L., Orozco, J. M., Erdos, M. R., Akiyama, J. A., van Bueren, K. L., Chines, P. S., Narisu, N., Black, B. L., Axel, V., Pennacchio, L. A., & Collins, F. S. (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(44), 17921–17926. https://doi.org/10.1073/PNAS.1317023110/SUPPL_FILE/SD01.XLS

Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A., & Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. *Science*, *339*(6120), 698–699. https://doi.org/10.1126/SCIENCE.1231382/SUPPL_FILE/1231382.PENGELLY.SM.PDF

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., de Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & 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

Peters, J. M., Tedeschi, A., & Schmitz, J. (2008). The cohesin complex and its roles in chromosome biology. *Genes & Development*, *22*(22), 3089–3114. https://doi.org/10.1101/GAD.1724308

Philipsen, S., & Hardison, R. C. (2018). Evolution of hemoglobin loci and their regulatory elements. *Blood Cells, Molecules & Diseases*, *70*, 2–12. https://doi.org/10.1016/J.BCMD.2017.08.001

Pott, S., & Lieb, J. D. (2015). What are super-enhancers? *Nature Genetics*, *47*(1), 8–12. https://doi.org/10.1038/NG.3167

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 2016 48:6*, *48*(6), 681–686. https://doi.org/10.1038/ng.3550

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. https://doi.org/10.1093/BIOINFORMATICS/BTQ033

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2010). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature 2010 470:7333*, *470*(7333), 279–283. https://doi.org/10.1038/nature09692

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. https://doi.org/10.1093/NAR/GKW257

Rinzema, N. J., Sofiadis, K., Tjalsma, S. J. D., Verstegen, M. J. A. M., Oz, Y., Valdes-Quezada, C., Felder, A.-K., Filipovska, T., Elst, S. van der, Ramos, Z. de A. dos, Han, R., Krijger, P. H. L., & Laat, W. de. (2021). Building regulatory landscapes: enhancer recruits cohesin to create contact domains, engage CTCF sites and activate distant genes. *BioRxiv*, 2021.10.05.463209. https://doi.org/10.1101/2021.10.05.463209

Rogozin, I. B., Makarova, K. S., Natale, D. A., Spiridonov, A. N., Tatusov, R. L., Wolf, Y. I., Yin, J., & Koonin, E. v. (2002). Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. *Nucleic Acids Research*, *30*(19), 4264. https://doi.org/10.1093/NAR/GKF549

Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. v., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers,

J., Vasile, E., Malik, S., Hnisz, D., Lee, T. I., Cisse, I. I., … Young, R. A. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. *Science (New York, N.Y.)*, *361*(6400). https://doi.org/10.1126/SCIENCE.AAR3958

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

Sartorelli, V., & Lauberth, S. M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. *Nature Structural & Molecular Biology 2020 27:6*, *27*(6), 521–528. https://doi.org/10.1038/s41594-020-0446-0

Schaft, J., Ashery-Padan, R., van Hoeven, F. der, Gruss, P., & Francis Stewart, A. (2001). Efficient FLP recombination in mouse ES cells and oocytes. *Genesis*, *31*(1). https://doi.org/10.1002/gene.1076

Schmidl, C., Rendeiro, A. F., Sheffield, N. C., & Bock, C. (2015). ChIPmentation: Fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature Methods*, *12*(10). https://doi.org/10.1038/nmeth.3542

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

Schübeler, D., Groudine, M., & Bender, M. A. (2001). The murine β-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11432. https://doi.org/10.1073/PNAS.201394698

Schwessinger, R., Suciu, M. C., McGowan, S. J., Telenius, J., Taylor, S., Higgs, D. R., & Hughes, J. R. (2017). Sasquatch: Predicting the impact of regulatory SNPs on transcription factor binding from cell- and tissue-specific DNase footprints. *Genome Research*, *27*(10). https://doi.org/10.1101/gr.220202.117

Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, *16*(1), 1–11. https://doi.org/10.1186/S13059-015-0831-X/TABLES/4

Shen, Y., Yue, F., Mc Cleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Ren, B., & Lobanenkov, V. v. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature 2012 488:7409*, *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). https://doi.org/10.1038/ng.3606

Sigova, A. A., Abraham, B. J., Ji, X., Molinie, B., Hannett, N. M., Guo, Y. E., Jangi, M., Giallourakis, C. C., Sharp, P. A., & Young, R. A. (2015). Transcription factor trapping by RNA in gene regulatory elements. *Science (New York, N.Y.)*, *350*(6263), 978. https://doi.org/10.1126/SCIENCE.AAD3346

Singh, G., Mullany, S., Moorthy, S. D., Zhang, R., Mehdi, T., Tian, R., Duncan, A. G., Moses, A. M., & Mitchell, J. A. (2021). A flexible repertoire of transcription factor binding sites and a diversity threshold determines enhancer activity in embryonic stem cells. *Genome Research*, *31*(4), 564–575. https://doi.org/10.1101/GR.272468.120/-/DC1

Smith, A. G. (1991). Culture and differentiation of embryonic stem cells. *Journal of Tissue Culture Methods 1991 13:2*, *13*(2), 89–94. https://doi.org/10.1007/BF01666137

Smith, A. J. H., Xian, J., Richardson, M., Johnstone, K. A., & Rabbitts, P. H. (2002). Cre-loxP chromosome engineering of a targeted deletion in the mouse corresponding to the 3p21.3 region of homozygous loss in human tumours. *Oncogene*, *21*(29). https://doi.org/10.1038/sj.onc.1205530

Sollaino, M. C., Paglietti, M. E., Loi, D., Congiu, R., Podda, R., & Galanello, R. (2010). Homozygous deletion of the major alpha-globin regulatory element (MCS-R2) responsible for a severe case of hemoglobin H disease. *Blood*, *116*(12), 2193–2194. https://doi.org/10.1182/BLOOD-2010-04-281345

Soutourina, J. (2017). Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology 2017 19:4*, *19*(4), 262–274. https://doi.org/10.1038/nrm.2017.115

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

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

Stark, R., & Brown, G. (n.d.). *DiffBind: Differential binding analysis of ChIP-Seq peak data*.

Tamary, H., & Dgany, O. (2020). Alpha-Thalassemia. *GeneReviews®*. https://www.ncbi.nlm.nih.gov/books/NBK1435/

Tang, F., Yang, Z., Tan, Y., & Li, Y. (2020). Super-enhancer function and its application in cancer targeted therapy. *Npj Precision Oncology 2020 4:1*, *4*(1), 1–7. https://doi.org/10.1038/s41698-020-0108-z

Thomas, H. F., Kotova, E., Jayaram, S., Pilz, A., Romeike, M., Lackner, A., Penz, T., Bock, C., Leeb, M., Halbritter, F., Wysocka, J., & Buecker, C. (2021). Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Molecular Cell*, *81*(5). https://doi.org/10.1016/j.molcel.2020.12.047

Thorne, N., Inglese, J., & Auld, D. S. (2010). Illuminating Insights into Firefly Luciferase and Other Bioluminescent Reporters Used in Chemical Biology. *Chemistry & Biology*, *17*(6), 646–657. https://doi.org/10.1016/J.CHEMBIOL.2010.05.012

Vernimmen, D., Gobbi, M. de, Sloane-Stanley, J. A., Wood, W. G., & Higgs, D. R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *The EMBO Journal*, *26*(8), 2041. https://doi.org/10.1038/SJ.EMBOJ.7601654

Vian, L., Pękowska, A., Rao, S. S. P., Kieffer-Kwon, K. R., Jung, S., Baranello, L., Huang, S. C., el Khattabi, L., Dose, M., Pruett, N., Sanborn, A. L., Canela, A., Maman, Y., Oksanen, A.,

Resch, W., Li, X., Lee, B., Kovalchuk, A. L., Tang, Z., … Casellas, R. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*, *173*(5), 1165-1178.e20. https://doi.org/10.1016/j.cell.2018.03.072

Viprakasit, V., Harteveld, C. L., Ayyub, H., Stanley, J. S., Giordano, P. C., Wood, W. G., & Higgs, D. R. (2006). A novel deletion causing α thalassemia clarifies the importance of the major human alpha globin regulatory element. *Blood*, *107*(9), 3811–3812. https://doi.org/10.1182/BLOOD-2005-12-4834

Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Research*, *35*(Database issue), D88. https://doi.org/10.1093/NAR/GKL822

Wallace, H. A. C., Marques-Kranc, F., Richardson, M., Luna-Crespo, F., Sharpe, J. A., Hughes, J., Wood, W. G., Higgs, D. R., & Smith, A. J. H. (2007). Manipulating the Mouse Genome to Engineer Precise Functional Syntenic Replacements with Human Sequence. *Cell*, *128*(1). https://doi.org/10.1016/j.cell.2006.11.044

Wang, X., Cairns, M. J., & Yan, J. (2019). Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Research*, *47*(22), 11481–11496. https://doi.org/10.1093/NAR/GKZ1038

Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, *153*(2). https://doi.org/10.1016/j.cell.2013.03.035

Wickham, H. (2016). *ggplot2*. https://doi.org/10.1007/978-3-319-24277-4

Xu, J., Shao, Z., Glass, K., Bauer, D. E., Pinello, L., van Handel, B., Hou, S., Stamatoyannopoulos, J. A., Mikkola, H. K. A., Yuan, G. C., & Orkin, S. H. (2012). Combinatorial assembly of developmental stage-specific enhancers controls gene expression programs during human erythropoiesis. *Developmental Cell*, *23*(4), 796–811. https://doi.org/10.1016/J.DEVCEL.2012.09.003

Yamagata, K., Nakayamada, S., & Tanaka, Y. (2020). Critical roles of super-enhancers in the pathogenesis of autoimmune diseases. *Inflammation and Regeneration*, *40*(1), 1–9. https://doi.org/10.1186/S41232-020-00124-9/TABLES/2

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., & Stark, A. (2014). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature 2014 518:7540*, *518*(7540), 556–559. https://doi.org/10.1038/nature13994

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, T., Zhang, Z., Dong, Q., Xiong, J., & Zhu, B. (2020). Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. *Genome Biology*, *21*(1), 1–7. https://doi.org/10.1186/S13059-020-01957-W/FIGURES/2

Zhang, Y., Buchholz, F., Muyrers, J. P. P., & Francis Stewart, A. (1998). A new logic for DNA engineering using recombination in Escherichia coli. In *Nature Genetics* (Vol. 20, Issue 2). https://doi.org/10.1038/2417

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), 1–9. https://doi.org/10.1186/GB-2008-9-9-R137/FIGURES/3

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). https://doi.org/10.1038/nature08531

Zuin, J., Roth, G., Zhan, Y., Cramard, J., Redolfi, J., Piskadlo, E., Mach, P., Kryzhanovska, M., Tihanyi, G., Kohler, H., Eder, M., Leemans, C., van Steensel, B., Meister, P., Smallwood, S., & Giorgetti, L. (2022). Nonlinear control of transcription through enhancer–promoter interactions. *Nature 2022 604:7906*, *604*(7906), 571–577. https://doi.org/10.1038/s41586-022-04570-y