Maps and mechanisms of three-dimensional genome organization

Seungsoo Kim

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Jay Shendure, Chair

William Noble

Christine Disteche

Program Authorized to Offer Degree:

Department of Genome Sciences

University of Washington

Abstract

Maps and mechanisms of three-dimensional genome organization

Seungsoo Kim

Chair of the Supervisory Committee:
Professor Jay Shendure
Department of Genome Sciences

The three-dimensional organization of the genome inside the nucleus both impacts and is influenced by its functions, including transcription and DNA replication. Recent technological advances, particularly the high-throughput sequencing adaptation of the chromosome conformation capture assay called Hi-C, have improved the genomic coverage and resolution of maps of 3D genome organization. However, it remains challenging to resolve the two homologous copies of the genome that exist in most eukaryotic cells. Furthermore, for many chromosomal structures, the mechanisms driving their formation remains unknown, in part due to the difficulty of testing many perturbations for effects on these structures. In my thesis work, I first utilized the Hi-C method on diverged hybrids to map the conformations of homologous chromosomes, using *Saccharomyces* yeasts as a model system. I then developed a pooled

mutational scanning method for studying chromosome conformation, and applied it to dissect the mechanisms underlying a novel inducible homolog pairing contact. This work both sheds light on the poorly understood process of mitotic homologous chromosome pairing and provides a powerful new approach for mechanistic studies of chromosome conformation.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank the many people who have, each in their own way, made this work possible. First, my advisor Jay, who has always been the best cheerleader and encouraged my optimistic view of science. Maitreya, who has been like an advisor to me and always put me in touch with the right people. Bill, Christine, and everyone in the 4D Nucleome Project who has helped me think about chromosome conformation. All current and past members of the Shendure lab, who have inspired me with your creativity and taught me all the molecular biology I know. Current and past members of the Dunham lab, who were always willing to answer my questions and taught me all I know about yeast. Ivan, whose idea to do Hi-C on hybrid yeast saved my graduate career, and prepared endless Hi-C libraries at a moment's notice. My collaborators Jason and Donna, and Sarah and Toshi, who have exemplified what collaborators should be like. My friends from college, my GS cohort, and other Seattle friends, who have kept me company through adventures outside the lab, whether puzzle challenges, eclipse-watching trips, or backpacking trips. I wouldn't be in academia if it weren't for Roy and Tami, who gave me my first taste of intellectual freedom in science. My family, who always nurtured my curiosity and taught me to believe myself. Last but not least, Katherine, who has been my constant companion through the highs and the lows.

Chapter 1. INTRODUCTION

Note: Chapters 1 and 4 were adapted from a manuscript currently in preparation.

1.1 ABSTRACT

Transcription factors (TFs) bind DNA in a sequence-specific manner and thereby serve as the protein anchors of 3D genome organization. Conversely, chromatin loops are thought to be a key mechanism by which TFs bound at enhancer regions of the genome activate distant target genes. However, our understanding of the mechanistic roles of TFs in 3D genome organization remains limited, in part due to this interdependency. In this review, we summarize the evidence for the diverse mechanisms by which TFs shape 3D genome organization. We further highlight outstanding questions, and finally, we discuss potential approaches for untangling the complex relationship between TF activity and 3D genome organization.

1.2 INTRODUCTION

Transcription factors (TFs) are proteins that play a key role in regulating gene expression, by recognizing and directly binding specific DNA sequences (Lambert et al., 2018). Upon binding DNA, TFs can activate transcription by directly (Chernukhin et al., 2007) or indirectly (via cofactors) bringing RNA Polymerase II and other transcription machinery to the promoter sequence at the start of a gene (Allen and Taatjes, 2015; Näär et al., 2001). Alternatively, they can repress transcription by recruiting corepressors or interfering with the binding of other TFs (Braun and Johnson, 1997; Moody et al., 2005; Nan et al., 1997; Sankaran et al., 2008). Some TF binding sites occur near promoters, but the overwhelming majority occur in clusters of binding

sites called enhancers, which are typically tens to hundreds of kilobases away from the promoters they activate (Neph et al., 2012).

This action-at-a-distance is largely thought to be mediated by the three-dimensional looping of DNA that brings the enhancer in physical proximity to the promoter (Furlong and Levine, 2018). Recent technical advances have enabled increasingly comprehensive and high-resolution views of chromosome conformation (Bonev et al., 2017; Lieberman-Aiden et al., 2009; Rao et al., 2017). Although the link between 3D genome organization and TF activity is undeniable, the detailed mechanistic and causal relationships between the two remain elusive. In what cases is physical enhancer-promoter contact frequency critical for transcriptional regulation? As DNA-binding proteins, TFs can directly and indirectly shape chromosome conformation in addition to being influenced by it. To what extent is chromosome conformation the consequence, rather than the cause, of TF activity? The diverse interactions among TFs and cofactors can each impact genome organization, further complicating the picture.

Here, we review the intricate interplay between TF activity and 3D genome organization, including both the mechanisms by which TFs shape 3D genome organization and the impacts of chromosome conformation on TF activity. We primarily focus on mammalian genomes in interphase, but also include discussion of other organisms where relevant.

1.3 Modes of TF action on the 3D genome

Transcription factors are defined by their ability to bind DNA, but generally function via interactions with other proteins and even RNAs (Lambert et al., 2018). These interactions all have the potential to impact chromosome conformation; in this section, we outline and categorize the mechanisms by which TFs can shape the 3D genome.

1.3.1 *Direct oligomerization*

The most intuitive example of TF-driven DNA loops is direct oligomerization of TFs (**Figure 1.1A**). Some of the earliest evidence of such loops came from studies of cooperative repressor protein binding to DNA in prokaryotes (Griffith et al., 1986). Several repressors, such as LacI (Priest et al., 2014) and GalR (Qian et al., 2012) in *Escherichia coli* and the lambda phage CI repressor (Dodd et al., 2004; Griffith et al., 1986), bind cooperatively not only at nearby binding sites, but can form larger oligomers that tether together distant sets of binding sites, both *in vitro* and *in vivo*.

Eukaryotic TFs are also capable of producing similar DNA loops via self-association *in vitro* (Su et al., 1991), although whether this self-association is structured and how this interaction behaves *in vivo* are less clear. Yin Yang 1 (YY1) is a ubiquitously expressed TF that binds both promoters and enhancers and forms homodimers, and thereby could mediate enhancer-promoter looping by a similar mechanism (Weintraub et al., 2017). CCCTC-binding factor (CTCF) is another ubiquitous TF with a key architectural role in mammals, and is thought to form dimers (and possibly oligomers) *in vitro* and *in vivo* (Pant et al., 2004; Yusufzai et al., 2004). However, as we discuss in subsequent sections, it is unclear whether this dimerization is the means by which CTCF shapes 3D genome organization.

1.3.2 *Cofactor oligomerization*

TFs can also form DNA loops by recruiting cofactor proteins that in turn form oligomers (**Figure 1.1B**). For example, LIM domain binding protein 1 (Ldb1) is a notable adaptor protein capable of dimerization that does not bind DNA directly, but is instead recruited to its target loci by TFs or cofactors. In mouse olfactory sensory neurons, it is recruited by the TFs Lhx2 and

Ebf1 (Monahan et al., 2019), whereas in mouse erythroid cells, the cofactor Lmo2 bridges Ldb1's interaction with TFs Gata1, Tal1, and E2A (Love et al., 2014b). Ldb1 recruitment is also sufficient for DNA looping, as evidenced by studies in which Ldb1 fused to a designed zinc finger protein targeting either the beta or gamma globin gene promoters induced looping with the locus control region (LCR) and subsequent gene activation (Deng et al., 2012, 2014).

1.3.3 *Condensate formation*

In addition to the strong protein-protein interactions at structured interfaces that typically underlie oligomerization, weak but multivalent interactions among intrinsically disordered regions (IDRs) can produce phase-separated liquid condensates (Banani et al., 2017), which could shape genome organization by tethering distant regions together (**Figure 1.1C**). TFs and coactivators are particularly enriched for IDRs (Liu et al., 2006a; Staby et al., 2017), and recent studies have demonstrated several examples in which they form condensates *in vitro* and *in vivo* (Boija et al., 2018; Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018). Enhancers sharing binding by specific sets of TFs and cofactors have been observed to colocalize in "nuclear microenvironments" enriched for those factors (Tsai et al., 2017), consistent with condensates shaping chromosome conformation. Similarly, the heterochromatin protein HP1a is capable of phase separating and of compacting DNA *in vitro*, consistent with a role in the self-association and compaction of heterochromatic regions of the genome (Larson et al., 2017; Strom et al., 2017).

1.3.4 *Interactions with loop extruders*

A unique class of interactions that plays an important role in 3D genome organization is interactions between TFs and cohesin and other Structural Maintenance of Chromosomes (SMC)

complexes implicated in loop extrusion (**Figure 1.1D**). SMC complexes can encircle DNA and actively pull a loop of DNA through the ring (Ganji et al., 2018; Terakawa et al., 2017), until they are blocked or unloaded. The loop extrusion model (Alipour and Marko, 2012; Fudenberg et al., 2016; Sanborn et al., 2015) is consistent with a growing amount of experimental evidence, most notably in interphase nuclei, where the blocking of cohesin complexes by CTCF explains the boundaries of topologically associating domains (TADs), which grow (Haarhuis et al., 2017) or disappear (Rao et al., 2017; Schwarzer et al., 2017) upon perturbations to cohesin unloading or loading, respectively. Other TFs may also be capable of altering cohesin loop positions: recent experiments mutating the RNA-binding domain of CTCF led to new loop boundaries at binding sites for Oct4 and other pluripotency factors in mouse embryonic stem cells (Hansen et al., 2018).

1.3.5 *Protein-RNA interactions*

In addition to protein-protein interactions, protein-RNA binding plays a role in structuring the 3D genome (**Figure 1.1E**). Both YY1 and CTCF, ubiquitously expressed TFs with structural roles, are known to bind noncoding RNAs (Jeon and Lee, 2011; Kung et al., 2015; Sigova et al., 2015). Specifically depleting these interactions, either by adding RNase *in vitro* or mutating the RNA-binding domain, weakens the ability of YY1 to form loops (Weintraub et al., 2017) and of CTCF to block cohesin (Hansen et al., 2018). These protein-RNA interactions provide a plausible mechanism by which noncoding RNAs like enhancer RNAs (Sigova et al., 2015) or long noncoding RNAs like Xist (Jeon and Lee, 2011; Kung et al., 2015) and Firre (Yang et al., 2015) might shape 3D genome organization.

1.3.6 *Chromatin modification*

Finally, TFs can impact 3D genome organization by modifying their chromatin context, and thus the recruitment of other TFs or proteins (**Figure 1.1F**). Many TFs recruit cofactors that modify DNA (e.g. methylation), histone tails (e.g. lysine methylation or acetylation), or nucleosome positioning, which can affect the binding of other TFs or cofactors (O'Malley et al., 2016; Yin et al., 2017; Zhu et al., 2018). TFs or proteins that either specifically bind or cannot bind to modified DNA (e.g. methyl-CpG binding protein 2 [MeCP2]) or histones (e.g. the acetyl-lysine binding BRD4) can then shape chromosome conformation through protein-protein and protein-RNA interactions. Indeed, chromatin states are associated with multiple scales of genome organization: DNA methylation can affect CTCF binding and thereby enhancer-promoter looping (Flavahan et al., 2016; Murrell et al., 2004), while larger genomic regions enriched in active or inactive chromatin marks self-associate in A or B compartments, respectively (Lieberman-Aiden et al., 2009). Even without recruiting cofactors, TFs can establish or maintain open chromatin by displacing nucleosomes (Zhu et al., 2018), and quantitative trait locus (QTL) analysis suggests that this is a key mechanism by which TFs influence 3D genome organization (Tehranchi et al., 2019). These indirect, chromatin-mediated effects may provide a means by which TFs with only transient binding to DNA may trigger longer-lasting changes to chromosome conformation and ultimately gene expression.

1.4 A MECHANISTIC VIEW OF 3D GENOME STRUCTURES

We next apply the principles outlined above to review the state of our mechanistic understanding of the various features of 3D genome organization. However, direct evidence for specific mechanisms is limited; thus, we also highlight key remaining questions.

1.4.1 *A/B compartments*

Each chromosome can be segmented into active A compartments and inactive B compartments, roughly 1-10 Mb in size, each of which preferentially associates with other compartments of the same identity (Lieberman-Aiden et al., 2009; Wang et al., 2017). This self-association is consistent with condensate formation by proteins specific to either active or inactive chromatin states (Rowley et al., 2017), such as Pol II or HP1a, respectively. Although the A/B compartment pattern is generally similar across cell types, reflecting the alternating pattern of gene-rich and gene-poor regions of the genome, the pattern also shifts with chromatin states upon differentiation (Stadhouders et al., 2018), lending support to this model. However, the relative contributions of specific factors remain unclear.

Furthermore, phase separation of active and inactive chromatin is antagonized by cohesin-mediated loop extrusion (Nuebler et al., 2018). Depleting the cohesin unloader leads to weaker compartments (Haarhuis et al., 2017), while depleting cohesin leads to stronger and finer compartments (Schwarzer et al., 2017).

1.4.2 *TADs and loop domains*

Zooming in further, compartments can be subdivided into topologically associating domains (TADs), roughly 0.1-1 Mb regions that preferentially self-associate and are usually bounded by CTCF binding sites and/or housekeeping gene promoters (Dixon et al., 2012; Nora et al., 2012). A growing body of evidence supports the role of loop extrusion by cohesin coupled with CTCF blocking of cohesin in the establishment of TADs (Fudenberg et al., 2016; Nora et al., 2017; Nuebler et al., 2018; Rao et al., 2017; Sanborn et al., 2015). However, the molecular

mechanism by which CTCF blocks cohesin remains unclear. The bias toward convergent CTCF motifs at TAD boundaries suggests that CTCF blocks cohesin in a directional manner (Rao et al., 2014; de Wit et al., 2015), but the structural basis of this bias is unknown. CTCF is thought to dimerize, but is this a side-effect of two molecules each blocking progression of a single cohesin complex, or does CTCF dimerization contribute to cohesin-blocking capacity? Furthermore, despite its constitutive expression, CTCF binding varies across cell types; what features of chromatin state drive this differential binding?

Furthermore, recent studies have highlighted the role of other factors in TAD establishment. A systematic study of the Dppa2/4 locus in mouse embryonic stem cells revealed that deletion of TAD boundaries is not sufficient to disrupt TAD structure, and internal early replication control elements (ERCEs) instead determine TAD structure (Sima et al., 2019). This is corroborated by other similar observations (Lupiáñez et al., 2015) and recent microscopy evidence that even upon cohesin depletion, TAD-like structures form in single cells but with random boundary locations (Bintu et al., 2018). ERCEs resemble enhancers and promoters, but do all enhancers and promoters contribute to TADs? If not, what features determine their relative contributions? If condensate formation is responsible, one prediction would be that high densities of binding or recruitment sites for the driving factors would be correlated to greater roles for TAD establishment (Hnisz et al., 2017).

1.4.3 *Architectural stripes*

Some TADs contain “architectural stripes” or frequently interacting regions (FIREs), regions that form many contacts throughout their domains. Stripes are associated with tissue-specific super-enhancers, and are thought to be formed by loop extrusion, through robust loading of cohesin accompanied by a cluster of unidirectional CTCF binding sites (a stripe anchor) that

blocks loop extrusion on one of the two ends of the loop (Kraft et al., 2019; Schmitt et al., 2016a; Vian et al., 2018). What then determines the genomic positioning of cohesin loading? Do the condensate-forming properties of super-enhancers play a role? In yeast, the RSC chromatin remodeler recruits cohesin loading (Muñoz et al., 2019), and in *Drosophila*, DNA replication complexes have been implicated (Pherson et al., 2019). Another potential contributor is that the recruitment of RNA polymerase II to active enhancers increases the mobility of those loci, thereby increasing the frequency of long-range contacts (Gu et al., 2018).

1.4.4 *Enhancer-promoter loops*

Not all active enhancers are near stripe anchors, and yet presumably act on distal promoters by DNA looping. Do all active enhancers preferentially form specific contacts with their cognate promoters, or is the "random" background contact frequency from being in the same TAD sufficient for function (**Figure 1.2**)? The prevalence of enhancer-promoter pairs with (1) architectural stripes at either the enhancer or promoter (**Figure 1.2A**), (2) focal enhancer-promoter contacts (**Figure 1.2B**), and (3) TAD background contact frequency (**Figure 1.2C**) remain difficult to test systematically, due to the paucity of functionally validated enhancer-promoter pairs. But in at least some cases, enhancers appear to form focal contacts with their target genes (Bonev et al., 2017; Gasperini et al., 2019).

What are the mechanisms of these contacts, and are all contacts functionally equivalent? In the canonical example of enhancer-promoter (E-P) looping at the beta-globin locus, loops are mediated by dimerization of the adaptor protein Ldb1 (Love et al., 2014b). An shRNA screen in mouse ESCs implicated the Mediator and cohesin complexes (Kagey et al., 2010), but how these shape E-P loops is unclear. Mediator can form condensates (Cho et al., 2018), along with BRD4 and other coactivators (Sabari et al., 2018), that can bridge enhancers and promoters, but it is

unclear whether this provides specificity among enhancers and promoters bound by specific TFs (Tsai et al., 2017). A key remaining question is what is the logic of condensate formation - how many distinct types of condensates are there, and what drives their specificity? Mediator and cohesin also interact *in vitro* (Kagey et al., 2010), but how do they interact *in vivo*? Does Mediator block loop extrusion by cohesin? A recent study suggested that YY1 forms E-P loops by binding both the enhancer and promoter and dimerizing (Weintraub et al., 2017), with an additional role for enhancer RNAs that aid YY1 recruitment (Sigova et al., 2015).

1.4.5 *Trans contacts between TADs and chromosomes*

Although recent attention has been focused on contacts within TADs, some of the first clues that aggregative processes like condensate formation might shape genome conformation came from observations of colocalization among co-regulated genes across multiple chromosomes (Schoenfelder et al., 2009). In mouse erythroid cells, genes regulated by Klf1 are enriched for colocalization with each other and with clusters of Pol II in so-called "transcription factories" (Schoenfelder et al., 2009). *In vivo* microscopy of *Drosophila* embryos confirmed similar association of homologous and ectopic *svb* alleles (Tsai et al., 2017). Condensate formation by TFs and associated cofactors is a likely mechanism of such contacts among enhancers and promoters (Boija et al., 2018; Chong et al., 2018), but remains to be directly tested, as mutations and drugs that disrupt condensate formation have not been tested for effects on trans contacts.

Oligomerizing TFs and cofactors also contribute to interchromosomal contacts. The cofactor Ldb1 mediates interchromosomal contacts among olfactory gene clusters in mouse olfactory sensory neurons (Monahan et al., 2019). However, in most cases, whether oligomerization *per se* shapes 3D genome organization is unclear. In heat-shocked yeast, the heat

shock response protein Hsf1 mediates clustering of Hsf1 target genes (Chowdhary et al., 2019), but whether Hsf1's ability to form trimers mediates these contacts remains unclear. Similarly, YY1 and CTCF are implicated in the transient homolog pairing of the X inactivation centers in differentiating mouse ES cells, where the role of CTCF is presumably unrelated to loop extrusion (Xu et al., 2007).

Oligomerizing factors do not form all possible contacts among their binding sites, suggesting that additional mechanisms must control *trans* contacts, particularly for highly restricted contacts like focal homolog pairing (Hogan et al., 2015; Kim et al., 2017; Xu et al., 2006). For example, Rgt1 mediates homologous pairing between *TDA1* promoters and *HXT3* promoters in saturated yeast, but not contacts between the two distinct promoters (Kim et al., 2018a). Future work testing perturbations of cofactors, specific domains, and catalytic positions will help distinguish whether these TFs mediate pairing through direct oligomerization, chromatin state, or other effects.

1.4.6 *Nuclear localization*

Many DNA-DNA contacts, including those in *cis* but particularly those in *trans*, reflect shared localization at a nuclear subcompartment. For example, gene-dense and highly transcribed regions form "hubs" of contacts at nuclear speckles (Chen et al., 2018b; Quinodoz et al., 2018), whereas silenced heterochromatic regions colocalize at nucleoli (Quinodoz et al., 2018) or the nuclear lamina (Guelen et al., 2008). Regions specifically targeted by the Polycomb repressive complex form contacts at Polycomb bodies (Denholtz et al., 2013).

In each case, the colocalization of specific genomic regions at these membraneless compartments is consistent with phase separated condensates, whether they are formed by TFs, cofactors, transcription machinery, and/or splicing machinery. Recent studies have begun to

demonstrate specific examples of factors that could drive phase separation. HP1a, which binds heterochromatin, is capable of forming condensates and compacting DNA *in vitro* (Larson et al., 2017), and displays the hallmarks of such condensates *in vivo* (Strom et al., 2017). However, each subcompartment often contains many enriched factors (Fong et al., 2013), and their contributions to condensate formation are unclear. Can a factor can be colocalized with such condensates but be dispensable for their formation? These compartments are also each associated with either active or inactive transcription, and the role of transcription or lack thereof in genomic localization is complex and will require further dissection (van Steensel and Furlong, 2019).

Many of these nuclear subcompartments are also associated with specific chromatin marks, such as H3K9me3 for heterochromatin; are marks are sufficient for, or simply correlated with, nuclear localization? In the case of H3K9me3, recruiting SUV39H1 to a lacO array was sufficient to reposition the locus to heterochromatin, but a chromodomain mutant capable of depositing H3K9me3 but unable to bind it was not (Wijchers et al., 2016). Thus, protein binding to specific histone modifications can play a key role in nuclear organization.

Also, what role does random genomic colocalization play in establishing nuclear subcompartments? The local abundance of binding sites for condensate-forming factors is a key parameter in determining their formation (Alberti, 2017; Hnisz et al., 2017), and once formed, a transient contact could be stabilized. In some cases, recruiting a nucleating factor to a single locus can be sufficient to create a condensate, e.g. for Cajal bodies (Wang et al., 2018), but this may not be true for other types of subcompartments.

1.5 THE IMPLICATIONS OF THE 3D GENOME FOR TF FUNCTION

Thus far, we have discussed the mechanisms by which TFs can shape 3D genome organization. In this section, we consider the converse—how does 3D genome organization affect TF activity?

1.5.1 *Enhancer-promoter specificity*

One major way in which chromosome conformation is thought to regulate TF activity is providing specificity in pairing between enhancers and the genes they target. This idea has led to efforts to use targeted assays of 3D genome organization, like ChIA-PET (Fullwood et al., 2009), HiChIP (Mumbach et al., 2016), promoter capture Hi-C (Javierre et al., 2016; Mifsud et al., 2015; Schoenfelder et al., 2015), and proximity ligation assisted ChIP-seq (PLAC-seq) (Fang et al., 2016) to associate enhancers (and their associated GWAS SNPs) with candidate target genes. Yet in some cases physical proximity is known to precede enhancer-promoter activation during development or *in vitro* differentiation (Ghavi-Helm et al., 2014; Stadhouders et al., 2018). Furthermore, enhancer-promoter pairing, both in the sense of the 3D genome and function, is not binary but quantitative; what contact frequency is sufficient for function?

Until recently, there has been insufficient data on enhancer-promoter activation to evaluate the correlation between physical and functional pairing. Two recent studies have begun to address this by using CRISPRi to perturb enhancer function at scale, coupled with either single-cell RNA-seq (Gasperini et al., 2019) or Flow-FISH to identify genes with altered expression (Fulco et al., 2019). In both studies, functional enhancer-promoter pairs exhibited more frequent physical contacts than controls. However, physical contacts alone are not sufficient to predict functional pairing; instead, a combination of 3D contacts and enhancer

activity, as measured by DNase hypersensitivity and H3K27ac signal, predicts an enhancer's relative contribution to a target genes' activation (Fulco et al., 2019). Notably, the ability to predict experimental enhancer perturbation data was only marginally impacted by excluding high-resolution Hi-C data, suggesting that genomic proximity, rather than specific TAD boundaries or focal enhancer-promoter loops, plays a major role in this specificity. However, this model does not explain why some promoters, despite proximity to strong enhancers, are not activated. In addition to enhancer activity levels, promoter activity levels (regulated by heterochromatin and interactions with the nuclear lamina, binding of TFs and chromatin accessibility, histone modifications, etc) could regulate the ability to be activated by enhancers (Zabidi and Stark, 2016).

Mechanistic details of physical enhancer-promoter pairing can impact enhancer function beyond simply providing specificity of pairing. For example, whether an enhancer can simultaneously pair with multiple promoters would impact whether or not those promoters need to compete for activation by the enhancer (Fukaya et al., 2016; Fulco et al., 2016). Although Hi-C and related datasets are uninformative with respect to temporal dynamics, the timing and stability of E-P loops might impact the variability in enhancer activation of target promoters.

1.5.2 *Biophysics of TF activity*

The spatial organization of the genome, and more broadly the nucleus, fundamentally impacts the biophysics of TF activity. The DNA binding activity of TFs is a function not only of the TF affinity for specific DNA sequences and the overall concentration of the TF, but also the localization of the TF within the nucleus. A cluster of binding sites for a TF, whether nearby in genomic sequence or brought together by DNA loops, could produce a locally elevated concentration of the TF that in turn allows binding even at low overall TF levels and/or at low-

affinity binding sites that differ considerably from the consensus motif (Mir et al., 2017, 2018; Tsai et al., 2017; Vockley et al., 2016). In some cases, such as the glucocorticoid response, the low-affinity binding sites are not alone sufficient for TF activity and instead amplify the activity of high-affinity sites (Vockley et al., 2016). But in other contexts, including *Drosophila* (Crocker et al., 2015; Tsai et al., 2017) and *Ciona* embryos (Farley et al., 2015, 2016), low-affinity binding sites are capable of, and even necessary for, proper tissue-specific expression.

Just as nuclear microenvironments can enrich TFs and coactivators, they can exclude factors as well (Strom et al., 2017). Precisely how heterochromatin and other repressive environments inhibit transcription is unclear; they might physically exclude TF and transcription machinery (Tjong et al., 2012), but TFs that do enter the subcompartment may be transiently trapped (Bancaud et al., 2009).

Nuclear architecture also shapes the temporal dynamics of TF activity and subsequent transcription. At many genes, transcription occurs in bursts (Suter et al., 2011), and transcription level is controlled through burst frequency rather than size (Bartman et al., 2016, 2019). At least in cases where the enhancer is relatively far from the promoter in genomic distance, enhancer-promoter contact is correlated with transcription (Chen et al., 2018a), while the size of the Pol II focus is correlated with transcript production (Cho et al., 2016). Thus, stable enhancer-promoter looping and perhaps even clustering among co-regulated genes could mediate robust transcription for crucial genes (Hnisz et al., 2017).

In addition to mediating robust expression, chromosome conformation can contribute to rare or monoallelic expression. In both endogenous and engineered systems, infrequent chromosomal contacts can lead to gene activation, perhaps by facilitating the assembly of TFs

(Apostolou and Thanos, 2008; Noordermeer et al., 2011). The stochasticity of these contacts can mediate random selection of a promoter to be activated (Guo et al., 2012; Monahan et al., 2019). However, functional roles for 3D genome organization has often been speculative, and establishing causality remains challenging. For example, homolog pairing of the X chromosomes coincides with the random choice of one chromosome for silencing in females (Xu et al., 2006), but recent tethering experiments do not support a causative role (Pollex and Heard, 2019).

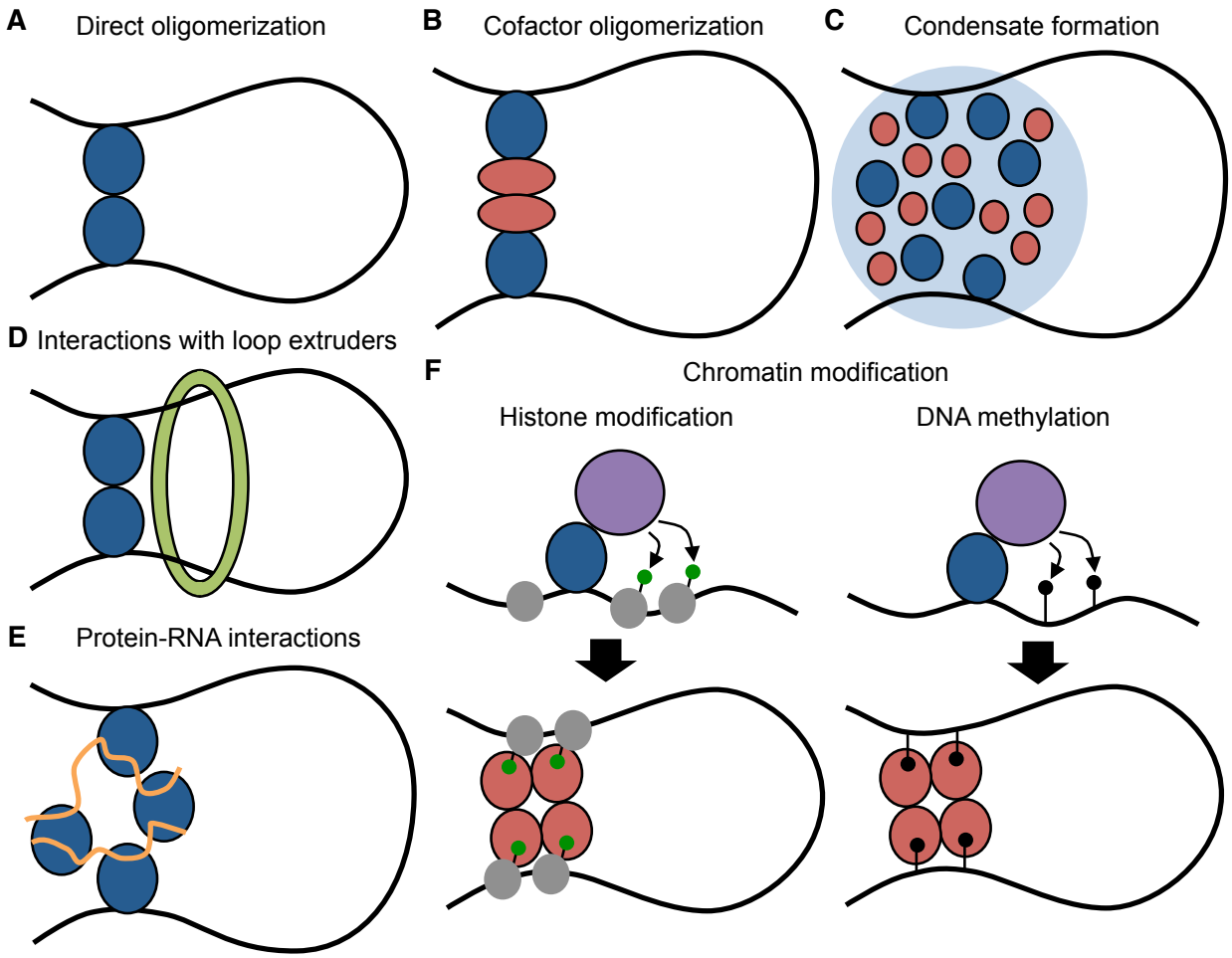

Figure 1.1. Modes of TF action on 3D genome organization.

Black lines indicate DNA; dark blue ovals represent TFs; salmon-colored ovals represent cofactors that do not directly bind DNA; green ring represents a loop extrusion complex; orange lines indicate RNA; purple ovals represent chromatin modifiers; grey ovals represent nucleosomes; green circles on stems indicate histone modifications; black circles on stems represent DNA methylation. (A) Direct oligomerization. (B) Cofactor oligomerization. (C) Condensate formation. (D) Interactions with loop extruders. (E) Protein-RNA interactions. (F) Chromatin modification, including histone modification (left panel) and DNA methylation (right panel).

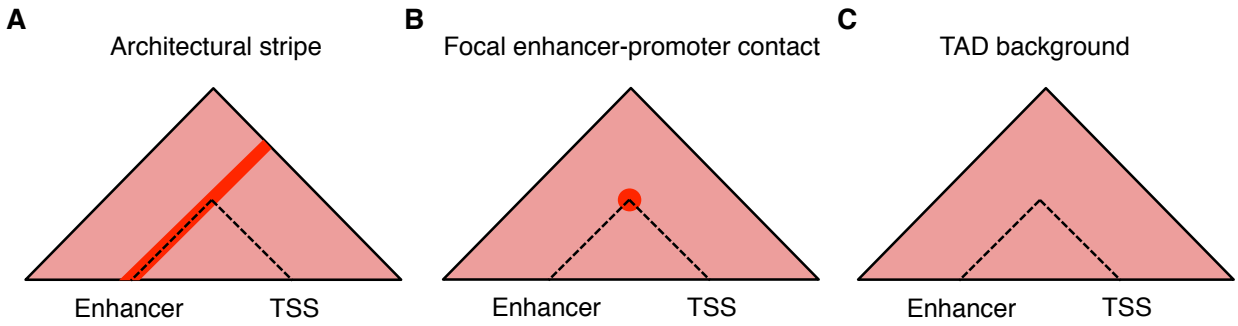

Figure 1.2. **Relative frequencies of enhancer-promoter loops.**

Schematics of contact frequency maps, zoomed in to a TAD (pink triangle); red areas indicate regions of increased contact frequency. Three possible scenarios for the contact frequency of a given enhancer-promoter (TSS) pair: (A) An architectural stripe is anchored at the enhancer, resulting in specificity of contacts for the promoter, but not the enhancer. (B) Focal enhancer-promoter contact, so that both the enhancer and promoter are enriched for contacts with the other. (C) TAD background rate of contacts, resulting in no contact specificity.

Chapter 2. THE DYNAMIC THREE-DIMENSIONAL ORGANIZATION OF THE DIPLOID YEAST GENOME

Note: Chapter 2 was published in *eLife* as Kim, S., Liachko, I., Brickner, D.G., Cook, K., Noble, W.S., Brickner, J.H., Shendure, J., and Dunham, M.J. (2017). The three-dimensional organization of the diploid yeast genome. eLife *6*, 23623.

2.1 ABSTRACT

The budding yeast *Saccharomyces cerevisiae* is a long-standing model for the three-dimensional organization of eukaryotic genomes. However, even in this well-studied model, it is unclear how homolog pairing in diploids or environmental conditions influence overall genome organization. Here, we performed high-throughput chromosome conformation capture on diverged *Saccharomyces* hybrid diploids to obtain the first global view of chromosome conformation in diploid yeasts. After controlling for the Rabl-like orientation using a polymer model, we observe significant homolog proximity that increases in saturated culture conditions. Surprisingly, we observe a localized increase in homologous interactions between the *HAS1-TDA1* alleles specifically under galactose induction and saturated growth. This pairing is accompanied by relocalization to the nuclear periphery and requires Nup2, suggesting a role for nuclear pore complexes. Together, these results reveal that the diploid yeast genome has a dynamic and complex 3D organization.

2.2 INTRODUCTION

The genome is actively organized in the nucleus in both space and time, and this organization impacts fundamental biological processes like transcription, DNA repair, and recombination (Taddei et al., 2010). The budding yeast *S. cerevisiae* has been a useful model for studying eukaryotic genome conformation and its functional implications (Taddei et al., 2010). The predominant feature of yeast 3D genome organization is its Rabl-like orientation (Jin et al., 1998) (**Figure 2.1A**): during interphase, the centromeres cluster at one end of the nucleus, attached to the spindle pole body, and chromosome arms extend outward toward the nuclear periphery where the telomeres associate (Schober et al., 2008; Therizols et al., 2010), like in anaphase. In addition, the ribosomal DNA array forms the nucleolus, opposite the spindle pole (Yang et al., 1989), splitting chromosome XII into two separate domains that behave as if they were separate chromosomes. This organization largely persists through the cell cycle (Jin et al., 1998) and even in stationary phase, albeit with increased telomere clustering and decreased centromere clustering (Guidi et al., 2015; Rutledge et al., 2015).

Genome-wide chromosome conformation capture methods like Hi-C have both confirmed these microscopy observations and permitted systematic analyses of the functional clustering of genomic elements like tRNA genes and origins of replication (Duan et al., 2010). However, multiple studies have argued that a simple volume-exclusion polymer model of chromosomes in a Rabl-like orientation is sufficient to explain microscopy and Hi-C data of the budding yeast genome (Tjong et al., 2012; Wong et al., 2012), at least in haploids grown under standard lab conditions. These studies have argued that even the functional clustering that is observed may simply be a consequence of the Rabl-like orientation coupled with biases in the

chromosomal positions of genomic elements, rather than active molecular interactions (Rutledge et al., 2015; Tjong et al., 2012).

Although this simplicity is attractive, diploidy and variable environmental conditions may add complexity to yeast genome conformation. In diploid yeast, homologous chromosomes can pair not only in meiosis, but also during mitotic growth (Burgess and Kleckner, 1999; Burgess et al., 1999; Dekker et al., 2002; Weiner and Kleckner, 1994), as they do in *Drosophila* (Metz, 1916). However, the extent of mitotic homolog pairing has been debated due to discrepancies between studies (Barzel and Kupiec, 2008). One explanation for these discrepancies is potential artifacts in the microscopy methods used to detect pairing. In fluorescence in situ hybridization (FISH), signal loss can lead to false inference of colocalization (Lorenz et al., 2003). It has also been suggested that tagging of genomic loci with repetitive arrays of GFP for live-cell imaging can directly cause pairing via GFP dimerization (Mirkin et al., 2014). Furthermore, the Rabl-like orientation alone can create the appearance of homolog pairing if not controlled for, by juxtaposing chromosomal loci at the same distance from centromeres, including homologous loci (Lorenz et al., 2003). Discrepancies in the extent of pairing between studies might be attributable to variation in pairing strength across the genome; however, mitotic homolog pairing has only been examined at a few loci.

In both haploid and diploid yeast, genome conformation can also change in response to environmental conditions. Genes that respond to signals like galactose induction (*GAL1*, *HXK1*), inositol starvation (*INO1*), oxidative stress, and heat shock (*HSP104*) have been observed by microscopy to relocate to the nuclear periphery upon activation via interactions with nuclear pore complexes (Ahmed et al., 2010; Brickner and Walter, 2004; Brickner et al., 2016; Casolari et al., 2004; Dieppois et al., 2006; Dultz et al., 2016; Taddei et al., 2006). Nuclear pore interactions can

mediate clustering of genes that share Gene Recruitment Sequences, including homologous alleles (Brickner et al., 2015, 2012, 2016; Randise-Hinchliff et al., 2016), and even impact the conformation of chromosomes well beyond the induced gene (Dultz et al., 2016). However, because such changes in conformation are primarily detected by microscopy, systematic studies of how inducible gene relocalization impacts global genome conformation have been lacking.

Here, we present a genome-wide analysis of diploid chromosome conformation in budding yeasts in multiple environmental conditions. We utilize hybrid yeasts resulting from mating diverged yeast species to perform homolog-resolved Hi-C. Our genomic approach allows us to more fully account for the Rabl-like orientation in assessing the extent of homolog pairing, and to detect whether some regions of the genome pair more strongly than others. We find that the strength of pairing varies across both growth conditions and the genome. Notably, the homologous *HAS1-TDA1* alleles on chromosome XIII pair specifically in galactose induction and saturated growth, but not during exponential growth in glucose. The condition-specific pairing is accompanied by relocalization to the nuclear periphery and in galactose requires Nup2, a component of the nuclear pore, suggesting a role for the nuclear pore complex. However, the genetic requirements of *HAS1-TDA1* relocalization and pairing differ from that of previously known relocalized genes, suggesting a potentially novel mechanism. Together, our results demonstrate the underappreciated complexity of the 3D organization of the yeast genome.

2.3 RESULTS

2.3.1 *Hi-C in hybrid yeasts provides a global view of diploid chromosome conformation*

We performed Hi-C on interspecific hybrids between diverged *Saccharomyces* species to obtain the first genome-wide view of chromosome conformation in diploid yeasts. The sequence identity of homologous chromosomes in diploid *S. cerevisiae* precludes observation of

interactions between them using sequencing-based methods. However, divergent *Saccharomyces* species can form stable hybrids (González et al., 2006; Mertens et al., 2015), e.g. between *S. cerevisiae* and *S. paradoxus* (90% nucleotide identity in coding regions (Kellis et al., 2003)) or its more distant relative *S. uvarum* (also known as *S. bayanus* var. *uvarum*; 80% nucleotide identity in coding regions (Kellis et al., 2003)). These interspecific hybrids are sufficiently diverged to allow straightforward sequence-level discrimination of homologs (**Figure 2.2A**) but have maintained nearly complete synteny (Fischer et al., 2000). For comparison, we also analyzed hybrids between *S. cerevisiae* strains Y12 and DBVPG6044, which are much less diverged (~99% nucleotide identity) (Liti et al., 2009). We confirmed the minimal impact of mapping and experimental artifacts by mapping Hi-C data from each individual species or strain (**Figure 2.2B-D**) and mixtures thereof (**Figure 2.1B**, **Figure 2.3**) to the hybrid reference genomes.

The most prominent features of Hi-C data from diploid yeast are the signatures of a Rabl-like orientation (**Figure 2.1C,D**). As in all Hi-C datasets, the contact map exhibits a strong diagonal signal indicating frequent intrachromosomal interactions between adjacent loci. In addition, pericentromeric regions interact frequently with one another, but infrequently with regions far from centromeres, as expected from the clustering of centromeres at the spindle pole body. Telomeric regions also preferentially interact, consistent with their clustering at the nuclear periphery. Finally, the rDNA-carrying chromosomes each behave as two separate chromosomes divided by the nucleolus, with frequent interactions on either side of the rDNA array but not across it.

2.3.2 *Homolog proximity exceeds the effects of the Rabl-like orientation*

In addition to these previously described phenomena, we observed an off-diagonal line of increased interaction suggestive of homolog pairing (**Figure 2.1C**). Homologous loci tend to be closer together than nonhomologous loci in multiple assays, including microscopy (Burgess et al., 1999; Weiner and Kleckner, 1994), recombination efficiency (Burgess and Kleckner, 1999), and chromatin conformation capture (Dekker et al., 2002). Mitotic homolog pairing could be the result of transient pairing between homologous nucleosome-free DNA (Danilowicz et al., 2009; Gladyshev and Kleckner, 2014) or interactions among proteins bound to DNA (Mirkin et al., 2014). However, it has also been suggested that the observation of homolog proximity is an artifact of the Rabl-like orientation or microscopy methods (Lorenz et al., 2003; Mirkin et al., 2014). This debate remains unresolved in part due to the targeted nature of previous studies, wherein each pair of homologous loci is only compared to a limited number of nonhomologous loci.

To systematically investigate whether homolog proximity can be explained by the Rabl-like orientation, we compared our experimental data from *S. cerevisiae* x *S. uvarum* hybrids to simulated data from a volume-exclusion polymer model of the Rabl-like orientation. This model did not explicitly encode homolog pairing (Tjong et al., 2012) and served as a negative control. We quantified homolog proximity by comparing the frequency of each interaction between a pair of homologous loci to the set of nonhomologous interactions involving either locus (**Figure 2.7**). This naïve comparison appears to suggest strong homolog proximity in both experiments (**Figure 2.6A**, left panel), but in fact, the equally strong signal from the polymer model suggests that the apparent signal is a consequence of the Rabl-like orientation. We therefore controlled for the Rabl-like orientation by restricting comparisons to interactions with loci at a similar distance

from the centromere (at a resolution of 32 kb), as previous studies have done (Burgess et al., 1999; Lorenz et al., 2003). Using this approach, we find that the polymer simulation still predicts strong homolog proximity (**Figure 2.6A**, middle panel), suggesting that the long-used approach of comparing homologous interactions to nonhomologous interactions at the same centromeric distance may not fully account for the Rabl-like orientation. Polymer models of the Rabl-like orientation suggest that short chromosomes interact preferentially, due to their dual telomeric tethering at the nuclear periphery and centromeric tethering at the spindle pole (Tjong et al., 2012). Therefore, we further restricted comparisons to loci on chromosome arms of similar length (within 25%). This additional restriction dramatically reduced the signal of homolog proximity for the polymer model, but not for the experimental data (**Figure 2.6A**, right panel).

Comparing homolog proximity across the genome, we noticed extensive interactions between the homologous chromosomes carrying the rDNA arrays (**Figure 2.6B**). To test whether this enrichment for interactions is due to sequence-dependent homolog pairing, we generated a translocation that swapped most of the centromeric half of *S. cerevisiae* chromosome XII with an equivalently sized portion of *S. cerevisiae* chromosome V, thereby moving the rDNA array to *S. cerevisiae* chromosome V. In this translocation-bearing strain, interactions between *S. uvarum* chromosome XII and *S. cerevisiae* chromosome V are enriched instead of *S. cerevisiae* chromosome XII (**Figure 2.6C,D** and **Figure 2.8A,B**), suggesting that homolog proximity of chromosomes carrying the rDNA arrays is due to the presence of the rDNA rather than the particular sequence of the chromosome that carries it. We propose that the rDNA-carrying chromosomes are uniquely positioned within the nucleus due to their tethering at the nucleolus (Duan et al., 2010) (**Figure 2.8C**). This shared tethering would then cause enhanced interactions

between the homologous proximal and distal segments of these chromosomes and inflate the signal for apparent homology-dependent pairing.

Based on these findings, we excluded the rDNA-carrying chromosomes from estimates of homolog proximity. Even with these stringent constraints, we find that the observed interaction between homologous alleles exceeds that predicted based on the Rabl-like orientation (**Figure 2.6E**). Of note, the left arm of chromosome III and the right arm of chromosome IX exhibit particularly strong homolog proximity (**Figure 2.6B**); proximity at chromosome III is possibly due to pairing of the silenced mating-type loci (Miele et al., 2009).

In all hybrids, homolog proximity is substantially greater in saturated cultures approaching stationary phase than in exponential growth (**Figure 2.6E**), consistent with previous observations (Burgess et al., 1999). One explanation for this result is differences in the strength of sequence-dependent homolog pairing between growth conditions, perhaps mediated by differences in nucleosome positioning and DNA-bound proteins. However, this difference could also be a consequence of the reduced cell cycling coupled with loss of homolog proximity during S-phase (Burgess et al., 1999) or smaller nuclear size in cells approaching stationary phase (Guidi et al., 2015). To test whether we also observe cell cycle dependence of homolog proximity, we performed Hi-C on nocodazole-arrested cells, which were previously reported to exhibit reduced homolog proximity (Burgess et al., 1999). We find that nocodazole arrest does not substantially reduce homolog proximity in the diverged hybrid *S. cerevisiae* x *S. uvarum* (**Figure 2.6E**). However, it remains possible that the lack of S-phase cells in saturated cultures contributes to the difference in homolog proximity between exponentially growing and saturated cultures.

We next sought to evaluate whether changes in nuclear size across growth conditions could explain the observed variation in homolog proximity. The nucleus is known to decrease in size in saturated cultures (Guidi et al., 2015), so we created alternate versions of the polymer model of the Rabl-like orientation with proportionally smaller nuclei, at 80% and 64% of the original size. In these models, smaller nuclear size led to decreased, rather than increased, homolog proximity (**Figure 2.6E**). These models suggest that the difference in homolog proximity between saturated and exponentially growing cultures cannot be explained by the effect of differences in nuclear size, and provide additional support for homolog pairing beyond the Rabl-like orientation.

2.3.3 *Relocalization of GAL1 upon galactose induction alters genome conformation*

We also searched our dataset for evidence of highly specific changes in genome conformation at the scale of individual genes. Microscopy studies have revealed inducible genes that relocate to the nuclear periphery upon activation due to association with nuclear pores, e.g. *GAL1* (Brickner et al., 2016; Casolari et al., 2004; Dultz et al., 2016), *INO1* (Brickner and Walter, 2004), *HXK1* (Taddei et al., 2006), *TSA2* (Ahmed et al., 2010), and *HSP104* (Dieppois et al., 2006), which can increase gene expression (Ahmed et al., 2010; Brickner and Walter, 2004; Brickner et al., 2016; Taddei et al., 2006). Although DNA interactions with components of the nuclear pore complex have been identified genome-wide by chromatin immunoprecipitation (Casolari et al., 2004), it remains unclear whether relocalization of specific genes impacts global genome conformation.

We first focused on the galactose metabolism gene *GAL1*. This gene and its neighbors *GAL7* and *GAL10* move upon galactose induction from their location near the spindle pole body to a nuclear pore complex at the nuclear periphery (Casolari et al., 2004; Dultz et al., 2016)

(**Figure 2.9A**). Consistent with this expectation, we found using Hi-C that both *GAL1* loci interacted less with pericentromeric regions upon galactose induction (**Figure 2.9B-D**). Despite previous reports that the homologous *GAL1* loci preferentially interact with each other during galactose induction (Brickner et al., 2016; Zhang and Bai, 2016), we do not see a clear signal for increased pairing (**Figure 2.10**), perhaps because of the high basal interaction frequency between pericentromeric loci or the divergence between *S. cerevisiae* and *S. uvarum*.

2.3.4 *Novel inducible pairing of homologous HAS1-TDA1 loci*

Having established that we could detect the known inducible relocalization of the *GAL1* gene, we looked for other specific changes in genome conformation in the well-studied environmental conditions of galactose induction and growth saturation (approaching stationary phase). Surprisingly, we observed markedly increased interactions between homologous loci surrounding the genes *HAS1* and *TDA1* (subsequently abbreviated as '*HAS1-TDA1* loci') on chromosome XIII under both growth saturation and galactose induction, compared to standard exponential growth in glucose (**Figure 2.11A**). In fact, under inducing conditions this interaction is among the strongest genome-wide, excluding pericentromeric and subtelomeric regions (top interaction out of over 83,000; **Figure 2.12**). No canonical galactose-induced genes are in or near this region. Nevertheless, this inducible homolog proximity appears to be evolutionarily conserved, as it occurs in all three tested interspecific hybrids, at least in saturated culture (**Figure 2.11A-C**; galactose not tested in all hybrids).

To explore whether this pairing depends on the presence of specific sequences, we created various deletions of the *S. cerevisiae* copy of the region, ranging from a 20 kb region from *NGL2* through *YMR295C* (**Figure 2.12A**) to a single 1 kb intergenic region containing the promoters for *HAS1* and *TDA1* (*HAS1pr-TDA1pr*; **Figure 2.11D**). Every deletion that included

this intergenic region reduced the interaction frequency between *HAS1-TDA1* homologs in saturated growth conditions back to uninduced levels, indicating that this inducible pairing is sequence-dependent (**Figure 2.11E** and **Figure 2.13**). In contrast, deletion of the *HAS1* coding sequence had minimal impact, which shows that the deletion construct itself did not impede inducible pairing (**Figure 2.11E** and **Figure 2.13**). To test whether the *HAS1pr-TDA1pr* region is sufficient to produce inducible pairing, we moved the *S. cerevisiae* copy of this region to the left arm of *S. cerevisiae* chromosome XIV. The ectopic *HAS1pr-TDA1pr* allele exhibited inducible interactions with the *S. uvarum HAS1pr-TDA1pr*, though not to the same extent as the endogenous allele (**Figure 2.14**). The diminished extent of inducible pairing may reflect the contribution of chromosomal homolog pairing, which would be disrupted in the ectopic location, or of additional regions that are not sufficient to produce pairing on their own. To verify whether this pairing occurs in homozygous *S. cerevisiae* diploids in addition to diverged hybrids, we labeled both *HAS1-TDA1* loci with integrated LacO arrays targeted by LacI-GFP and measured the distance between them in a population of cells by confocal microscopy (**Figure 2.11F**). Consistent with our Hi-C data, the *HAS1-TDA1* homologs were closer together in galactose-induced and saturated cultures than in glucose (**Figure 2.11G,H**).

2.3.5 *Nuclear pores play a role in HAS1-TDA1 homolog pairing*

Based on previous studies of relocalized genes (Brickner et al., 2012, 2016), we hypothesized that pairing between the homologous *HAS1-TDA1* loci might be mediated by interactions of both alleles with nuclear pores (**Figure 2.16**). Therefore, we tested whether the *HAS1-TDA1* loci are relocalized to the nuclear periphery in a condition-dependent manner. We tagged the *HAS1-TDA1* locus with a LacO array as before and counted the proportion of cells in which *HAS1-TDA1* colocalized with the mCherry-labeled nuclear membrane, in haploid *S.*

cerevisiae. Indeed, the *HAS1-TDA1* locus shifted to the nuclear periphery upon galactose induction and in saturated culture conditions (**Figure 2.15A**). To confirm whether this inducible reorganization was dependent on association with nuclear pores, we repeated our analysis in strains with deletions of nuclear pore components *NUP2* or *NUP100*, or pore-associated protein *MLP2* (**Figure 2.15A**). As in other cases of gene relocalization, Nup2 but not Nup100 was required for peripheral localization of the *HAS1-TDA1* locus. However, unlike other relocalized genes (Ahmed et al., 2010; Brickner et al., 2016; Luthra et al., 2007), *HAS1-TDA1* locus relocalization did not require Mlp2, suggesting that the *HAS1-TDA1* locus may interact with nuclear pores via a distinct mechanism. We performed these initial analyses in haploids, to facilitate deletion of nuclear pore components, but pairing cannot occur in haploids. Thus, we confirmed that the *HAS1-TDA1* loci are peripherally relocalized in diploids, by reanalyzing the images we used to measure distances between *HAS1-TDA1* alleles in diploid cells (**Figure 2.15B**). We then asked whether pairing of the *HAS1-TDA1* alleles only occurs at the periphery, by determining the proportion of cells in each category of peripheral localization that have paired *HAS1-TDA1* alleles (**Figure 2.15C**). In fact, *HAS1-TDA1* alleles can remain paired in galactose in the nucleoplasm. However, this need not imply that the nuclear pores do not play a role in pairing. Previous studies of gene relocalization to nuclear pores have reported that the cell cycle affects when genes relocalize to the nuclear periphery; namely, genes tend to move to the nucleoplasm during S-phase (Brickner and Brickner, 2010). Alleles can remain paired in the nucleoplasm, but cannot actively pair during S-phase (Brickner et al., 2012). In agreement with these studies, the peripheral localization of *HAS1-TDA1* loci that we observe also exhibits cell cycle dependence (**Figure 2.15D,E**), which may explain the presence of pairing in the nucleoplasm.

To test whether the nuclear pore complex is required for pairing as well as relocalization of the *HAS1-TDA1* loci, we performed Hi-C on an *S. cerevisiae* x *S. uvarum* hybrid strain with a homozygous deletion of *NUP2*. In this strain, *HAS1-TDA1* pairing was not observed in galactose, as expected, but still occurred at full strength in saturated growth (**Figure 2.15F**). These data indicate that Nup2 is required for *HAS1-TDA1* homolog pairing in galactose but not in saturated culture, suggesting distinct and/or additional mechanisms of pairing. To test biochemically whether the *HAS1-TDA1* loci interact with nuclear pores, we performed chromatin immunoprecipitation (ChIP) sequencing on the nuclear basket protein Nup60, which unlike the dynamic Nup2 cannot dissociate from the nuclear pore complex, tagged with the tandem affinity purification (TAP) tag (Ghaemmaghami et al., 2003), in haploid *S. cerevisiae* grown in either glucose or galactose. As expected, we observed a clear enrichment of the galactose metabolism gene cluster *GAL1-GAL10-GAL7* in the immunoprecipitated DNA from cells grown in galactose, compared to those grown in glucose (**Figure 2.15G**). In contrast, we observed little if any such enrichment of the *HAS1-TDA1* locus (**Figure 2.15H**). We also performed qPCR on the same ChIP DNA, which gave the same results (**Figure 2.15I** and **Figure 2.17**). Together, these data suggest that although *HAS1-TDA1* locus relocalization and pairing requires Nup2 in galactose, *HAS1-TDA1* differs from *GAL1* in how it interacts with the nuclear pore complex and may pair via alternative mechanisms as well, particularly in saturated growth.

2.3.6 *Transcription at the HAS1-TDA1 locus*

Nuclear pore association is thought to play a role in transcriptional regulation, generally but not always leading to greater or faster activation (Akhtar and Gasser, 2007; Taddei, 2007; Taddei et al., 2010). To explore how nuclear pore association might affect transcription at the *HAS1-TDA1* locus, we performed RNA sequencing on haploid *S. cerevisiae* grown in glucose,

galactose, and saturated growth conditions. *HAS1* is downregulated in both pairing conditions (**Figure 2.18A**), whereas *TDA1* is weakly upregulated in both conditions (**Figure 2.18B**). However, transcriptional changes at *HAS1* and *TDA1* are relatively unremarkable; in both galactose and saturated growth, dozens to hundreds of genes are more strongly up- or downregulated than *HAS1* or *TDA1* (**Figure 2.18C,D**). For example, *GAL1* is upregulated nearly 1000-fold in galactose (**Figure 2.18C**). This suggests that nuclear pore association is not solely a function of strong transcriptional activation.

2.4 DISCUSSION

Homologous chromosomes pair along their lengths leading up to and during meiosis, but may recognize each other and preferentially interact even in normal mitotic growth, perhaps to facilitate homology-directed repair or prepare for meiosis under stressful conditions. However, whether and to what extent this mitotic homolog pairing occurs has remained controversial, in part due to the lack of genome-wide data and the apparent homolog pairing caused by the Rabl-like orientation. We performed Hi-C in diverged hybrid diploid yeast, which allowed us to resolve homologous chromosomes (**Figure 2.1**) and thus infer homolog pairing strength on a genome-wide basis (**Figure 2.6B**). After using a polymer model of the Rabl-like orientation to calibrate our estimates, we find that even in hybrid diploids with homologs diverged to less than 80% nucleotide identity, homologous chromosomes do interact preferentially during mitotic growth, albeit subtly (**Figure 2.6E**). It would be interesting to compare the strength of pairing across hybrids with varying levels of divergence; however, our homolog pairing analysis requires filtering and stratifying genomic regions and thus may not be directly comparable across different reference genomes. Our data do not necessarily imply end-to-end chromosome alignment as occurs in meiosis. Instead, our data indicates an increased frequency of contact

between homologous loci above the expectation based on random collisions, with substantial variation across the genome. Nevertheless, that such distant homology is sufficient for at least some homolog interactions is perhaps remarkable, and may hint at the role of DNA-bound proteins, which are more conserved than DNA, in mediating the interactions. Homolog pairing strength also depends on both growth conditions and genomic location, sometimes jointly: the homologous *HAS1-TDA1* loci on chromosome XIII pairs during saturated growth and galactose induction, but not exponential growth in glucose (**Figure 2.11**). This region is not remarkably conserved, suggesting that homolog pairing is at least partly due to specific interactions mediated by proteins, rather than direct DNA-DNA homology interactions (Danilowicz et al., 2009; Gladyshev and Kleckner, 2014).

In all tested hybrids, the *HAS1-TDA1* locus exhibits surprisingly strong homolog proximity (**Figure 2.11A-C**; **Figure 2.12B**). How does *HAS1-TDA1* pairing occur, and why? The nuclear pore component Nup2 seems to play a role, though not exclusively, in mediating pairing. The *HAS1-TDA1* locus moves to the nuclear periphery under pairing conditions, and both this relocalization and pairing are Nup2-dependent in galactose (**Figure 2.15A,F**). However, Nup2 is not required for *HAS1-TDA1* pairing in saturated growth (**Figure 2.15F**). Together, the Nup2-independence of *HAS1-TDA1* pairing in saturated growth, the Mlp2-independence of *HAS1-TDA1* peripheral localization, and the lack of *HAS1-TDA1* enrichment in Nup60 ChIP suggest that *HAS1-TDA1* may interact with nuclear pores by a mechanism distinct from the previously studied *GAL1* and *INO1*, and possibly by different mechanisms in galactose and in saturated culture. However, more experiments are needed to fully elucidate the role and mechanism of the nuclear pore complex in *HAS1-TDA1* homolog pairing. Regardless of the molecular mechanism, nuclear pore interactions may confine the *HAS1-TDA1* alleles to the

relatively small space near the nuclear periphery, thus speeding up the rate at which they randomly contact each other (**Figure 2.16**). Once in physical proximity, additional mechanisms such as protein-protein interactions between DNA-binding proteins could prolong the duration of contact, even after the alleles are no longer at a nuclear pore. Indeed, the presence of paired *HAS1-TDA1* alleles in the nucleoplasm suggests that such nuclear pore-independent pairing mechanisms may act at *HAS1-TDA1* (**Figure 2.15C**). Interestingly, a recent study showed that Nup2 is involved in meiotic homolog pairing (Chu et al., 2017), suggesting that Nup2 may more generally play a role in homolog pairing.

Why do the homologous *HAS1-TDA1* alleles pair and relocalize to the periphery, and why does this interaction appear to be unique? Many genes associate with nuclear pores and relocalize to the nuclear periphery upon activation, including *GAL1*, but we do not observe strong pairing of *GAL1* homologs. It is possible that Hi-C may be failing to capture pairing at *GAL1* due to its pericentromeric location, but these data may also reflect particularly strong pairing at the *HAS1-TDA1* loci. Divergence between *S. cerevisiae* and *S. uvarum*, particularly in their galactose metabolism pathways (Roop et al., 2016), may also contribute to the lack of pairing at the *GAL1* locus. Lack of *GAL1* pairing need not correspond to lack of peripheral localization (Brickner et al., 2016); in our hybrid, the *S. cerevisiae* and *S. uvarum GAL1*-binding proteins may be able to each interact with the nuclear pores but not with each other.

Which gene is driving *HAS1-TDA1* homolog pairing, *HAS1* or *TDA1*, or both? Given the association between nuclear pore interactions and transcriptional activation, transcriptional changes in growth conditions that induce pairing may provide a clue. The genes at the *HAS1-TDA1* locus, *HAS1* and *TDA1*, demonstrate opposing changes in gene expression in galactose and saturated culture: *HAS1* is downregulated, whereas *TDA1* is upregulated. The upregulation

of *TDA1*, though not particularly strong in magnitude (**Figure 2.18C,D**) may be of functional importance. Tda1 is a kinase required for phosphorylation of Hxk2, the primary hexokinase in yeast (Kaps et al., 2015; Kettner et al., 2012). Unphosphorylated Hxk2 can interact with Mig1 to repress various alternative carbon source metabolism genes in the presence of glucose. Phosphorylation of Hxk2 by Tda1 in low glucose conditions prevents its interaction with Mig1 and thus leads to release from glucose repression. While we have not yet tested whether disrupting peripheral localization would affect *TDA1* or *HAS1* transcription, we hypothesize that nuclear pore interaction may aid the upregulation of *TDA1* in response to low glucose concentrations, perhaps by facilitating efficient transcription or mRNA export from the nucleus.

Other questions remain about the mechanism and functional implications of *HAS1-TDA1* pairing and peripheral relocalization. Increased transcription may itself contribute to localization at nuclear pores via interactions between nascent mRNA and mRNA processing and export factors at nuclear pores (Akhtar and Gasser, 2007), and may be involved in establishment of *HAS1-TDA1* pairing as it is for *GAL1* (Brickner et al., 2016). However, given the abundance of other genes with similar or greater changes in transcription that do not pair (**Figure 2.18**), transcription alone likely cannot explain our data. For some genes, nuclear pore interactions mediate rapid reactivation in a phenomenon termed epigenetic transcriptional memory (D'Urso and Brickner, 2017; D'Urso et al., 2016; Light et al., 2010); it is also possible the nuclear pore interactions with *HAS1-TDA1* may be involved in epigenetic transcriptional memory. The pairing of the *HAS1-TDA1* alleles may also serve a distinct function, potentially including *trans* gene regulation like at *GAL1* (Zhang and Bai, 2016), but further experiments are needed to test this possibility.

The principles and functional implications of genome conformation remain open questions. Although the budding yeast *S. cerevisiae* is thought to have a simple genome organization, it serves as a versatile and relevant model system amenable to integrating multiple approaches to studying genome conformation, including Hi-C, polymer simulations, live-cell imaging, and genetic perturbations. While yeast nuclear organization may differ from that of other eukaryotes in some ways, our findings may nevertheless be applicable to other organisms: recent studies in the fruit fly *Drosophila* have provided evidence for the generality of the role of nuclear pores in transcriptional regulation first observed in yeast (Pascual-Garcia et al., 2017). Our study illustrates both the utility of combining orthogonal methodologies and that we have much more to learn about genome organization, even in the simple budding yeast.

2.5 MATERIALS AND METHODS

2.5.1 *Strain construction*

All yeast strains used in this study are listed in **Supplementary file 1**. All primers used in this study, including those used for generation and validation of strains, are listed in **Supplementary file 2**.

Hybrid strains were created by mating haploid strains and then performing auxotrophic selection.

The ScV-ScXII translocation *S. cerevisiae* x *S. uvarum* strain was generated by first creating the translocation in the haploid *S. cerevisiae* strain BY4742, followed by mating with haploid *S. uvarum*. A cassette containing *hphMX* followed by the first half of *URA3*, an artificial intron, and a *lox71* site was amplified from pBAR3 (Levy et al., 2015) and integrated into the intergenic region between *YLR150W* and *YLR151C*. A second cassette containing a *lox66* site, an artificial intron, the second half of *URA3*, and *natMX* was amplified from pBAR2-natMX

(pBAR2 (Levy et al., 2015) with *natMX* in place of *kanMX*) and integrated into the intergenic region between *YER151C* and *YER152C*. The translocation was induced by transforming the resulting strain with the galactose-inducible Cre plasmid pSH47-kanMX (pSH47 (Güldener et al., 1996) with *kanMX* in place of *URA3*), and then inducing Cre recombination by plating on YP + galactose medium. Successful translocation strains were selected by growing in medium lacking uracil, and verified by PCR across the translocation junctions. This *S. cerevisiae* strain was then mated with *S. uvarum* strain ILY376.

Heterozygous deletion strains were made in *S. cerevisiae* x *S. uvarum* hybrids, by replacing regions of interest with the *hphMX* cassette. Homozygous deletion strains were made by making deletions in haploids and then mating the haploid strains. Strains were verified by PCR across each deletion junction.

The knock-in strain was made by integrating the *HAS1pr-TDA1pr* region followed by the *natMX* cassette into the region between *YNL266W* and *YNL267W* (*PIK1*) on *S. cerevisiae* chromosome XIV in the *S. cerevisiae* x *S. uvarum* hybrid YMD3269 (*HAS1pr-TDA1pr* deletion).

Plasmids pAFS144 (Straight et al., 1996), p5LacIGFP (Randise-Hinchliff et al., 2016), pER04 (Randise-Hinchliff et al., 2016) and pFA6a-kanMX6 (Longtine et al., 1998) have been described. To tag *HAS1-TDA1* with the LacO array, 1 kb downstream of the *HAS1* ORF was PCR amplified and TOPO cloned to create pCR2.1-*HAS1*_3'UTR. Plasmid p6LacO128-*HAS1* was made by inserting *HAS1* from pCR2.1- *HAS1*_3'UTR into p6LacO128 (Brickner and Walter, 2004).

2.5.2 *Hi-C*

Cells were grown overnight, shaking at 30˚C (room temperature for *S. uvarum*) in YPD medium (1% yeast extract, 2% peptone, 2% dextrose), YP + raffinose (2%), or YP + galactose

(2%). For saturated culture samples, they were crosslinked at this point by resuspension and incubation in 1% formaldehyde in PBS for 20 minutes at room temperature. Crosslinking was quenched by addition of 1% w/v solid glycine, followed by incubation for 20 minutes and a PBS wash. For all other experiments, fresh cultures were inoculated to OD_{600} = 0.1 in appropriate medium and grown to OD_{600} = 0.6-0.8. Exponential growth samples were crosslinked at this point, while for nocodazole-arrested samples, cultures were supplemented with 15μg/mL nocodazole and grown at 30˚C for 2 hours following addition of drug prior to crosslinking. Arrested cultures were checked by flow cytometry. For mixture controls, samples were mixed prior to crosslinking.

Hi-C libraries were created as described (Burton et al., 2014) with the exceptions that the restriction endonuclease Sau3AI or HindIII was used to digest the chromatin and the Kapa Hyper Prep kit (Kapa Biosystems, Wilmington, MA) was used to create the Illumina library instead of the Illumina TruSeq kit. Libraries were pooled and sequenced on an Illumina NextSeq 500 (Illumina, San Diego, CA), with 2x80 bp reads for interspecific hybrids and 2x150 bp reads for intraspecific *S. cerevisiae* hybrids. Hi-C libraries were similar across the two restriction enzymes and biological replicates (**Figure 2.4**). All Hi-C libraries are listed in **Supplementary file 3**.

2.5.3 *Reference genomes*

The *S. cerevisiae* references and annotations were downloaded from the Saccharomyces Genome Database (version R64.2.1). The *S. paradoxus* and *S. uvarum* references and annotations were downloaded from saccharomycessensustricto.org (Scannell et al., 2011) but modified to correct misassemblies evident based on synteny and Hi-C data (**Figure 2.5**). *S. paradoxus* chromosome IV was rearranged so bases 1-943,469 were followed by 1,029,253-1,193,028, then 1,027,718-1,029,252, then 943,470-1,027,717 in reverse order, followed by the

remainder of the chromosome. *S. uvarum* chromosome III was rearranged so bases 219,500 onward were placed at the beginning (left end) of the chromosome, followed by the first 219,399 bases, and then new sequence determined by Sanger sequencing with primers CATTCCCATTTGTTGATTCCTG and GGATTCTATTGTTGCTAAAGGC : TAATAAGGAAGAACTGCTTATTCTTAATTATTTCTACCTACTAAACTAACTAATTATC AACAAATATCATCTATTTAATAGTATATCATCACATGCGGTGTAAGAGGATGACATA AAGATTGAGAAACAGTCATCCAGTCTAATGGAAGCTCAAATGCAAGGGCTGATAAT GTAATAGGATAATGAATGACAACGTATAAAAGGAAAGAAGATAAAGCAATATTATT TTGTAGAATTATCGATTCCCTTTTGTGGATCCCTATATCCTCGAGGAGAA. *S. uvarum* chromosomes X and XII were also swapped, based on homology to *S. cerevisiae*. The *S. cerevisiae* Y12 and DBVPG6044 strain references were sequenced to 145- and 315-fold coverage using the PacBio (Pacific Biosciences, Menlo Park, CA) single-molecule, real-time (SMRT) sequencing platform with P6-C2 chemistry. Each genome was assembled with FALCON (Chin et al., 2016), version June 30, 2015 hash: cee6a58, and polished with Quiver (Chin et al., 2013) version 1.1.0 to generate chromosome-length contigs (with the exclusion of chromosome XII, which was split at the rDNA array, and Y12 chromosome XIV, which was split into one large and one small contig). To call centromeres in *S. paradoxus*, we searched the region on each chromosome between the genes homologous to those nearest the centromeres in *S. cerevisiae* (e.g. *YEL001C* and *YER001W* on chromosome V) for the sequence motif N_2TCAC(A/G)TG$N_{95\text{-}100}$CCGAAN_6 (based on an alignment of *S. uvarum*, *S. mikatae*, and *S. kudriavzevii* centromeres (Scannell et al., 2011)) or its reverse complement. When this motif was absent (chromosomes VII and VIII), we called the centromere as the middle 120 bp of the

region. To call centromeres in Y12 and DBVPG6044, we mapped the *S. cerevisiae* S288C centromere sequences to the new references.

2.5.4 *Theoretical mappability analysis*

Simulated reads for each hybrid genome (as in experimental data, 80 bp for interspecific hybrids and 150 bp for the intraspecific *S. cerevisiae* hybrid) were generated by taking sequences of the read length at 10 bp intervals. These reads were then remapped to the hybrid genome using bowtie2 (Langmead and Salzberg, 2012) with the --very-sensitive parameter set. The proportion of reads that mapped with mapping quality ≥ 30 to the correct location was then calculated.

2.5.5 *Hi-C data analysis*

Sequencing reads were first pre-processed using cutadapt (Martin, 2011): reads were quality-trimmed (option -q 20), trimmed of adapter sequences, and then trimmed up to the ligation junction (if present), excluding any read pairs in which either read was shorter than 20 bp after trimming (option -m 20). The two reads in each read pair were then mapped separately using bowtie2 (Langmead and Salzberg, 2012) with the --very-sensitive parameter set. For interspecific hybrids, reads were mapped to a combined reference containing both species references, where if secondary mappings were present the best alignment must have a score ≥ 10 greater than the next best alignment. For intraspecific *S. cerevisiae* hybrids, reads were mapped separately to both strain references, keeping only read pairs in which both reads mapped to both references—perfectly to one reference and with ≥ 2 mutations including ≥ 1 substitution to the other. PCR duplicates (with identical fragment start and end positions) were removed, as were read pairs mapping within 1 kb of each other or in the same restriction fragment, which represent either unligated or invalid ligation products. The genome was then binned into 32 kb fragments

(except the last fragment of each chromosome), and the number of read pairs mapping to each 32 kb genomic bin was counted based on the position of the restriction sites that were ligated together. Due to gaps in the reference genomes of *S. uvarum*, some repetitive sequences were only represented once and therefore artifactually mapped uniquely; therefore, reads mapping to annotated repetitive sequences were masked from further analysis. Similarly, gaps in the *S. paradoxus* reference led to mismapping of reads to the corresponding *S. cerevisiae* sequence; therefore, for *S. cerevisiae* x *S. paradoxus* libraries we masked regions in the *S. cerevisiae* genome where > 1 read from a *S. paradoxus* Hi-C library mapped, and vice versa. We took a similar approach to mask regions of the Y12 and DBVPG6044 references that were prone to mismapping, as estimated by haploid Y12 and DBVPG6044 Hi-C libraries. For knock-in experiments, the *HAS1pr-TDA1pr* region was masked to account for its altered genomic location. The resulting matrices were then normalized by excluding the diagonal (interactions within the same genomic bin), filtering out rows/columns with an average of less than 1 read per bin, and then multiplying each entry by the total number of read pairs divided by the column and row sums.

2.5.6 *Polymer model*

The volume-exclusion polymer model of the Rabl-like orientation was a modified version of the Tjong et al. tethering model (Tjong et al., 2012). Briefly, beads representing segments of the genome are randomly positioned and then adjusted until constraints (e.g. consecutive beads must be adjacent, and no two beads can occupy the same space) are met. The model was extended from 16 chromosomes to 32, with the lengths of the *S. cerevisiae* and *S. uvarum* chromosomes. The parameters for nuclear size, centromeric constraint position and size, telomeric constraint at the nuclear periphery, and nucleolar position and size were scaled by a

factor of 1.25 to reflect the roughly doubled volume of diploid nuclei (cell volume correlates with ploidy (Mortimer, 1958), and nuclear volume correlates with cell volume (Jorgensen et al., 2007)). To test the effect of smaller nuclei, all parameters were scaled by a factor of 0.8 or 0.64 from this initial diploid model. For each model, the modeling procedure was repeated 20,000 times to create a population of structures. From this population, we simulated Hi-C data by calling all beads within 45 nm of each other as contacting each other, and then counting the number of contacts between each pair of 32 kb bins. The resulting matrix of counts was normalized using the same pipeline as the experimental Hi-C data.

2.5.7 *Homolog proximity analysis*

In order to assess homolog proximity genome-wide, we first determined which bins represented interactions between homologous sequences, and then compared the normalized interaction frequencies in those bins compared to a set of "comparable" nonhomologous bins.

In the interspecific hybrids, we determined homology by counting the number of starts or ends of one-to-one homologous gene annotations falling into each bin. Genes whose "SGD" and "BLAST" gene annotations differed were ignored. To find homologous interaction bins for genomes 1 and 2, for each bin of genome 1 we considered the bin in genome 2 where the most homologous gene ends fell to be homologous.

In the intraspecific hybrids where inter-strain mapping was much more reliable, we simulated 150 bp reads from the Y12 genome at 10 bp intervals, then mapped them to the DBVPG6044 reference. Here, for each bin of genome 1 we considered the bin in genome 2 where the most reads mapped with MAPQ $\geq$ 30 to be homologous.

To eliminate minor "homology" arising from repetitive sequences (e.g. telomeres), we excluded isolated homologous interaction bins lacking any other homologous interaction bins within 2 bins. To fully exclude homologous interactions from our estimates of nonhomologous interactions, any interaction bins within 2 bins of homologous interaction bins were excluded from analyses.

After determining homologous bins, we compared each homologous bin to other intergenome interactions (i.e. between chromosomes from different species/strains) involving one of the two genomic bins involved in the homologous interaction. To control for the effects of the Rabl-like orientation, we further filtered the nonhomologous interaction bins for those in which the centromeric distance (in units of 32 kb bins) was equivalent, and then for those in which the chromosome arm lengths of the two loci were within 25% of each other (in units of 32 kb bins). We also considered exclusion of the rDNA carrying chromosome XIIs as well as the centromeric bins, for which we could not fully control chromosome arm lengths. In all cases, we only considered homologous bins with at least two comparable nonhomologous bins.

To estimate genome-wide homolog proximity, we compared the sum of normalized interaction frequencies across the homologous bins to those of an equal number of randomly chosen nonhomologous bins, one comparable to each homologous bin, with replacement. We repeated this 10,000 times to obtain a distribution of genomic homolog proximity.

To obtain a view of homolog proximity strength across the genome, we compared the normalized interaction frequency in each homologous bin to the median of that in the similar nonhomologous bins, and then plotted the ratio of homologous/nonhomologous across the *S. cerevisiae* genome.

2.5.8 *Confocal microscopy*

Gene positioning at the nuclear periphery and inter-allelic clustering were determined as described previously (Brickner and Walter, 2004; Brickner et al., 2016; Egecioglu et al., 2014). Briefly, cells bearing an array of 128 Lac operators integrated downstream of the *HAS1* coding sequence and expressing both the ER04 mCherry membrane marker (Egecioglu et al., 2014) and the GFP-LacI (Robinett et al., 1996) were imaged on a Leica SP5 line-scanning confocal microscope (Leica Microsystems, Wetzlar, Germany).

Cultures were grown in synthetic minimal media with 2% glucose or 2% galactose overnight at 30˚C with constant shaking and harvested in log phase (OD_{600} < 0.5) or late log/stationary phase (OD_{600} > 1.0). Unless noted, cultures were grown in the designated media overnight prior to imaging. Cultures were concentrated by brief centrifugation, and then 1 µl was spotted onto a microscope slide for visualization.

For all experiments, cells were illuminated at 10-15% power with 488 nm and 561 nm using argon and diode pumped solid state lasers, respectively. Stacked images of ~150 µm x 150 µm fields were collected; ~15-20 *z*-slices of 0.34 µm thickness each. The optical thickness of the slices is ~0.73 µm. The *z*-slice in which the green dot(s) is most focused and bright is selected for analysis (**Figure 2.11F**).

For peripheral localization experiments, cells in which the center of the dot colocalizes with the nuclear envelope, as measured by mCherry fluorescence, are scored as peripheral. All other cells are scored as nucleoplasmic. Cells in which the dot was at the top or bottom of the nucleus were excluded. Each experiment was performed three times, counting ~30 cells per replicate. The percent of cells scored as peripheral was averaged and the standard error of the mean was calculated. Student's *t*-test was used to compare these distributions.

To monitor inter-allelic clustering of the *HAS1-TDA1* locus, haploid strains bearing the LacO array integrated downstream of *HAS1* were mated to create a diploid strain. These strains were imaged as above and, in cells in which the two alleles were either in the same *z*-slice or adjacent *z*-slices, the distance between the centers of the dots was measured using LAS AF software. Cells in which the two dots were not in the same or adjacent *z*-slices, or cells in which the two dots were unresolvable, were excluded. For each experiment, 100 cells were measured and both the distribution of distances among 0.15 µm bins and the fraction of cells in which the two alleles were < 0.55 µm was calculated. To compare distributions, the Wilcoxon Rank Sum test was used. To compare the fraction of the cells in which the two alleles were < 0.55 µm, Fisher's exact test was used.

2.5.9 *Chromatin immunoprecipitation*

Chromatin immunoprecipitation was performed as in (Egecioglu et al., 2014), with modifications. Nup60-TAP yeast (Ghaemmaghami et al., 2003) were grown overnight, diluted to OD_{600} = 0.125-0.15 in 50 ml medium, then grown to OD_{600} = 0.75-0.85 at 30˚C in either YPD (for glucose samples) or YP + 2% galactose (for galactose samples), then crosslinked with 1% formaldehyde (v/v) for 5 minutes at room temperature. Crosslinked cells were quenched in 150 mM glycine, washed twice in Tris-buffered saline, and then stored at -80˚C.

Crosslinked cell pellets were resuspended in 500 µl lysis buffer with 1x cOmplete Protease Inhibitor tablet (Roche, Basel, Switzerland), and then 700 µl of 500 µm acid-washed glass beads were added. The cells were vortexed for 12 minutes total, in cycles of 2 minutes shaking and 2 minutes resting on ice. The lysate was pelleted and resuspended in fresh lysis buffer, and then sonicated for 3x10 min runs on a Diagenode Bioruptor (Diagenode, Liège,

Belgium), on high power with cycles of 30 seconds on and 30 seconds off, to an average of ~300 bp. The sonicate was cleared by centrifugation, and the resulting supernatant was split into an input aliquot (1/20 of IP volume) and two halves for the IP and mock-IP (BSA instead of antibody). For each sample, 5 µg of anti-TAP antibody (Thermo Fisher Scientific, Waltham, MA; #CAB1001, Lot #RL240352) was used with 10 µl Dynabeads Protein A (Thermo). The antibody or BSA was incubated with pre-washed beads for 2 hours and then washed twice in fresh lysis buffer before being added to the lysate and then incubated overnight. After washing the beads four times in lysis buffer and eluting, the eluate was reverse crosslinked and then treated with RNase A and Proteinase K. DNA was purified using Zymo ChIP DNA Clean & Concentrator (Zymo Research, Irvine, CA) and eluted in 30 µl water.

IP/mock-IP samples were diluted 1:8 and inputs were diluted 1:320, and then 4 µl were used in each 10 µl qPCR reaction. qPCRs were performed in triplicate in 384-well plates on a ViiA7 (Applied Biosystems, Foster City, CA), with Kapa Robust 2G Hot Start 2x master mix (thermocycling as recommended, with 20 second extension for 40 cycles) and 0.2x SYBR Green I dye. CT values were calculated and normalized to a genomic DNA standard curve using the ViiA7 software. IP/input ratios were normalized to those for the negative control *PRM1*.

ChIP-seq libraries were prepared using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences, Ann Arbor, MI), from equal volume pools of the biological replicates, either 1 ng total from input samples or 21 µl of total IP sample. Input and IP libraries were amplified for 6 and 7 cycles, respectively. Libraries were sequenced to ~5-6 million read pairs using 2x75 bp reads on an Illumina MiSeq.

2.5.10 *ChIP-seq analysis*

Sequencing reads were first pre-processed using cutadapt (Martin, 2011): reads were quality-trimmed (option -q 20), trimmed of adapter sequences, excluding any read pairs in which either read was shorter than 28 bp after trimming (option -m 28). The two reads in each read pair were then mapped jointly to the sacCer3 *S. cerevisiae* reference using bowtie2 (Langmead and Salzberg, 2012) with the --very-sensitive parameter set, requiring the two reads to be within 2000 bp of each other (option -X 2000). Fragments (read pairs) in which both reads had a mapping quality score of at least 30 were then deduplicated by fragment start and end positions and then aggregated into a coverage track using bedtools (Quinlan and Hall, 2010). Genome browser tracks were generated using the UCSC Genome Browser (Kent et al., 2002).

2.5.11 *RNA sequencing*

BY4741 yeast were grown overnight in YPD (for exponential growth and saturated samples) or YP + 2% galactose (for galactose samples) at 30˚C, then diluted to OD_{600} = 0.1-0.125 in 50 ml medium, and then grown to OD_{600} = 0.5-0.6, pelleted and stored at -80˚C. RNA was purified using acid phenol extraction, and then treated with the DNA-*free* DNase kit (Thermo). Illumina libraries were then prepared using the TruSeq RNA Library Prep Kit v2 (Illumina), with 6 cycles of amplification. Libraries were sequenced to ~15-20 million read pairs using 2x75 bp reads on an Illumina NextSeq 500.

2.5.12 *RNA-seq analysis*

Sequencing reads were first pre-processed using cutadapt (Martin, 2011): reads were quality-trimmed (option -q 20), trimmed of adapter sequences, excluding any read pairs in which either read was shorter than 28 bp after trimming (option -m 28). The two reads in each read pair

were then mapped jointly to the sacCer3 *S. cerevisiae* reference using bowtie2 (Langmead and Salzberg, 2012) with the --very-sensitive parameter set, requiring the two reads to be within 500 bp of each other (option -X 500). Fragments (read pairs) in which both reads had a mapping quality score of at least 30 were overlapped with annotated genes using HTSeq (Anders et al., 2015). Global fold-change analyses were performed using DESeq2 (Love et al., 2014a).

2.5.13 *Code availability*

Code for all bioinformatic analyses is available at https://github.com/shendurelab/HybridYeastHiC (Kim, 2017).

2.5.14 *Data availability*

GEO accession number: GSE88952

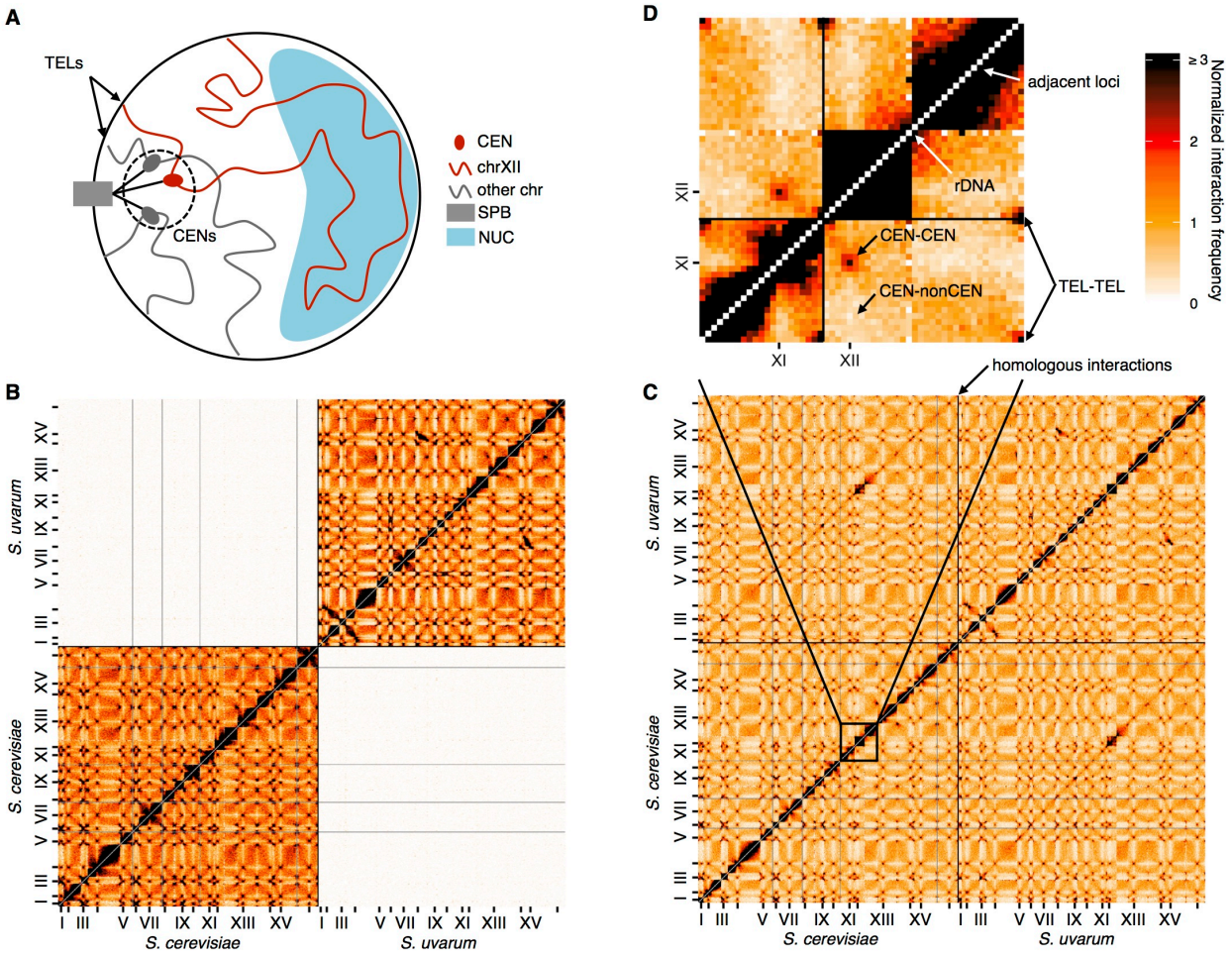

Figure 2.1. Diverged hybrids provide a genome-wide view of diploid chromosome conformation.

(A) Schematic of the Rabl-like orientation. CEN, centromere; SPB, spindle pole body; TEL, telomere; NUC, nucleolus. (B) Hi-C contact map for saturated *S. cerevisiae* and *S. uvarum* mixture control, at 32 kb resolution. Each axis represents the *S. cerevisiae* genome followed by the *S. uvarum* genome in syntenic order, separated by a black line. Ticks indicate centromeres. Odd-numbered centromeres are labeled. Rows and columns with insufficient data are colored grey. (C) Hi-C contact map for saturated *S. cerevisiae* x *S. uvarum* hybrid, as in (B). (D) The portion of the map outlined in black in (C), is enlarged with annotated features of the Rabl-like orientation.

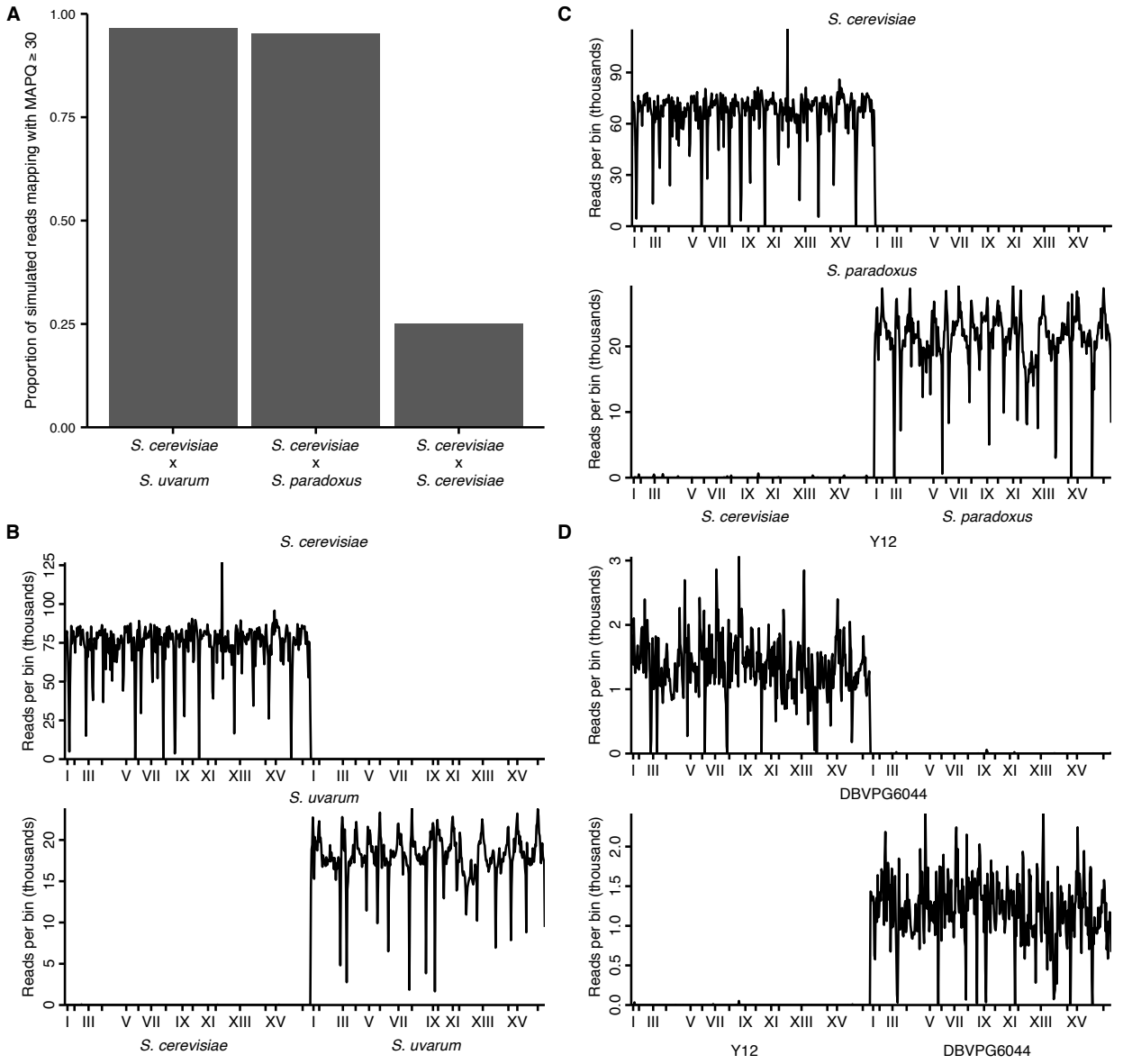

Figure 2.2. Mappability of hybrid yeast genomes.

(A) Proportion of simulated reads from each hybrid (80 bp for interspecific and 150 bp for intraspecific hybrids, in 10 bp windows across the reference genome) that can be remapped correctly to the reference with a mapping quality score (MAPQ) of at least 30. (B) Number of reads (in thousands) from separate *S. cerevisiae* (left) or *S. uvarum* (right) Hi-C libraries mapping to each 32 kb genomic bin in the *S. cerevisiae* x *S. uvarum* hybrid reference genome. *x* ticks indicate centromeres; odd-numbered centromeres are labeled. (C) Same as (B) for *S. cerevisiae* (left) and *S. paradoxus* (right) mapping to *S. cerevisiae* x *S. paradoxus*. (D) Same as (B) for *S. cerevisiae* Y12 (left) and *S. cerevisiae* DBVPG6044 (right) mapping to Y12 x DBVPG6044.

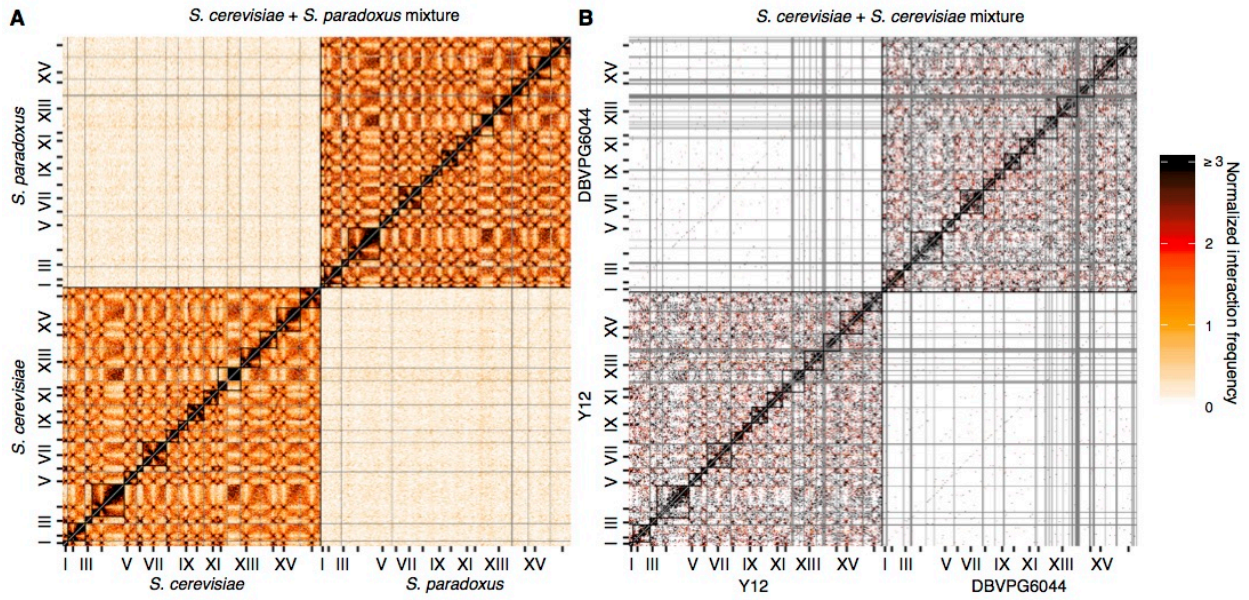

Figure 2.3. Mixture control experiments for *S. cerevisiae* x *S. paradoxus* and *S. cerevisiae* x *S. cerevisiae* hybrids.

(A) Hi-C contact map for a mixture of *S. cerevisiae* and *S. paradoxus* in exponential growth, at 32 kb resolution. Each axis represents the *S. cerevisiae* genome followed by the *S. paradoxus* genome, in syntenic order, separated by a black line; tick marks indicate centromere positions, and odd-numbered chromosome centromeres are labeled. Intrachromosomal interactions are outlined by black squares along the diagonal. Rows and columns with an average of less than 1 read pair per bin were filtered out, and are colored grey. (B) Hi-C contact map for a mixture of *S. cerevisiae* Y12 and *S. cerevisiae* DBVPG6044 haploids in exponential growth, as in (A).

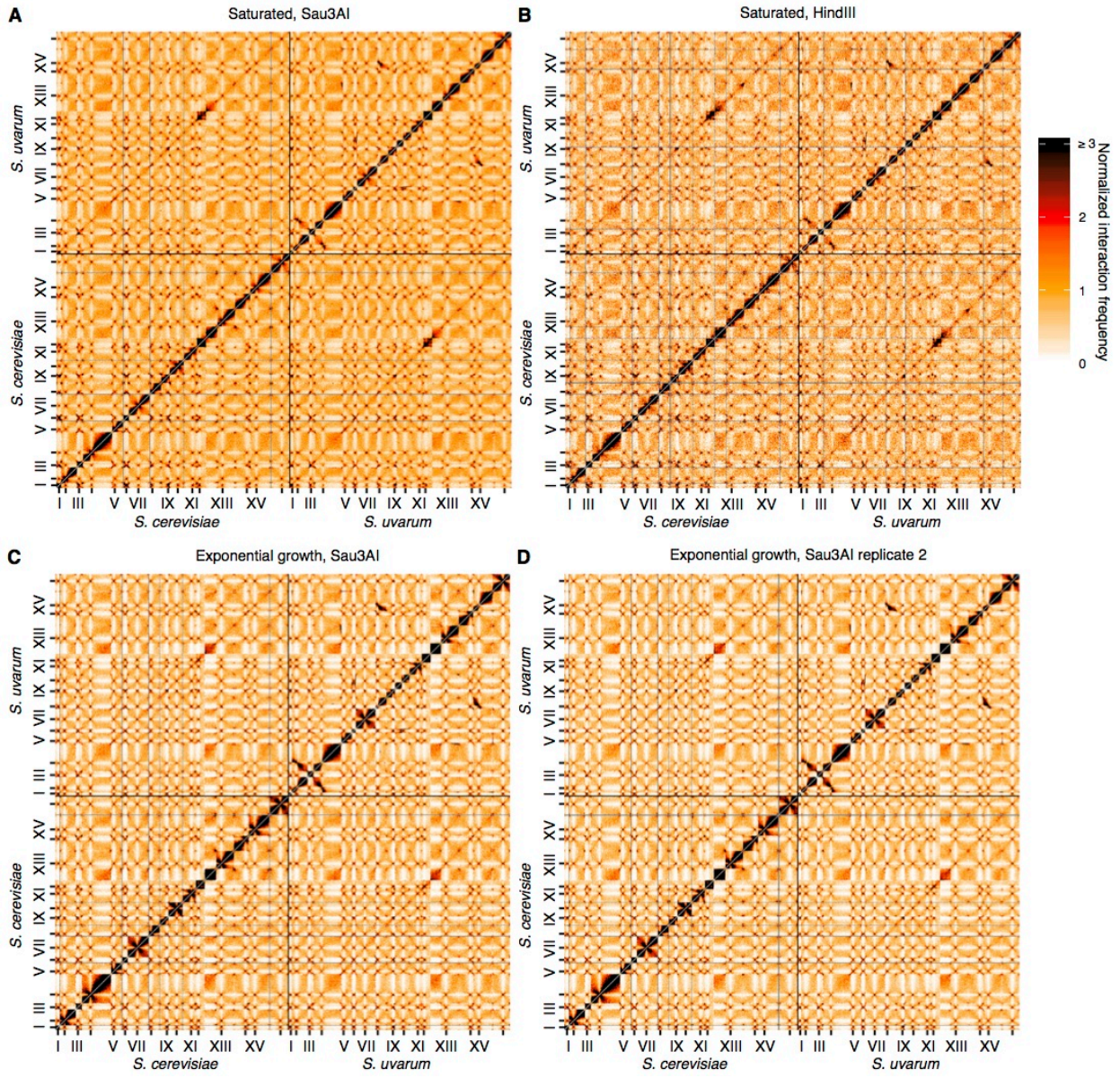

Figure 2.4. Reproducibility of Hi-C across replicates and restriction enzymes.

(A and B), Hi-C contact maps at 32 kb resolution for the *S. cerevisiae* x *S. uvarum* hybrid in saturated cultures, using the restriction enzyme Sau3AI (A) or HindIII (B). Each axis represents the *S. cerevisiae* genome followed by the *S. uvarum* genome, in syntenic order, separated by a black line; ticks indicate centromeres, and odd-numbered centromeres are labeled. Rows and columns with insufficient data are colored grey. (C and D) Same as (A) and (B) for two biological replicates of *S. cerevisiae* x *S. uvarum* hybrid in exponential growth, using the restriction enzyme Sau3AI.

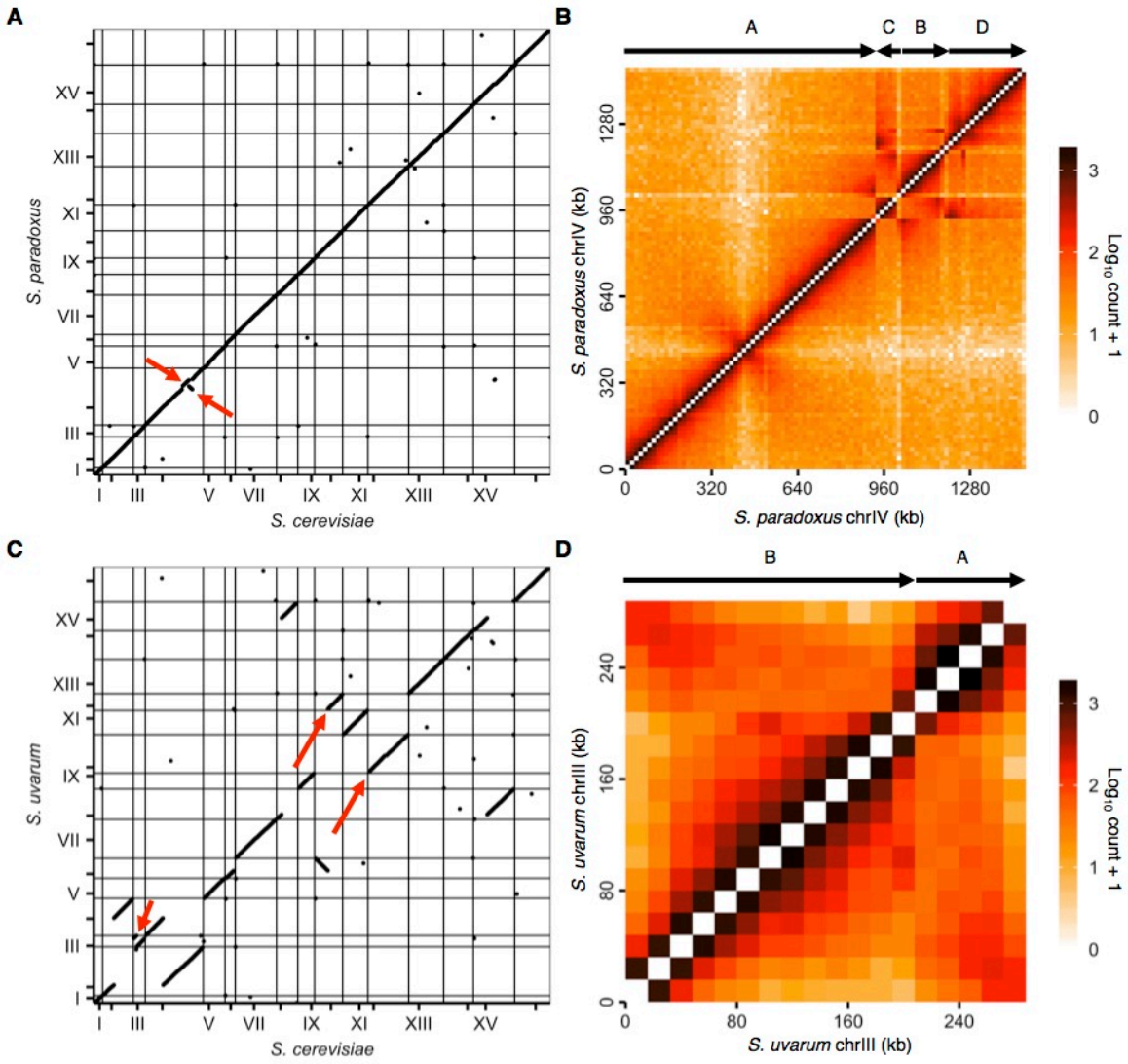

Figure 2.5. Revisions to *S. paradoxus* and *S. uvarum* reference genomes.

(A) Start and end positions of homologous genes on the original *S. cerevisiae* and *S. paradoxus* reference genomes. Each point represents the start or end of a homologous gene pair. Horizontal and vertical lines indicate starts of chromosomes, and ticks indicate centromere positions. Odd numbered chromosomes are numbered. Red arrows indicate unexpected rearrangements. (B) Raw contact map for *S. paradoxus* chromosome IV using the original reference genome. Arrows above the heat map represent segments of the chromosome, labeled A-D in order and orientation of synteny. (C) Same as (A) for *S. cerevisiae* and *S. uvarum*. Red arrows indicate an unexpected rearrangement in chromosome III, and *S. uvarum* chromosomes X and XII, which are homologous to *S. cerevisiae* chromosome XII and X, respectively. (D) Same as (B) for *S. uvarum* chromosome III.

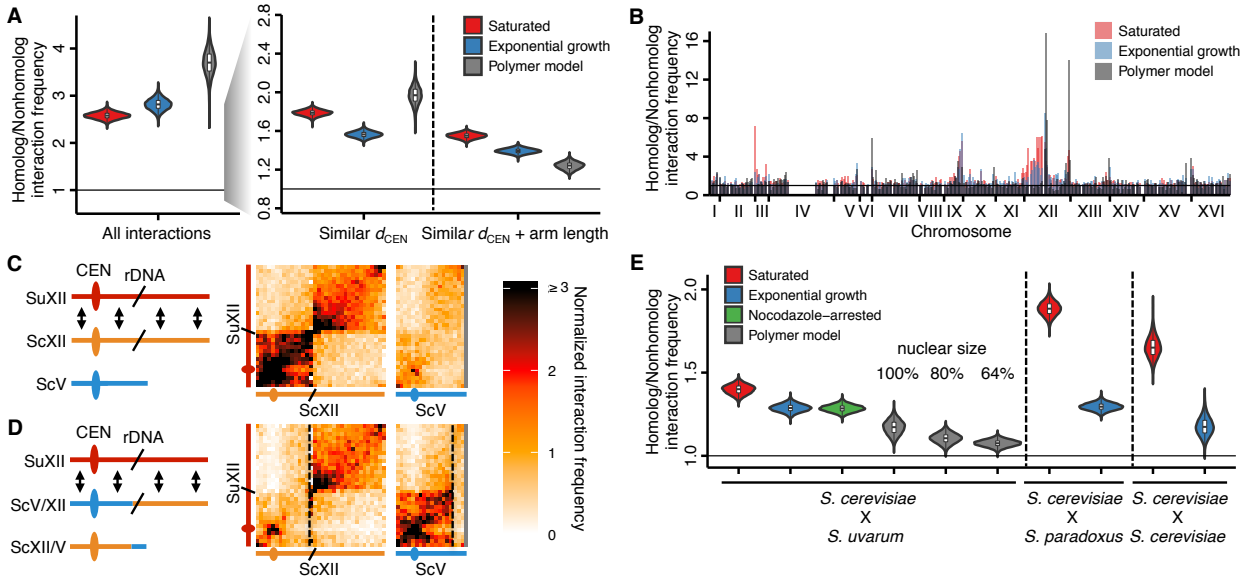

Figure 2.6. Homolog proximity exceeds predicted effects of Rabl-like orientation.

(A) Violin plot of the distribution of 10,000 sampled estimates of genomic homolog proximity (ratio of homologous to nonhomologous interaction frequencies) in the *S. cerevisiae* x *S. uvarum* hybrid, as a function of increasing comparison stringency (left to right) to account for Rabl-like orientation (**Figure 2.7**). Saturated culture data are shown in red, exponential growth in blue, and simulated data from a homology-agnostic polymer model in grey. d_{CEN}, distance from centromere. Boxplot indicates median and interquartile range. Whiskers correspond to the highest and lowest points within 1.5× interquartile range. (B) Variation in homolog proximity across the *S. cerevisiae* x *S. uvarum* hybrid genome at 32 kb resolution, in saturated culture (red), exponential growth (blue), and the polymer model (grey). Nonhomologous interactions were restricted to similar centromere distance and chromosome arm length. Bins with insufficient data (< 2 comparisons) are left blank. Data are plotted by *S. cerevisiae* genome position. *x* ticks indicate ends of chromosomes. (C and D) Schematics and Hi-C contact maps (at 32 kb resolution) of interactions between *S. uvarum* chromosome XII (SuXII) and either *S. cerevisiae* chromosome XII (ScXII) or *S. cerevisiae* chromosome V (ScV), in wild-type *S. cerevisiae* x *S. uvarum* hybrids (C) and a strain with a translocation between ScXII and ScV (D), both in saturated cultures. Exponential growth data are shown in **Figure 2.8**. Ovals indicate centromeres and slanted lines indicate the locations of rDNA arrays. Double-headed arrows indicate enhanced interactions. Dashed lines indicate translocation breakpoints. (E) Violin plot of homolog proximity across conditions, polymer models, and hybrids. *S. cerevisiae* x *S. cerevisiae* indicates hybrid between Y12 and DBVPG6044 strains. Calculated as in (A), but excluding chromosome XII and all 32 kb bins that include centromeres. Saturated culture data are shown in red, exponential growth in blue, nocodazole-arrested in green, and polymer models in grey. Boxplot indicates median and interquartile range. Whiskers correspond to the highest and lowest points within the 1.5× interquartile range.

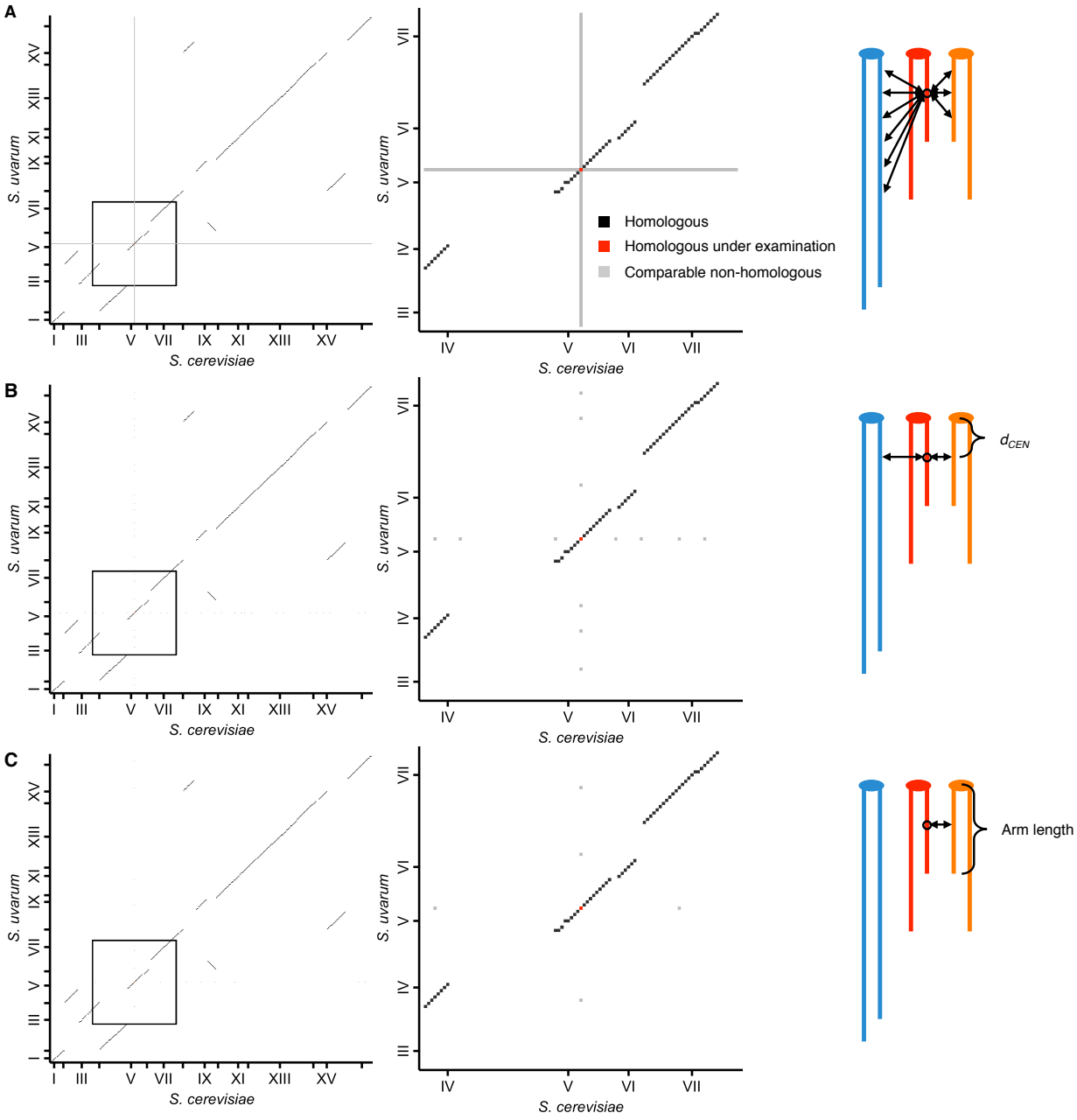

Figure 2.7. Schematic of homolog proximity analysis.

Representation of how homologous interactions (black squares) were compared to various subsets of nonhomologous interactions (grey squares), either including all interactions with either homologous locus (red squares) (A), restricted to interactions with loci at a similar centromeric distance, or d_{CEN} (same number of 32 kb bins) (B), or restricted to interactions with loci at a similar centromeric distance and on a chromosome arm of similar length (within 25%) (C). Left panels show all interactions between the *S. cerevisiae* and *S. uvarum* genomes; middle panels show enlarged view of the area outlined in the left panels. Tick marks indicate centromere positions. Right panels represent the nonhomologous interactions being used for comparison;

different colors represent nonhomologous chromosomes, and double-headed arrows represent interactions with the locus of interest (circled).

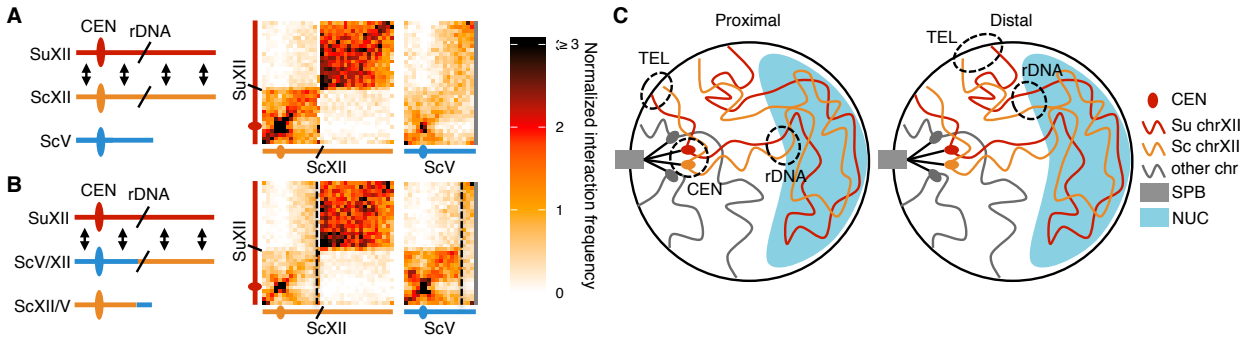

Figure 2.8. rDNA-carrying chromosomes interact preferentially due to shared tethering.

(A and B) Schematics and contact maps of rDNA-carrying chromosomes *S. uvarum* chromosome XII (SuXII) and *S. cerevisiae* chromosome XII (ScXII), and *S. cerevisiae* chromosome V (ScV), in wild-type *S. cerevisiae* x *S. uvarum* hybrids (A) and a strain with a translocation between ScXII and ScV (B), both in exponential growth. Ovals indicate centromeres and slanted lines indicate the rDNA arrays. Double-headed arrows indicate enhanced interactions. Dashed lines in the contact maps indicate the translocation breakpoints. (C) Schematic of how the rDNA-carrying chromosomes *S. cerevisiae* (Sc) chrXII and *S. uvarum* (Su) chrXII preferentially interact due to shared tethering. The proximal halves (left diagram) of the chromosomes, which contain the centromeres (CEN), are tethered at the spindle pole body (SPB) at their centromeres, at the periphery at their telomeres (TEL), and at the nucleolus (NUC) at their rDNA arrays (rDNA). The distal halves (right diagram) are tethered at their telomeres and rDNA, but not their centromeres. These combinations of tethering points are not found in other chromosomes (shown in grey). Su, *S. uvarum*; Sc, *S. cerevisiae*.

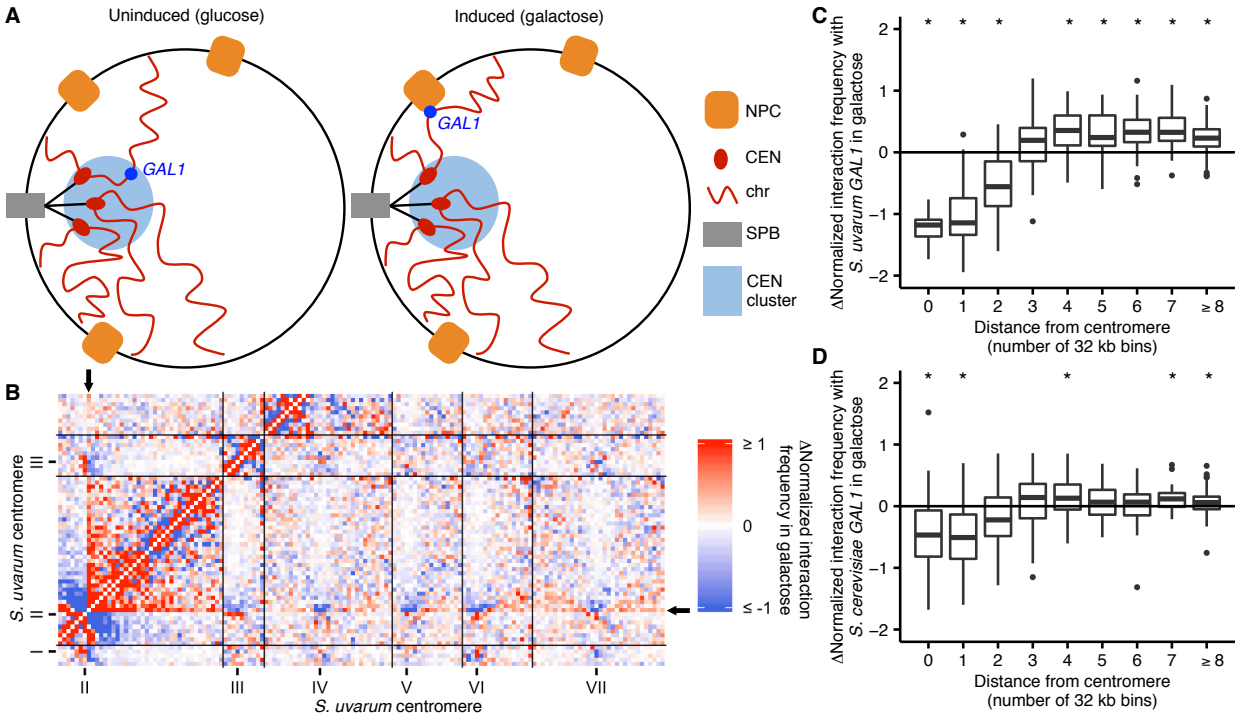

Figure 2.9. *GAL1* shifts away from centromeres upon galactose induction.

(A) Schematic of *GAL1* positioning (dark blue) in glucose (left) and galactose (right). NPC, nuclear pore complex; CEN, centromere; chr, chromosome; SPB, spindle pole body. (B) Example region of differential Hi-C map of *S. cerevisiae* x *S. uvarum* hybrids in galactose vs. glucose, at 32 kb resolution. Interactions that strengthen in galactose are in red, while those that weaken are in blue. Ticks indicate centromeres; black lines indicate chromosomes. Arrows indicate location of *S. uvarum GAL1*. (C) Boxplot of the difference in *S. uvarum GAL1* interaction frequency in galactose vs. glucose across the *S. cerevisiae* x *S. uvarum* genome, excluding intrachromosomal interactions and binned by distance from the centromere (in 32 kb bins). Whiskers correspond to the highest and lowest points within the 1.5× interquartile range. **P* < 0.05 after Bonferroni correction (*n* = 9); Mann-Whitney test. Note: some outliers are beyond the plot range and are not shown. (D) Same as (C) for *S. cerevisiae GAL1*.

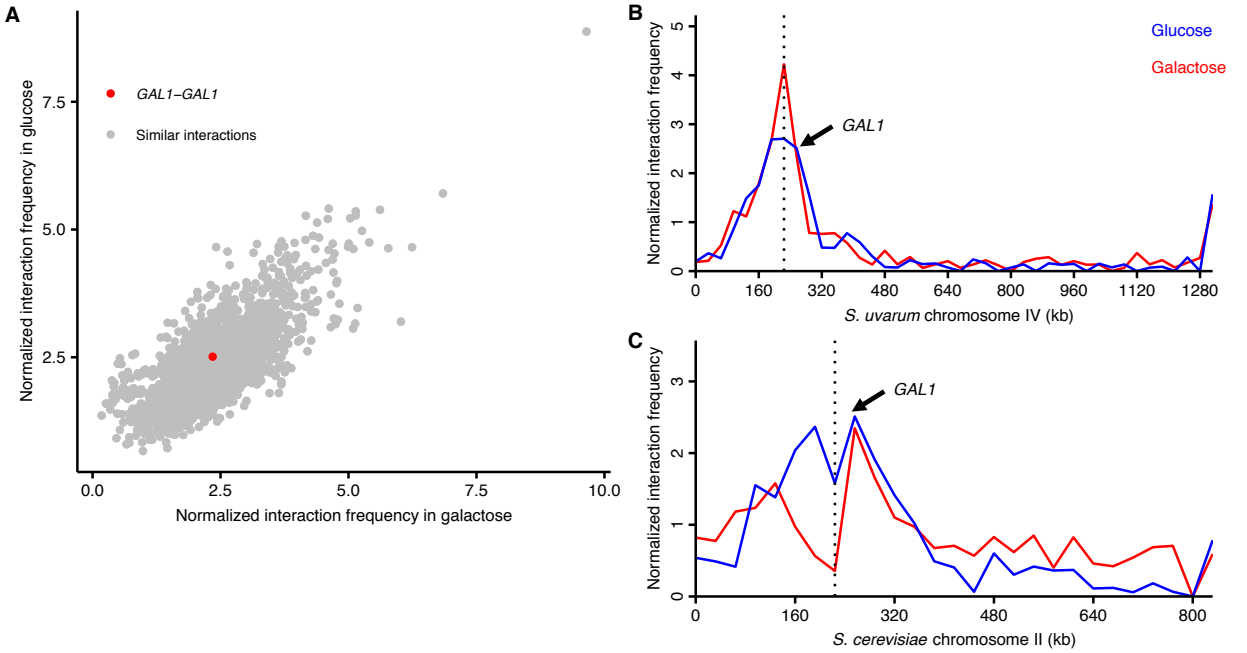

Figure 2.10. *GAL1* homologs do not detectably pair during galactose induction.

(A) Scatter plot of normalized interaction frequencies in galactose (*x*-axis) and in glucose (*y*-axis), between the *GAL1* homologs (in red) and other interactions between loci at the same centromeric distance (in grey). All interactions are between 32 kb bins. (B) Normalized interaction frequencies between the *S. cerevisiae GAL1* and the *S. uvarum* chromosome IV in glucose (blue) and galactose (red). Arrow points toward interaction between *GAL1* homologs, and the dashed line indicates the location of the centromere. (C) Same as (B) for the *S. uvarum GAL1* and the *S. cerevisiae* chromosome II.

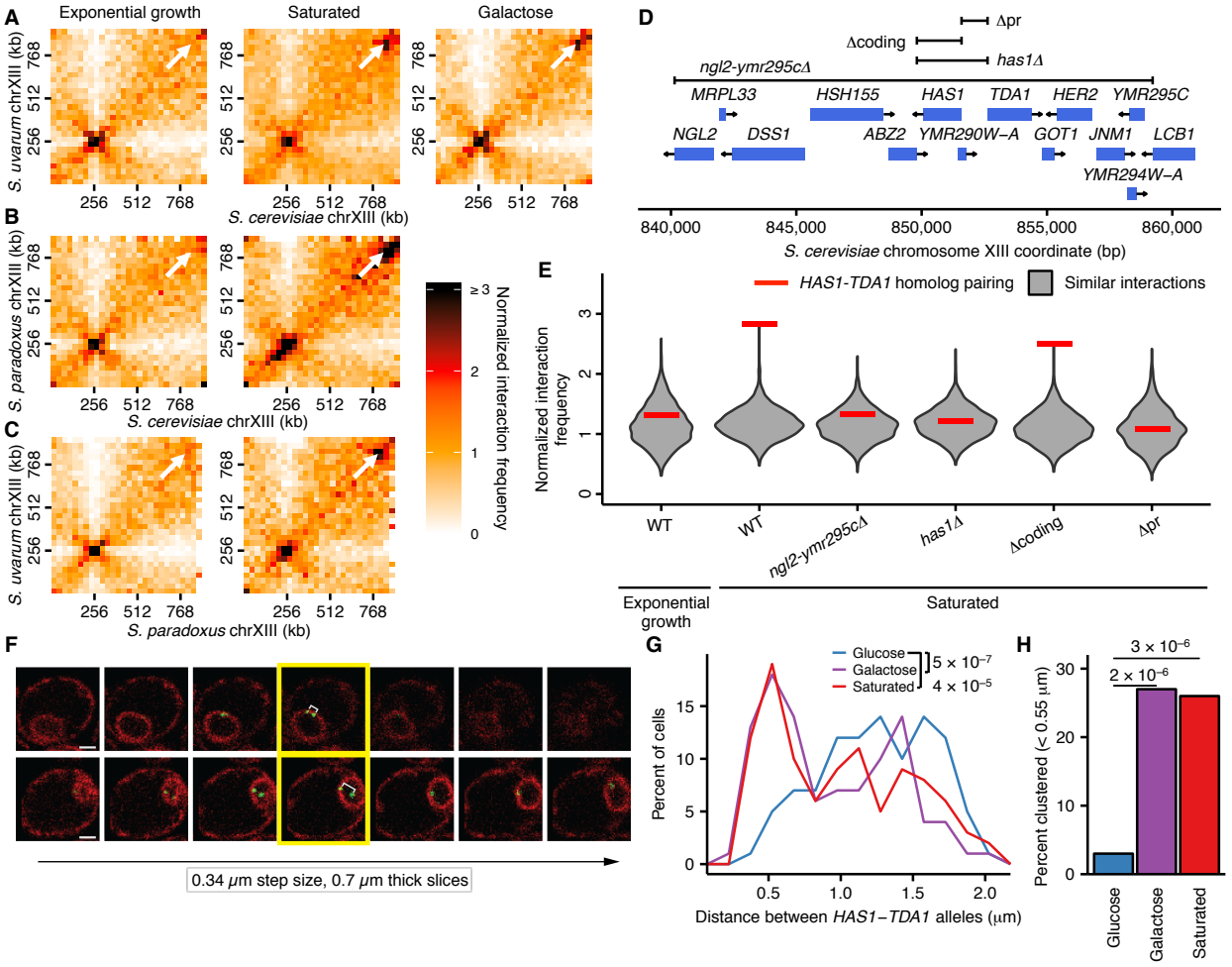

Figure 2.11. Inducible pairing of *HAS1-TDA1* homologs is evolutionarily conserved and sequence-specific.

(A, B, and C) Hi-C contact maps of chromosome XIII interactions at 32 kb resolution in *S. cerevisiae* x *S. uvarum* (A), *S. cerevisiae* x *S. paradoxus* (B), and *S. paradoxus* x *S. uvarum* (C) hybrids in exponential growth (left column), saturated cultures (middle column), and in *S. cerevisiae* x *S. uvarum* hybrids (A), galactose (right column). White arrows indicate the interaction between the homologous *HAS1-TDA1* loci. (D) Genome browser shot of open reading frames (ORFs; blue boxes) and tested deletions (brackets) in the *S. cerevisiae* region surrounding the genes *HAS1* and *TDA1*, from positions 840,000-860,000 (**Figure 2.12A**). Arrows indicate ends and directionalities of ORFs. (E) Strength of *HAS1-TDA1* homolog pairing at 32 kb resolution (red lines) compared to similar interactions (grey violin plots; i.e. interactions between an *S. cerevisiae* locus and an *S. uvarum* locus, where both loci are ≥ 15 bins from a centromere and ≥ 1 bin from a telomere, and not both on chromosome XII) in wild-type and deletion strains of *S. cerevisiae* x *S. uvarum*. See **Figure 2.13** for Hi-C contact maps of deletion strains. (F) Two example z-stacks of images used to measure distances between *HAS1-TDA1* alleles tagged with LacO arrays targeted by LacI-GFP (shown in green), with membranes labeled

by ER04 mCherry (shown in red). The yellow outline indicates the images chosen for analysis. White brackets indicate the measured distance. Scale bar = 1 μm. (G) Distributions of the distance between the *HAS1-TDA1* alleles measured by microscopy in *S. cerevisiae* diploids in glucose (blue), galactose (purple), and saturated cultures (red). n = 100 for each condition. *P*-values were calculated using the Wilcoxon rank sum test. (H) Frequency of *HAS1-TDA1* alleles less than 0.55 μm apart, measured as in (G), in glucose (blue), galactose (purple), or saturated cultures (red). *P*-values were calculated using Fisher's exact test.

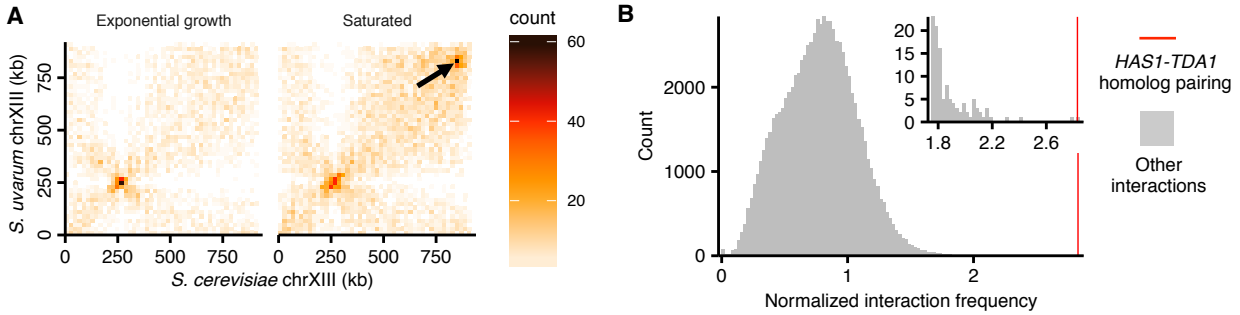

Figure 2.12. Exceptional inducible homolog pairing at *HAS1-TDA1* locus.

(A) Raw Hi-C contact maps at 20 kb resolution for *S. cerevisiae* and *S. uvarum* chromosome XIII. Arrow points to strongest interaction in saturated culture excluding regions near centromeres, at positions 840,000-860,000 on the *S. cerevisiae* chromosome XIII, the target of deletion studies. (B) Histogram comparing *HAS1-TDA1* homolog pairing interaction frequency (red line) to all other interactions (grey) between an *S. cerevisiae* locus and an *S. uvarum* locus (32 kb bin), with both loci ≥ 3 bins from a centromere, > 1 bin from a telomere, and not both on an rDNA-carrying chromosome. Inset shows enlarged view of right end of plot.

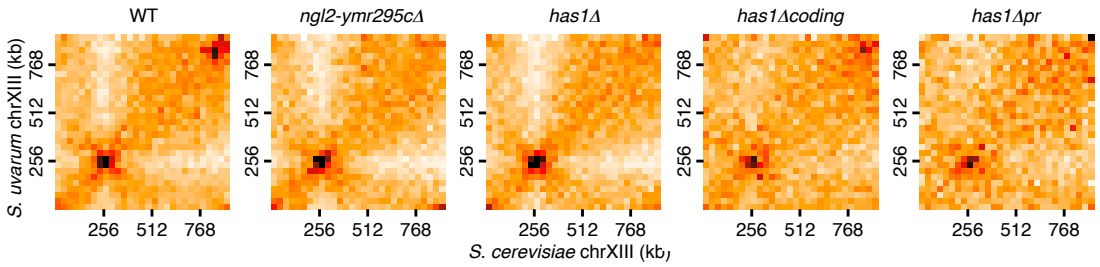

Figure 2.13. *HAS1-TDA1* homolog pairing does not shift nearby upon deletion.

Hi-C contact maps of chromosome XIII interactions at 32 kb resolution in *S. cerevisiae* x *S. uvarum* hybrids in saturated cultures, in wild-type (WT) and deletion strains. See **Figure 2.11D** for deletion boundaries.

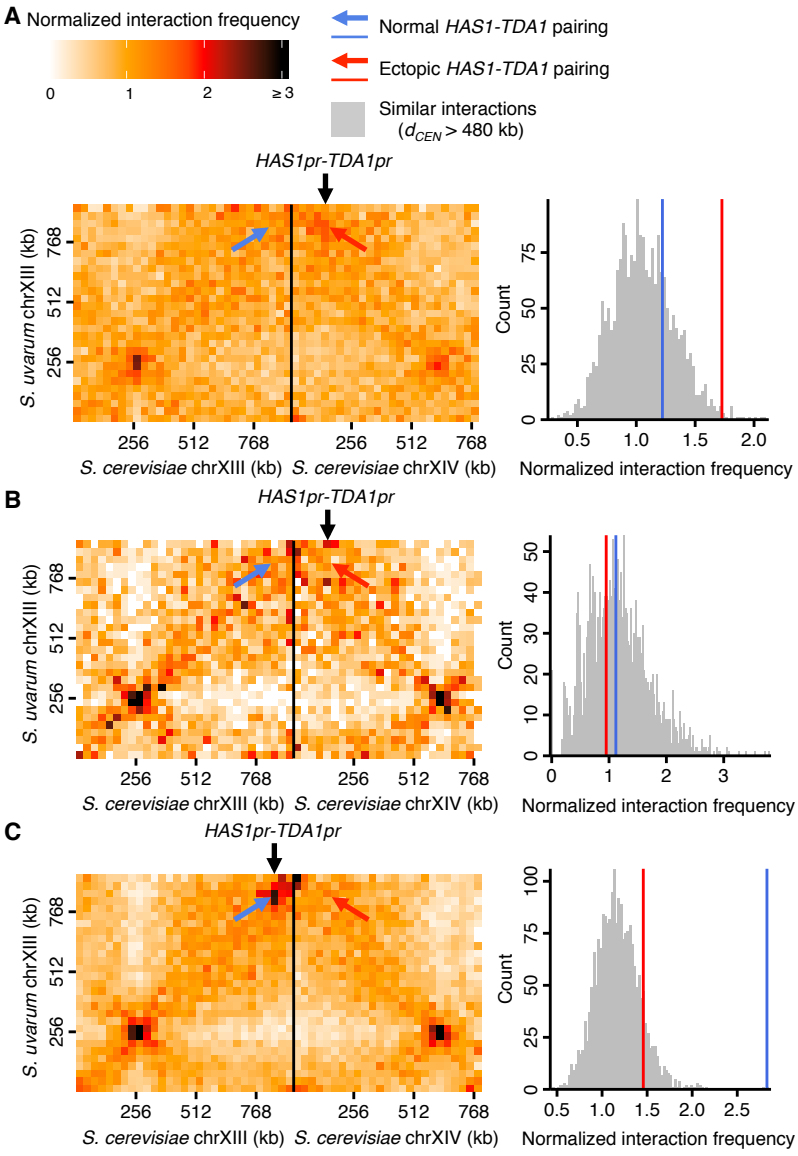

Figure 2.14. *HAS1-TDA1* homolog pairing is recapitulated ectopically by the *HAS1* and *TDA1* promoters.

Genomic relocation of the *S. cerevisiae HAS1pr-TDA1pr* 1 kb region causes inducible ectopic pairing with the *S. uvarum HAS1-TDA1* allele. (Left) Hi-C contact maps of interactions between *S. uvarum* chromosome XIII and *S. cerevisiae* chromosomes XIII and XIV in the *S. cerevisiae* x *S. uvarum* hybrid with the *S. cerevisiae HAS1pr-TDA1pr* region moved to *S. cerevisiae* chromosome XIV (location indicated by vertical black arrow), in saturated culture (A) and in exponential growth (B), compared to the wild-type hybrid (C) in saturated culture. Interactions between the *S. uvarum HAS1-TDA1* locus and the original *S. cerevisiae HAS1-TDA1* locus are indicated by a blue arrow, whereas interactions with the new *HAS1pr-TDA1pr* locus are indicated by a red arrow. (Right) Histograms comparing the pairing frequency of the *S. uvarum HAS1-TDA1* locus and either the normal *S. cerevisiae HAS1-TDA1* locus (blue line) or the

ectopic *HAS1pr-TDA1pr* locus (red line) to the frequency of all similar interactions, i.e. those between an *S. cerevisiae* locus and an *S. uvarum* locus (32 kb bin), with both loci ≥ 15 bins from a centromere (or d_{CEN} > 480 kb), > 1 bin from a telomere, and not both on an rDNA-carrying chromosome.

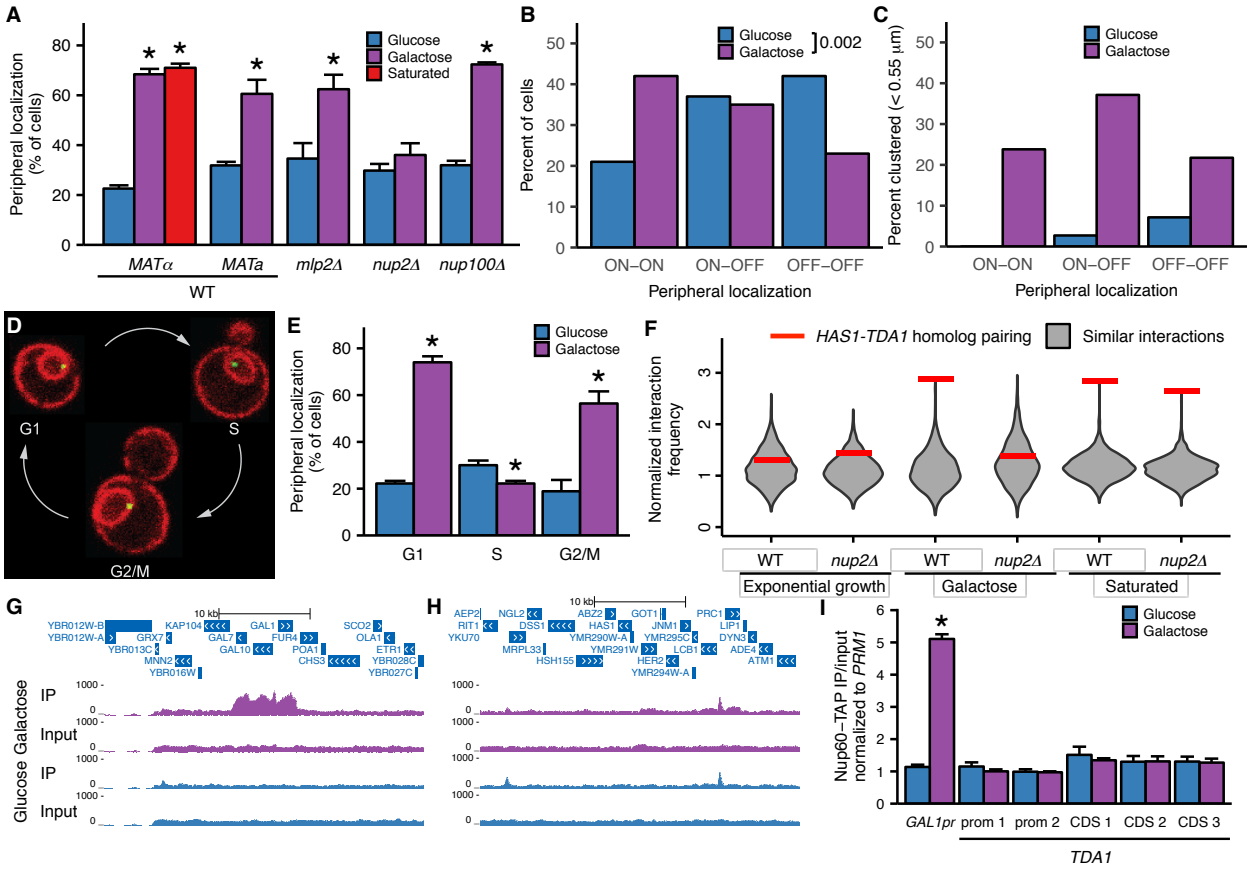

Figure 2.15. Inducible peripheral localization and pairing of *HAS1-TDA1* alleles involve nuclear pore interactions.

(A) Proportions of haploid *S. cerevisiae* cells exhibiting peripheral *HAS1-TDA1* localization in strains with and without deletions of nuclear pore components, in glucose (in blue), galactose (in purple), or saturated culture (in red). Experiments were performed in biological triplicate, with $n \geq 30$ per experiment. *$P < 0.05$, Student's *t*-test. Center values and error bars represent mean ± s.e.m. (B) Proportion of diploid *S. cerevisiae* cells exhibiting two (ON-ON), one (ON-OFF), or zero *HAS1-TDA1* alleles with peripheral localization, in glucose (blue) and galactose (purple). *P*-value calculated using chi-squared test. (C) Proportion of diploid *S. cerevisiae* cells with *HAS1-TDA1* alleles clustered (< 0.55 µm apart), in glucose (blue) and galactose (purple) as a function of the peripheral localization of *HAS1-TDA1* alleles. Same images used as in (B) and **Figure 2.11G,H**. (D) Example images of cells in G1, S, and G2/M phases of cell cycle. (E) Proportions of haploid *S. cerevisiae* cells exhibiting peripheral *HAS1-TDA1* localization in different phases of the cell cycle, in glucose (in blue) and galactose (in purple). *$P < 0.05$, Student's *t*-test. Center values and error bars represent mean ± s.e.m. (F) Strength of *HAS1-TDA1* homolog pairing at 32 kb resolution (red lines) compared to similar interactions (grey violin plots; i.e. interactions between an *S. cerevisiae* locus and an *S. uvarum* locus, where both loci are ≥ 15 bins from a centromere and ≥ 1 bin from a telomere, and not both on chromosome XII) in wild-type and homozygous nup2Δ strains of *S. cerevisiae* x *S. uvarum*. (G and H) Nup60-TAP ChIP-seq read

coverage tracks for IP and input in galactose (purple) and glucose (blue), zoomed into a 35 kb region surrounding *GAL1-GAL10-GAL7* on chromosome II (G), and a 35 kb region surrounding *HAS1* and *TDA1* on chromosome XIII (H). (I) Nup60-TAP ChIP qPCR as IP/input normalized to the negative control *PRM1*, for the positive control *GAL1pr* and five sets of primers in *TDA1*. Primer sequences are provided in **Supplementary file 2**. Center values and error bars represent mean ± s.e.m. of three biological replicates. **P* < 0.05, Student's *t*-test.

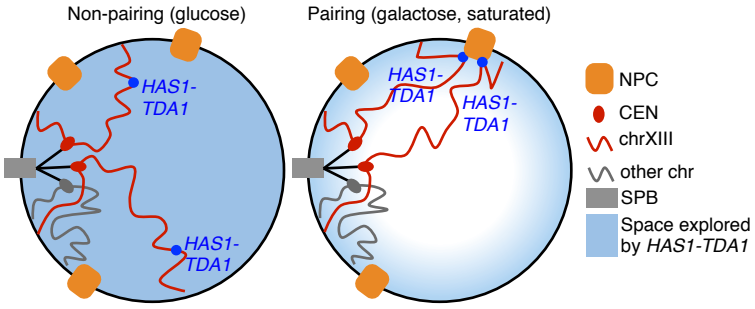

Figure 2.16. Schematic of how nuclear pore association mediates homologous *HAS1* pairing.

NPC, nuclear pore complex; CEN, centromere; chr, chromosome; SPB, spindle pole body.

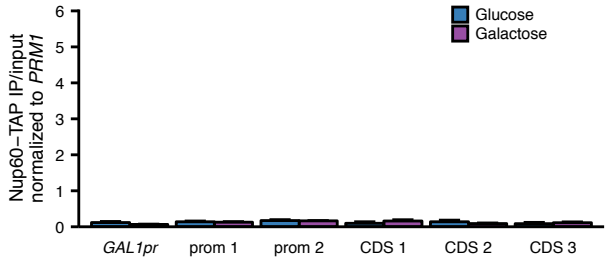

Figure 2.17. Mock-IP on Nup60-TAP.

Nup60-TAP ChIP qPCR using BSA instead of anti-TAP antibody, plotted as IP/input normalized to the negative control *PRM1*, for the positive control *GAL1pr* and five sets of primers in *TDA1*. Primer sequences are provided in **Supplementary file 2**. Center values and error bars represent mean ± s.e.m. of three biological replicates. No values are significant ($P < 0.05$, Student's *t*-test).

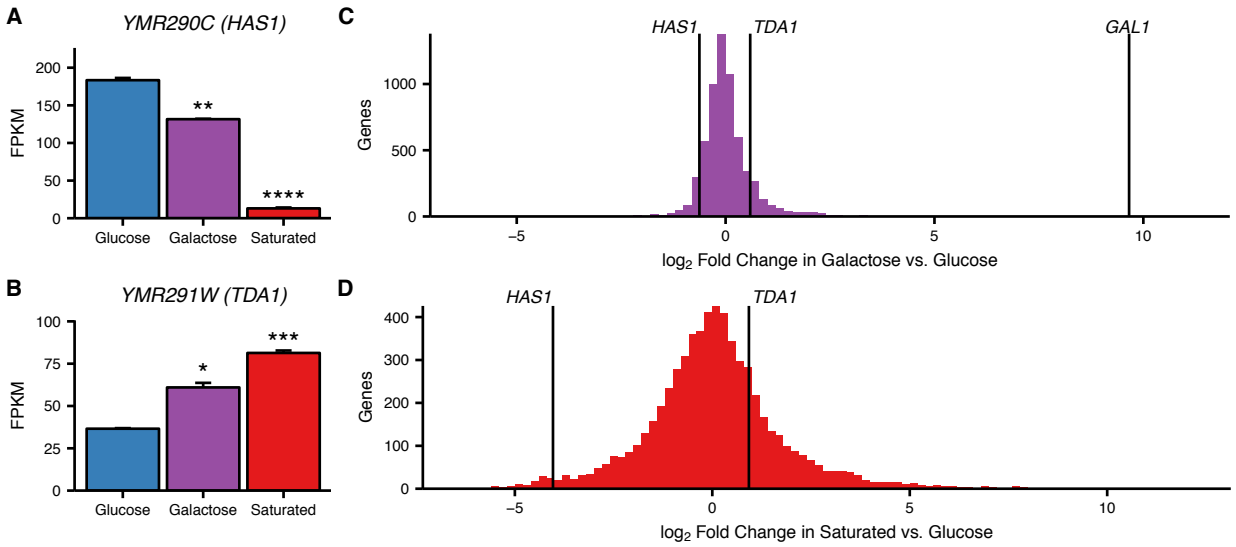

Figure 2.18. Transcriptional changes in galactose and saturated culture.

(A and B) Bar plots of gene expression in haploid *S. cerevisiae* grown in glucose, galactose, or to saturation, for *HAS1* (A) and *TDA1* (B). Asterisks indicate *P*-values < 0.05 (*), 0.01 (**), 0.001 (***), or 0.0001 (****), Student's *t*-test. Center values and error bars represent mean ± s.e.m. of three biological replicates (C and D) Histogram of $\log_2$ fold change in gene expression in galactose (C) or saturated growth (D) compared to glucose. Vertical lines indicate values for *HAS1*, *TDA1*, and *GAL1* (C only).

Chapter 3. A COMBINATION OF TRANSCRIPTION FACTORS MEDIATES INDUCIBLE INTERCHROMOSOMAL CONTACTS

Note: Chapter 3 was adapted with minimal modification from a manuscript currently under review at *eLife*.

3.1 ABSTRACT

The genome forms specific three-dimensional contacts in response to cellular or environmental conditions. However, it remains largely unknown which proteins specify and mediate such contacts. Here we describe an assay, MAP-C (Mutation Analysis in Pools by Chromosome conformation capture), that simultaneously characterizes the effects of hundreds of *cis* or *trans*-acting mutations on a chromosomal contact. Using MAP-C, we show that inducible interchromosomal pairing between *HAS1pr-TDA1pr* alleles in saturated cultures of *Saccharomyces* yeast is mediated by three transcription factors, Leu3, Sdd4 (Ypr022c), and Rgt1. The coincident, combined binding of all three factors is strongest at the *HAS1pr-TDA1pr* locus and is also specific to saturated conditions. We applied MAP-C to further explore the biochemical mechanism of these contacts, and find they require the structured regulatory domain of Rgt1, but no known interaction partners of Rgt1. Altogether, our results demonstrate MAP-C as a powerful method for dissecting the mechanistic basis of chromosome conformation.

3.2 INTRODUCTION

The three-dimensional organization of the genome within the nucleus is structured but dynamic (Bonev and Cavalli, 2016). Although many features of this conformation are largely

conserved across cell types and conditions (Rao et al., 2014; Schmitt et al., 2016a), some chromatin loops and contacts form specifically in response to signals such as differentiation (Bonev et al., 2017; Monahan et al., 2019; Schmitt et al., 2016a; Stadhouders et al., 2018), changes in nutrient availability (Brickner et al., 2015, 2016), heat shock (Chowdhary et al., 2019, 2017), drugs (D'Ippolito et al., 2018), meiosis (Muller et al., 2018), or circadian rhythms (Kim et al., 2018b). This dynamic three-dimensional organization of the genome plays a role in regulating gene expression in diverse organisms. In multicellular organisms, active developmental gene promoters form long-range loops with specific enhancer elements (Bonev et al., 2017), and this looping is in some cases sufficient for transcriptional activation (Deng et al., 2014). In the budding yeast *Saccharomyces cerevisiae*, a well-studied model of genome conformation, genes targeted to nuclear pores are activated (Taddei et al., 2006), while those at the nuclear periphery are repressed (Andrulis et al., 1998).

Transcription factors (TFs) are attractive candidates for orchestrating such dynamic changes in chromatin conformation, given their site-specific DNA binding and changes in abundance or activity in response to differentiation and cellular signals (Lambert et al., 2018). For many conditions, it remains unknown exactly which TFs bind to any given locus. Although binding site motifs are known for many TFs, motif searches poorly predict TF binding (Guertin and Lis, 2010; Jolma et al., 2015; Le et al., 2018; Levo et al., 2015; Liu et al., 2006b; Slattery et al., 2014). Even if the set of TFs bound to each locus is known, it is unclear which TFs are capable of forming chromosomal contacts. DNA-bound TFs can also recruit other cofactor proteins that can mediate chromosomal contacts (Deng et al., 2012; Monahan et al., 2019; Song et al., 2007), but our understanding of TF-cofactor interactions remains incomplete.

Among chromosomal contacts and loops, interchromosomal contacts are less well-understood. This is in part due to the relative paucity of interchromosomal contacts in Hi-C and other 3C (chromosome conformation capture) data, which results from their greater contact distance (Maass et al., 2018) and chromosomal self-association into territories (Cremer and Cremer, 2010). Nevertheless, many distinct classes of interchromosomal contacts are known, including clustering of transcriptionally active genes (Mitchell and Fraser, 2008; Osborne et al., 2004; Schoenfelder et al., 2009), associations with nuclear bodies (Quinodoz et al., 2018), interactions among developmental enhancers and promoters (Lomvardas et al., 2006; Monahan et al., 2019), and mitotic homologous chromosome pairing in organisms ranging from yeast (Burgess et al., 1999) to flies (Henikoff and Dreesen, 1989; Joyce et al., 2016; Morris et al., 1999) and mammals (Xu et al., 2006). However, our understanding of the molecular mechanisms of these contacts remains incomplete.

Many known mechanisms for establishing 3D chromosome conformation may act on both intrachromosomal loops and interchromosomal contacts. The emerging consensus model for such DNA-DNA interactions involves loop extrusion by cohesin and other Structural Maintenance of Chromosomes (SMC) factors, which is thought to primarily mediate intrachromosomal loops (Alipour and Marko, 2012; Rao et al., 2014; Rowley and Corces, 2018; Sanborn et al., 2015; Swygert et al., 2019). However, SMC complexes are also capable of mediating interactions between multiple DNA molecules, such as between sister chromatids (Michaelis et al., 1997). Interchromosomal contacts (Monahan et al., 2019) and intrachromosomal contacts (Weintraub et al. 2017; Deng et al. 2012) can also be mediated by dimerization of structured proteins. In addition to these well-defined strong molecular interactions, weak interactions such as those underlying phase separation of nuclear factors such

as transcription factors (Boija et al., 2018; Chong et al., 2018), coactivators (Cho et al., 2018), RNA polymerase (Boehning et al., 2018), and heterochromatin proteins (Larson et al., 2017; Strom et al., 2017) may play an important role in shaping 3D genome organization. How these and other mechanisms synergize remains an open question.

Mitotic (or somatic) homologous chromosome pairing is the preferential association of homologous pairs of loci in mitotically dividing cells. Homolog pairing occurs along the length of the genome in *Drosophila* (Joyce et al., 2016), but is more subtle in yeast and other organisms, where the association is often transient and/or genomically localized (Xu et al., 2006). Fluorescence *in situ* hybridization screens in flies have nominated various pairing and anti-pairing factors that modulate the strength of homolog pairing (Joyce et al., 2012), but the precise mechanisms by which these factors regulate pairing are largely unknown. In mammals, X chromosome pairing is mediated by CTCF and Oct4 (Donohoe et al., 2009), in conjunction with transcription (Xu et al., 2007). However, cases of highly localized homolog pairing remain rare. Furthermore, the distinctions between homolog pairing and non-allelic interactions between repetitive elements (Gladyshev and Kleckner, 2017; Mirkin et al., 2014) remain unclear.

We recently identified a novel example of an inducible, localized interchromosomal contact between homologous copies of the *HAS1pr-TDA1pr* locus in diploid *Saccharomyces* yeasts (Kim et al., 2017). This interaction occurs in saturated culture conditions, requires the 1 kb intergenic region between the *HAS1* and *TDA1* coding sequences, and is detectable by both Hi-C and microscopy. The condition-specificity and dependence on intergenic sequence led us to hypothesize that one or more TFs might mediate this pairing. Although yeast TF binding is well-characterized for standard growth conditions (Badis et al., 2008), TF binding has not been systematically measured in saturated culture conditions. Meanwhile, computational predictions

of TF binding sites are insufficiently specific, e.g. within the 1 kb region required for *HAS1pr-TDA1pr* pairing, dozens of TFs have at least one motif match. Furthermore, even if the TFs bound to this region were known, it would remain unclear which subset played a role in mediating inducible interchromosomal pairing.

Here we describe a method that enables the simultaneous testing of hundreds of *cis* or *trans*-acting mutations for their effects on a chromosomal contact of interest. As a proof of concept, we applied this method, which we call Mutation Analysis in Pools by Chromosome conformation capture (MAP-C), to characterize the molecular components mediating *HAS1pr-TDA1pr* pairing. We first perform saturating mutagenesis of the regulatory region that mediates the interchromosomal pairing (*cis* MAP-C) to identify sequence motifs required for pairing that potentially correspond to TF binding sites. We then test the effects of knocking out over one hundred TFs (*trans* MAP-C), and confirm that three—Leu3, Sdd4 (Ypr022c), and Rgt1—are necessary for inducible interchromosomal pairing. We verify their binding by chromatin immunoprecipitation, and find that *HAS1pr-TDA1pr* exhibits the strongest coincident, combined binding by all three factors across the genome in a condition-specific manner. We further use *trans* MAP-C to interrogate how interaction partners and domains of Rgt1 regulate pairing. Finally, we make an initial attempt to characterize the functional consequences of *HAS1pr-TDA1pr* pairing. Taken together, our results demonstrate how a combination of TFs can mediate inducible interchromosomal pairing. Furthermore, our study shows the utility of a pooled mutant approach to studying both the *cis* and *trans* dependencies of chromosome conformation.

3.3 RESULTS

3.3.1 *A pooled approach to systematically dissect chromosome conformation*

In order to identify and dissect the molecular mechanisms underlying chromosome conformation, experiments involving perturbations (*e.g.* mutations) are needed. However, despite numerous advances in chromosome conformation capture (3C) technology over the last two decades (de Wit and de Laat, 2012; Schmitt et al., 2016b), each experiment characterizes a single sample, which limits the number of genes or *cis*-regulatory elements that can be disrupted (Monahan et al., 2019; Nora et al., 2017; Schwarzer et al., 2017; Weintraub et al., 2017).

To address this limitation and enable systematic screens, we developed MAP-C, an assay in which hundreds of mutations are simultaneously tested for their effects on a single chromosomal contact of interest (**Figure 3.1A**). In the *cis* version of MAP-C, which we describe first, these mutations are targeted to one of the regions involved in the chromosomal contact. In the *trans* version of MAP-C, mutations can be spread across the genome, as long as they are associated with a unique barcode sequence at the chromosomal contact site.

The first step of *cis* MAP-C is to generate an allelic series of a region of interest, which can be achieved in a cost-effective manner via array-synthesized oligonucleotide pools or error-prone PCR. Alternatively, if desired, variants can be generated individually and then pooled prior to conducting MAP-C. The resulting mutant pool is then integrated into the genome and subjected to the 3C assay (Dekker et al., 2002). The region containing the genetic variants is amplified using two different primer pairs: the first (3C library) amplifies a specific ligation product, and the second (genomic library) amplifies regardless of ligation. These amplification products are deeply sequenced to measure the abundance of each variant in the 3C library, which

is normalized to its abundance in the genomic library. The relative extent to which sequence variants participate in the chromosomal contact of interest is proportional to their normalized representation in the 3C library.

3.3.2 *A cluster of TF motifs is necessary and sufficient for HAS1pr-TDA1pr pairing*

As a first test of *cis* MAP-C, we sought to systematically dissect the conserved pairing between *HAS1pr-TDA1pr* homologs in diploid *Saccharomyces* yeasts grown to saturation. We recently used Hi-C of *S. cerevisiae* x *S. uvarum* hybrids (<80% nucleotide identity) to discover this homolog pairing interaction, and furthermore identified a 1,038 bp noncoding region that was necessary and sufficient for pairing (Kim et al., 2017).

To find a minimal subsequence of the 1,038 bp *HAS1pr-TDA1pr* region that is sufficient to pair with other *HAS1pr-TDA1pr* alleles, we replaced the native *S. cerevisiae HAS1pr-TDA1pr* locus with a library containing each of 861 tiling 178 bp subsequences of the 1,038 bp region (along with a G418 resistance cassette and restriction site), in *S. cerevisiae* x *S. uvarum* hybrid yeast. We then performed *cis* MAP-C for pairing of the modified locus with the *S. uvarum* copy of *HAS1pr-TDA1pr* on a saturated culture of the pool, in two replicates (**Figure 3.1B**). Compared to the genomic libraries, the 3C libraries were highly enriched for a narrow region spanning ~500 to ~700 bp from the *HAS1* coding sequence, with a plateau between ~525 to ~675 bp, consistent with that ~150 bp region being the only subsequence shorter than 178 bp sufficient for pairing (termed the "minimal pairing region" below). To confirm that this pattern of enrichment is specific to *HAS1pr-TDA1pr* homolog pairing rather than underlying all of its chromosomal contacts, we repeated the assay with a pair of primers amplifying a ~10 kb intrachromosomal contact (**Figure 3.2A and B**). In this "off-target" control, coverage from the 3C library matched that of the genomic control, suggesting that most variants are capable of

intrachromosomal looping, but only those containing the minimal pairing region are capable of interchromosomal pairing with the other *HAS1pr-TDA1pr* allele.

Since our initial experiment was performed in the endogenous genomic context, other DNA sequences outside but near the *HAS1pr-TDA1pr* locus could be required in addition to the minimal pairing region. We therefore tested whether inserting a 184 bp sequence that included the minimal pairing region into an ectopic location, the gene *FIT1* (*YDR534C*), would induce pairing with the native *HAS1pr-TDA1pr* locus in saturated cultures of haploid *S. cerevisiae* (**Figure 3.2C**). As a negative control, we inserted an equivalently sized subsequence insufficient for pairing into the same locus (**Figure 3.2C**). Indeed, insertion of the minimal pairing region led to a >30-fold increase in 3C signal for pairing with *HAS1pr-TDA1pr* as compared to the negative control (**Figure 3.1C**).

We next sought to obtain a base-pair resolution map of the DNA sequences necessary for pairing. We used error-prone PCR to generate variants of a 207 bp region (161 bp excluding fixed primer sequences) containing the minimal pairing region, with an average of 1.49 substitutions (range 0-14) per template (**Figure 3.2D**). We inserted this variant library in place of the native *S. cerevisiae HAS1pr-TDA1pr* sequence as before, and performed *cis* MAP-C. The ratio of total substitution abundance in the 3C and genomic libraries can be plotted at each mutagenized position. This identified six clusters of two or more adjacent positions showing strong depletion of substitutions in the 3C libraries, indicating that they are required for *HAS1pr-TDA1pr* pairing (**Figure 3.1D**). We inspected motifs in the region to identify candidate TFs that might mediate the pairing (**Figure 3.3 and 3)**. Given the abundance of potential matches, we prioritized motifs with high-scoring matches, motifs with conserved positions corresponding to those most important for pairing, and motifs occurring in multiple clusters. The first two clusters

together aligned to a Leu3 motif (**Figure 3.1E**), the third to several similar motifs, including Sdd4 (Ypr022c) and Mig1 (**Figure 3.1F**), and the last three to Rgt1 motifs in both orientations (**Figure 3.1G-I**). These motifs, which span 47 bp, include 23 of the 24 positions most depleted for mutations in the 3C libraries (**Figure 3.2E**), and all clusters of two or more adjacent such positions (**Figure 3.1D**). None of these mutations had a strong effect on intrachromosomal looping (**Figure 3.2F**), and all of the clusters were reproduced using an alternative mutagenesis strategy (programmed 3 bp substitutions) (**Figure 3.5**). Interestingly, a fourth Rgt1 motif and a second Sdd4/Mig1 motif mutagenized only in our validation experiment were not required for pairing, suggesting that either not all motifs in this region are bound by the same TFs or not all bound TFs are involved in mediating homolog pairing at this locus.

Thus, using *cis* MAP-C, we identified a ~150 bp subsequence of *HAS1pr-TDA1pr* sufficient for pairing, containing five required TF motif occurrences. If these TF motifs are together sufficient for pairing, we would expect that 1) they are only observed in a cluster in the minimal pairing region and not elsewhere in the *HAS1pr-TDA1pr*, and 2) they are present in a cluster in the *S. uvarum* copy of this region, and potentially other *Saccharomyces* as well. Indeed, these motifs are clustered together only in the central region of *HAS1pr-TDA1pr*; remarkably, this pattern holds across all *Saccharomyces* species (**Figure 3.6**).

3.3.3 *Three transcription factors are required for pairing*

Although we identified the TF motifs required for *HAS1pr-TDA1pr* homolog pairing at base-pair resolution with *cis* MAP-C, the redundancy among TF motifs made it difficult to definitively identify the TFs involved (**Figure 3.1F**). To address this, we developed a modified version of MAP-C that is capable of assaying *trans* mutations spread across the genome, such as gene knockouts, for their effects on a specific chromosomal contact (*trans* MAP-C). With *trans*

MAP-C, each mutation is uniquely associated with a short barcode sequence near a region involved in the chromosomal contact of interest, which is then assayed by 3C. Mutations that affect pairing frequency modulate the abundance of their corresponding barcodes in the 3C products, which can be readily quantified by deep sequencing. To study the *trans* requirements of *HAS1pr-TDA1pr* pairing, we ectopically integrated barcoded versions of the minimal pairing region, wherein the barcode identifies which TF is knocked out, and assayed their capacity to pair with the native copy of *HAS1pr-TDA1pr* (**Figure 3.7A**).

We first tested this approach in a pilot set of 10 TF knockouts by inserting the minimal pairing region (**Figure 3.2C**) into the common KanMX drug resistance cassette used to replace each deleted gene in the haploid yeast deletion collection (Giaever et al., 2002). We then assayed these constructs for interactions with the native *HAS1pr-TDA1pr* region, using the existing gene-specific barcodes to measure strain abundances in each library. In this approach, because the pairing sequences are inserted into different genes throughout the genome, the frequency of pairing is confounded by the potential effects of the genomic location of the pairing sequence. Therefore, for each of the 10 TFs we targeted, we included as controls up to 6 neighboring genes, which should have a similar genomic location effect as the targeted gene (**Figure 3.8A and B**). We hypothesized that due to the Rabl orientation of yeast chromosomes, in which centromeres are clustered together (Duan et al., 2010), centromere-proximal regions would interact less with *HAS1pr-TDA1pr*, which is centromere-distal. Indeed, the most centromere-distal gene knockouts interacted ~4-fold more with the *HAS1pr-TDA1pr* locus than the most centromere-proximal gene knockouts (**Figure 3.8C**). Of the 7 TF gene knockouts that were measured in our assay (3 dropped out during library construction, including *LEU3*), six had no substantial difference in pairing strength compared to their genomic neighbors (**Figure 3.8D**). However, deletion of

RGT1 led to a ~20-fold decrease in pairing strength, consistent with Rgt1 binding its cognate motifs in the minimal pairing region (**Figure 3.7B and C**).

Next, we expanded our *trans* MAP-C screen to include the majority of known nonessential TFs with known binding motifs (de Boer and Hughes, 2012). To avoid the confounding effect of genomic location, we inserted a barcoded pairing sequence construct into a fixed locus instead of into the gene knockout locations. We associated each of these barcodes with the cognate knockout by individually transforming each knockout strain with a unique barcode, in a 96-well plate format. We tested a total of 109 TF gene knockouts, as well as 15 nuclear pore complex components, 8 fitness neutral negative controls, and a wild-type control, with multiple barcode replicates for controls and expected hits (**Figure 3.9A**). As expected, most barcoded strains were equally abundant in the 3C and genomic libraries, indicating that the corresponding knockout did not impact *HAS1pr-TDA1pr* pairing. However, *LEU3* and *RGT1* knockouts were depleted ~4-fold from the 3C libraries, suggesting that they are required for pairing (**Figure 3.7B**). In addition, deletion of *MOT3* modestly decreased pairing (~2.5-fold); however, its binding motif is not present in the *HAS1pr-TDA1pr* region, suggesting an indirect role. Two other knockouts, *VHR1* and *CBF1*, appeared to also decrease pairing, but were at low abundances and might reflect noise (**Figure 3.9B**).

Surprisingly, none of the TFs required for pairing in either *trans* MAP-C experiment had a motif matching the sequence CCCCAC (the third cluster of positions required for pairing; **Figure 3.1F, 2B, and Figure 3.8C**). However, two putative TFs with high scoring motif matches, *YPR022C* (*SDD4*) and *YGR067C*, were excluded in the initial screens due to their lack of annotations. Therefore, we repeated our fixed-locus TF knockout screen with a limited set of genes, including the two putative TFs and additional replicates for the hits *MOT3*, *VHR1*, and

CBF1. We found that *SDD4* is indeed required for pairing, suggesting that it is the *trans*-acting factor that binds the CCCCAC motif (**Figure 3.7C**). *MOT3* once again exhibited modest depletion, suggesting a minor, perhaps indirect, role in pairing, whereas *VHR1* and *CBF1* displayed no depletion (**Figure 3.7C**).

The strong concordance between the genomic base-pairs required for pairing (identified by *cis* MAP-C) and the DNA binding motifs of *trans* factors required for pairing (identified by *trans* MAP-C) suggests that Leu3, Sdd4, and Rgt1 bind the *HAS1pr-TDA1pr* to mediate pairing. To test this hypothesis, we performed chromatin immunoprecipitation (ChIP) for the tandem affinity purification (TAP) tagged versions of Leu3, Sdd4, and Rgt1 (Ghaemmaghami et al., 2003), in haploid *S. cerevisiae* yeast under both saturated and exponential growth conditions. By qPCR using two different primer pairs, all three TFs strongly bound to the *HAS1pr-TDA1pr* pairing region in saturated conditions, but near background levels in exponential growth (**Figure 3.7** Pairing factors do not play major roles in HAS1 or TDA1 transcriptional regulation. We then performed ChIP sequencing to determine TF binding genome-wide. Consistent with our quantitative PCR measurements, all three TFs showed robust ChIP-seq peaks at *HAS1pr-TDA1pr* in saturated conditions. In contrast, only Leu3 demonstrated a significant peak in exponential growth, with 20-fold weaker enrichment (2.5-fold vs. 51-fold in saturated conditions; **Figure 3.7F**). For all three TFs, saturated conditions produced ChIP-seq peaks with greater enrichments over the input controls and more robust enrichment of the expected motifs, suggesting generally more extensive DNA binding (**Figure 3.10 and Figure 3.11**). These global trends could be a result of technical artifacts as well as biological differences; however, all samples were treated identically using a protocol not optimized for saturated culture conditions, and the *HAS1pr-TDA1pr* peak showed particularly strong condition-specificity (**Figure 3.10**).

Based on the convergence of *cis* MAP-C, *trans* MAP-C, and ChIP-seq data, we conclude that Leu3, Rgt1, and Sdd4 directly bind to both alleles of the *HAS1pr-TDA1pr* minimal pairing region under saturated growth conditions and thereby mediate inducible interchromosomal contacts between them.

3.3.4 *Combinatorial transcription factor binding specifies strong pairing*

We next explored whether the combination of Leu3, Sdd4, and Rgt1 binding explains the uniqueness of *HAS1pr-TDA1pr* pairing. If the clustered binding of Leu3, Sdd4, and Rgt1 is necessary and sufficient to cause pairing in saturated culture conditions, either 1) no loci other than *HAS1pr-TDA1pr* should have all three TFs bound on both the *S. cerevisiae* and *S. uvarum* copies, or 2) other loci that do have all three TFs bound should also exhibit pairing. Because DNA binding data was only available for the *S. cerevisiae* genome, we tested the first possibility by scanning the *S. cerevisiae* and *S. uvarum* genomes for clusters of the three motifs using permissive thresholds for motif matches and allowing up to 200 bp between motifs (see **Materials and Methods**), and then assessed ChIP-seq data at these clusters. The promoters of four genes, *TDA1* but also *HXT3*, *YKR075C*, and *SKS1*, harbored a motif cluster containing all three motifs in both *S. cerevisiae* and *S. uvarum* (**Figure 3.12A**). We also assessed two additional loci, *MIG1pr* and *ILV2pr*, that lacked one of the three motifs in *S. uvarum*. Of these six loci, *TDA1pr* exhibited the strongest total ChIP-seq signal in saturated conditions (**Figure 3.12B and C**), even when extending the search to motif clusters with only one or two of the three motifs in *S. cerevisiae* (**Figure 3.13**). Furthermore, *YKR075Cpr* was the only other locus with robust binding of all three TFs in *S. cerevisiae*, but the only Leu3 motif in the *S. uvarum* copy of the region overlaps a stronger Rgt1 motif and may not result in Leu3 binding. We also noticed that the *TDA1pr* cluster was the most compact (i.e. shortest maximum distance between motifs);

whether this plays a role in pairing is unclear. We infer that the combinatorial binding of Leu3, Sdd4, and Rgt1 on both homologs specifies inducible homolog pairing in saturated cultures.

Although *TDA1pr* exhibits the strongest combinatorial TF binding, we wondered whether any other motif clusters pair inducibly like the *HAS1pr-TDA1pr* locus, albeit perhaps more weakly. To address this question, we leveraged our previously published Hi-C datasets (Kim et al., 2017), along with new Hi-C experiments for the high-pairing strain background in which we performed our pooled mutant experiments, to compare the strength of pairing at each homologous motif cluster and assess whether this pairing is condition-specific (**Figure 3.14**). We observed that *HXT3pr* appears to form inducible homologous contacts in saturated culture conditions (**Figure 3.12D**), despite weak Sdd4 and no Leu3 binding. We did not detect pairing at any other loci (**Figure 3.14**). Across several strain backgrounds, the *HXT3* promoters exhibited 1.6- to 2.7-fold increased interaction frequencies compared to other similar interchromosomal pairs of loci (at least 480 kb from a centromere and excluding subtelomeric regions) in saturated culture conditions (**Figure 3.12E**), but only at baseline levels during exponential growth in rich medium. This is weaker than the pairing between *HAS1pr-TDA1pr* alleles (**Figure 3.14**), suggesting that combinatorial TF binding is not strictly necessary for inducible pairing but may facilitate particularly strong pairing.

The identification of two pairs of loci, *HAS1pr-TDA1pr* and *HXT3pr*, exhibiting homolog pairing opened the possibility that there might be cross-pairing (nonhomologous contacts) between *HAS1pr-TDA1pr* and *HXT3pr*. To test this possibility, we extracted the interaction frequencies among the four loci (*i.e.* two pairs) from our Hi-C data. In all saturated culture datasets where homolog pairing was present, inter-locus interactions were substantially weaker and similar in frequency to those in non-pairing conditions (**Figure 3.13**). Together with our

previous experiment demonstrating that two identical copies of the minimal pairing region can pair even at non-allelic locations in haploid *S. cerevisiae* (**Figure 3.1C**), these data suggest that the interchromosomal pairing mediated by Leu3, Sdd4, and Rgt1 is sequence-specific beyond the simple presence of the same TF binding sites.

3.3.5 *HAS1pr-TDA1pr pairing is regulated by Rgt1 abundance, recruitment of Tup1/Ssn6, and competing domains*

We next sought to explore the mechanisms that regulate pairing. We hypothesized that TF expression levels might regulate the strength of *HAS1pr-TDA1pr* pairing. To test this hypothesis, we analyzed RNA-seq data for haploid *S. cerevisiae* in pairing and non-pairing conditions (saturated and exponentially growing cultures, respectively) (Kim et al., 2017). *RGT1* and *SDD4* were upregulated in saturated cultures, ~2-fold and ~6-fold, respectively, whereas *LEU3* transcript levels remained constant (**Figure 3.17A**). These results are consistent with the hypothesis that increased transcription of the TFs mediating pairing, particularly *RGT1*, regulates the strength of the pairing interaction.

Based on these results, we wondered whether overexpression of any one of the three proteins would be sufficient to produce pairing in non-saturated culture conditions. We used the Z_3EV estradiol induction system (McIsaac et al., 2014) to individually overexpress *S. cerevisiae* Leu3, Sdd4, or Rgt1 , and measured pairing between the native *HAS1pr-TDA1pr* loci using 3C in *S. cerevisiae* x *S. uvarum* hybrids growing exponentially in rich medium (**Figure 3.17B**). In all three strains, a 2 h estradiol induction led to no increase in pairing strength despite between 2- and 10-fold increases in transcript levels (**Figure 3.18A**). As an alternative test, we used galactose induction to overexpress epitope-tagged *RGT1*, and observed a decrease in pairing strength relative to a strain lacking the overexpression cassette (**Figure 3.18B**). These results are

consistent with no single TF being sufficient for pairing; however, Leu3 and Rgt1 are both known to change in conformation (Sze et al., 1992) or phosphorylation state (Kim et al., 2003) in different conditions, so it remains possible that overexpression of a single TF in the appropriate state suffices for pairing.

Rgt1 is known to interact with several cofactors that affect its DNA binding and transcriptional repression activities: the Tup1/Ssn6 co-repressor complex and the proteins Mth1 and Std1 (Polish et al., 2005). Our experiments thus far had not distinguished whether Rgt1's pairing activity is directly mediated by physical interactions among molecules of Rgt1 or indirectly, through these or other interaction partners. To address this, we performed *trans* MAP-C for individual deletions of these four interacting partners of Rgt1, along with the same positive and negative controls as before (**Figure 3.17C**). In addition, based on the glutamine and asparagine-rich domains present in Sdd4 and Rgt1, we tested deletion of *RNQ1*, a Q/N-rich peptide known to influence the oligomerization of other Q/N-rich proteins (Derkatch et al., 2004). Deletion of *RNQ1* had no effect on pairing, suggesting that *HAS1pr-TDA1pr* pairing is not mediated by Q/N-rich domains. Deletion of *TUP1* or *SSN6* both led to increased pairing, indicating that the recruitment of the Tup1/Ssn6 co-repressor complex inhibits pairing. This is consistent with its known inhibition of Rgt1's DNA binding activity (Roy et al., 2013), which is presumably required for pairing. Deletion of *MTH1* or *STD1* had a minimal negative effect on pairing. These results suggest that the role of Rgt1 in pairing is not simply to recruit cofactors that mediate pairing; instead, its pairing activity may compete with its transcriptional repression activity.

We hypothesized that a particular domain of Rgt1, coupled with its DNA-binding activity, might be responsible for *HAS1pr-TDA1pr* pairing. To test this idea, we generated a

series of 10 amino acid deletions spanning most of the Rgt1 protein (positions 91-1030, excluding several deletions that dropped out during strain construction) and performed *trans* MAP-C to test the pairing function of the Rgt1 mutants (**Figure 3.17D**). Surprisingly, much of the C-terminal half of Rgt1 was required for *HAS1pr-TDA1pr* pairing. These "pairing domains" are closely aligned to the regions of Rgt1 predicted to be highly structured, largely through alpha helices (**Figure 3.17D**). The regions of Rgt1 required for pairing also correspond loosely to those required for regulation of activation vs. repression function through allosteric changes in protein conformation in response to glucose-regulated phosphorylation (Polish et al., 2005) (**Figure 3.19**).

To further compare the roles of Rgt1 protein domains and phosphorylation in *HAS1pr-TDA1pr* pairing, we generated deletions of the Rgt1 zinc finger, Q/N-rich, and C-terminal domains, and phosphodepletion mutants S88A and S758A, and performed *trans* MAP-C to test their pairing function (**Figure 3.17E and F**). As expected, deletion of the zinc finger DNA-binding domain or the C-terminal domain led to background pairing levels, equivalent to lacking Rgt1 altogether. Surprisingly, deletion of the Q/N-rich domain led to stronger pairing than the wild-type control (with *RGT1* integrated at the same ectopic location), suggesting that the Q/N-rich domain inhibits pairing. Both the S88A and S758A mutations had a weak decrease in pairing, indicating that although these mutations are capable of disrupting Rgt1 activator function and intramolecular interactions (Polish et al., 2005), they do not individually play major roles in regulation of pairing in saturated cultures.

Together, our results suggest multiple potential modes by which the transcription factor Rgt1 regulates *HAS1pr-TDA1pr* pairing: its own expression level, recruitment of Tup1/Ssn6, and the competing activities of its Q/N-rich and C-terminal regulatory domains.

3.3.6 *Pairing factors do not play major roles in HAS1 or TDA1 transcriptional regulation*

We have thus far characterized the molecular mechanism and regulation of *HAS1pr-TDA1pr* homologous pairing in saturated cultures. However, the question of why this pairing occurs (i.e. what biological function it serves, if any) remains outstanding. Now that we have identified the precise DNA base pairs and proteins involved in *HAS1pr-TDA1pr* pairing, we are equipped to test whether Leu3, Sdd4, and Rgt1 play a role in transcription at *HAS1* or *TDA1* in saturated cultures. Furthermore, we can employ *S. cerevisiae x S. uvarum* hybrids, in which we can discriminate the two homologous copies of each gene, in order to distinguish the effects on transcription of disrupting the pairing DNA sequence *in cis* (presumably through impaired *cis* regulation) vs. *in trans* (presumably through impaired pairing). To this end, we generated mutant hybrid strains carrying a wild-type copy of *S. uvarum HAS1pr-TDA1pr* and a copy of *S. cerevisiae HAS1pr-TDA1pr* with either a wild-type (WT) genotype, mutated Leu3 binding site (leu3), or two mutated Rgt1 sites (rgt1x2) (**Figure 3.20A**). We then performed RNA-seq in saturated culture conditions. The Rgt1 binding site mutations led to higher levels of *S. cerevisiae TDA1*, consistent with Rgt1 acting as a repressor under low glucose conditions (**Figure 3.20B**), but had no significant effect on either *S. uvarum HAS1* or *TDA1*. Surprisingly, despite the conserved binding site and strong ChIP-seq signal for Leu3 in *HAS1pr-TDA1pr*, disrupting its binding site had no significant effect on the transcript levels of any gene (**Figure 3.20C**).

To further characterize the transcriptional roles of Leu3, Sdd4, and Rgt1 in saturated culture, we also performed RNA-seq on deletion strains for each of these genes, along with a wild-type control, grown to saturation. As expected, each deleted gene was highly down-regulated in the mutant strains (**Figure 3.21A**). In addition, many known targets of Rgt1 and genes downstream of ChIP-seq peaks were highly upregulated in *rgt1*Δ (**Figure 3.22**). Focusing

on the *HAS1-TDA1* locus, none of the deletion strains had altered *HAS1* expression, and only the *rgt1*Δ strain had even slightly altered *TDA1* expression (**Figure 3.20D**). These results are consistent with our earlier results from mutating TF binding sites, and together indicate that despite the importance of Leu3, Sdd4, and Rgt1 binding for pairing at the *HAS1pr-TDA1pr* locus in saturated culture conditions, they do not play a major role in transcriptional regulation as detected by standard polyA mRNA sequencing.

3.4 DISCUSSION

In summary, we developed Mutation Analysis in Pools by Chromosome conformation capture (MAP-C), a method to simultaneously test hundreds of mutations for their effects on a chromosomal contact of interest. MAP-C can be used to identify the precise sequences that are necessary for a contact (*cis* MAP-C) as well as the factors that are necessary to mediate the contact (*trans* MAP-C). Here we applied both versions of MAP-C to dissect the mechanism of inducible interchromosomal pairing between *HAS1pr-TDA1pr* alleles in budding yeast. Using a combination of gain-of-function and loss-of-function screens, we demonstrate that a trio of transcription factors—Leu3, Sdd4, and Rgt1—mediates pairing between clusters of binding sites.

Our results begin to elucidate the mechanisms of condition-specific interchromosomal contacts and homolog pairing, which have often been elusive (Mirkin et al., 2013). We have not yet fully defined the biochemical mechanism of pairing—it is possible the TFs interact directly and/or indirectly—but so far, no known interaction partners or cofactors have proven essential for this interaction. Unlike more prevalent nuclear-pore mediated gene relocalization and homolog pairing (Brickner et al., 2012; Randise-Hinchliff and Brickner, 2016), *HAS1pr-TDA1pr* pairing in saturated cultures does not appear to require the nuclear pore complex (**Figure 3.7B**).

Instead, the Tup1/Ssn6 repressor complex recruited by Rgt1 negatively regulates pairing (**Figure 3.17C**), perhaps by inhibiting the DNA-binding activity of Rgt1 (Roy et al., 2013). Within Rgt1, the zinc finger DNA-binding domain and C-terminal regulatory domain are required for pairing, while the Q/N-rich region, which contains the activation domain (Polish et al., 2005), inhibits pairing (**Figure 3.17F**). These results point toward specific interactions among structured domains mediating pairing, rather than phase separation by intrinsically disordered regions.

Intriguingly, Leu3 also contains a regulatory domain whose conformation responds to environmental cues (Sze et al., 1992; Wang et al., 1997). We speculate that these regulatory domains, which are capable of intramolecular interactions (Polish et al., 2005), may mediate interchromosomal pairing. Furthermore, if these regulatory domains are indeed capable of direct dimerization or oligomerization, they may mediate not only interchromosomal contacts, but also cooperative TF binding at clusters of binding sites (Kim et al., 2003). We note that of Leu3, Sdd4, and Rgt1, only Leu3 is known to form strong dimers at each of its binding sites. However, even weak homotypic intermolecular interactions could increase the duration of interchromosomal contacts as well as increase local TF concentrations, and thereby increase DNA binding activity. It remains unclear whether heterotypic interactions among Leu3, Sdd4, and Rgt1 occur.

Another question is whether the contacts are stoichiometric, *e.g.* one-to-one, or instead mediated by aggregation of one or more TFs via weak interactions. Notably, we observe homotypic pairing between homologs of *HAS1pr-TDA1pr* and *HXT3pr* but not heterotypic pairing (**Figure 3.15**), suggesting that Rgt1 does not indiscriminately form contacts between all of its binding sites. However, this is not because the entire homologous context is necessary for pairing, as we found that the pairing sequence from *HAS1pr-TDA1pr* is sufficient to induce

ectopic pairing in haploid *S. cerevisiae* (**Figure 3.1C**). Instead, a possible explanation is that only pairs of loci with a series of motifs in similar orders and orientations form frequent contacts, consistent with a stoichiometric interaction model. This added specificity beyond the presence of TF motifs could explain the lack of pairing between *HAS1pr-TDA1pr* and other sites of Leu3, Sdd4, and Rgt1 binding (**Figure 3.16**).

Despite our detailed molecular characterization of *HAS1pr-TDA1pr* homolog pairing, we still do not know its biological function, if any. We hypothesized that it might contribute to transcriptional regulation at either *HAS1* or *TDA1*. In particular, *TDA1* is a kinase that phosphorylates Hxk2, the main hexokinase in yeast, and thereby inhibits its ability to mediate glucose repression together with Mig1 (Kaps et al., 2015; Kettner et al., 2012) We speculated that pairing might help activate *TDA1*, which is mildly upregulated at the transcriptional level in saturated conditions (Kim et al., 2017). However, we find that Leu3 and Sdd4 play no detectable role in regulating transcript levels of either gene in saturated culture conditions, while Rgt1 mildly represses *TDA1* in *cis* but not in *trans* (**Figure 3.20**). There are several possible reasons why we may not have detected a phenotypic effect from disrupting the pairing region or factors: 1) even low levels of pairing are sufficient to mediate its biological function; 2) its function is not in saturated culture conditions per se, but affects either exit from or re-entry into those conditions, similar to transcriptional memory at *GAL1* and *INO1* (Brickner et al., 2015; Sood et al., 2017); 3) its function only affects the dynamics of transcription (or other processes) but not steady-state RNA levels (Zhang and Bai, 2016). However, we also acknowledge the possibility that there is no function *per se*, but that pairing is instead a side-effect of conservation of biochemical features of the pairing TFs. If the same domains are required for both cooperative binding and pairing, then evolutionarily conserved pairing could result from selection on

cooperative binding. Interestingly, the TF Hsf1, which forms trimers, and in flies also exhibits cooperativity between trimers (Xiao et al., 1991), was also recently shown to mediate interchromosomal contacts in yeast (Chowdhary et al., 2019). More studies of the biochemical basis of cooperative TF binding are required to test this hypothesis.

Is *HAS1pr-TDA1pr* pairing unique? Based on Hi-C, we previously found *HAS1pr-TDA1pr* to exhibit the strongest interchromosomal interactions in saturated cultures of *S. cerevisiae* x *S. uvarum* hybrids, excluding centromeres, telomeres, and the chromosomes carrying the rDNA arrays (Kim et al., 2017). Using ChIP-seq and motif analyses, we found that a requirement for robust adjacent DNA-binding by Leu3, Sdd4, and Rgt1 on both the *S. cerevisiae* and *S. uvarum* copies was sufficient to identify *HAS1pr-TDA1pr*. However, we also expect that the limited resolution of Hi-C and the sequence divergence in interspecific hybrids would both limit our sensitivity for detecting homolog pairing. We hypothesize that more cases of localized pairing exist, as we found with *HXT3pr* (**Figure 3.12D and E**). We imagine that other TFs capable of interchromosomal contacts, like Hsf1 (Chowdhary et al., 2019), are also capable of mediating homolog pairing, and the condition-specificity of these contacts suggests that other conditions may also exhibit similar contacts but have not yet been explored.

Although our study was focused on *HAS1pr-TDA1pr* pairing in budding yeast, both *cis* and *trans* MAP-C should be applicable to other loci and organisms. The main constraints in experimental design are that 1) the introduced mutations or associated barcodes must be included in the 3C PCR product, and 2) the region of interest should not be digested by the restriction enzyme. As we have implemented this approach, the region targeted by saturation mutagenesis was limited to <250 bp to allow for Illumina sequencing, but this could be extended using either barcode association, similar to our *trans* knockout screens, or long-read sequencing methods.

Also, we focused on a single pair-wise interaction, but it is also possible to assay a mutant pool for multiple interactions, by using multiple primer pairs. MAP-C should be applicable to intrachromosomal contacts as well as interchromosomal ones, albeit with a potentially higher background of nonspecific contacts. It would be interesting to apply MAP-C to dissect the *cis* and *trans* regulators of enhancer-promoter loops in mammals, and thereby distinguish the contributions of cohesin and CTCF (Guo et al., 2012), general looping factors like YY1 (Weintraub et al., 2017), site-specific transcription factors (Nolis et al., 2009), and other cofactors. *Trans* MAP-C could also allow mutational scanning of TFs to clarify the biochemical mechanisms by which these transcription factors mediate chromosomal contacts.

MAP-C leverages the high throughput of saturation mutagenesis and mutant collections to allow systematic dissection of chromosome conformation. We tested up to ~1,000 variants at a time, but with larger-scale experiments, it should be possible to test even more variants. A major potential strength of our approach is that unlike cellular high-throughput genetic screens (Fowler and Fields, 2014; Gasperini et al., 2016; Shalem et al., 2015), it resolves the functional consequences of mutations at the allelic level, and thus is not confounded by heterozygosity (Patwardhan et al., 2009). As we continue to map chromosome conformation at high resolution across ever-expanding numbers of cell types and conditions, MAP-C will provide a scalable approach to dissect the molecular mechanisms underlying specific contacts.

3.5 Materials and Methods

3.5.1 *Yeast strains and culture*

Yeast strains used in this study are described in the **Key Resources Table**. Yeast were cultured at 30C, with the exception of *S. uvarum* strains, which were grown at room temperature.

Cultures were grown shaking overnight to OD_{600} > 5 for saturated culture samples, or diluted to OD_{600} ~ 0.125 and grown to OD_{600} = 0.5-0.8 for exponential growth samples. Estradiol inductions were performed by addition of beta-estradiol to 1 μM final concentration (or equivalent volume of ethanol for negative control) to OD_{600} = 0.5 cultures grown in YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v dextrose) and grown for 2 hours. Galactose induction was performed by growth in synthetic complete medium (without uracil for selection for overexpression plasmid) with 2% v/v raffinose to OD_{600} = 0.75 followed by addition of 2% galactose and subsequent growth for 1.5 hours. For comparison, yeast were grown in synthetic complete medium with or without uracil with 2% glucose or 2% raffinose. *S. cerevisiae* x *S. uvarum* hybrids were generated by standard mating and auxotrophic or drug selection procedures. Yeast transformations were performed using a modified Gietz LiAc method (Pan et al., 2004).

3.5.2 *Mutant library generation*

Subsequences. All 178 bp subsequences of the intergenic region between the *S. cerevisiae HAS1* and *TDA1* coding sequences were synthesized in an Agilent (Agilent Technologies, Santa Clara, CA) array-synthesized oligonucleotide pool (sequences included in **Supplementary file 1**), and then amplified and cloned into a vector just downstream of a KanMX cassette followed by a DpnII restriction site (GATC) using NEBuilder HiFi (New England Biolabs, Ipswich, MA). These plasmids, which contain homology to the *S. cerevisiae HAS1* and *TDA1* sequences, were linearized by restriction digestion and transformed into YMD3919 (deletion of *S. cerevisiae has1pr-tda1pr*). Transformants (typically ~5,000 per experiment) were selected on G418 medium for a total of at least 4 days (including two rounds of scraping and replating a portion

onto a fresh plate), and then used to inoculate a 50 ml culture in YPD (1% w/v yeast extract, 2% w/v peptone, 2% w/v dextrose).

Error-prone PCR. The primers TTCACCGCCTGCTATCATCC and GAATCGGCGGAATAACCTAACACG were used in a PCR reaction using the Agilent GeneMorph II kit, using as template 0.1 ng of a PCR product generated with the same primers from *S. cerevisiae* genomic DNA. The error-prone PCR products were then cloned, transformed, and selected as described above. The library contained ~63,000 transformants.

Programmed 3 bp substitutions. For each set of 3 consecutive base-pairs between positions 532-675 (inclusive), we randomly chose three trinucleotide substitutions such that among them, each nucleotide change (*e.g.* A->C) was included once at each position and so that no DpnII restriction sites (GATC) were created. These sequences were synthesized on an array (sequences included in **Supplementary file 1**) and processed as described above. The library contained ~38,000 transformants.

In-gene knockout screen. The selected gene knockout strains from the MATalpha yeast deletion collection (Giaever et al., 2002), in addition to the knockouts of the three nearest genes on either side excluding those absent or known to be slow-growing in the yeast deletion collection (**Figure 3.8**), were pooled and transformed *en masse* using a PCR construct with homology arms for the TEFb promoter and terminator from the KanMX deletion cassette flanking the pairing sequence (**Figure 3.2B**), an EcoRI restriction site, and the *URA3* selectable marker gene. The entire pool, containing ~12,500 transformants, was selected on synthetic complete medium without uracil and processed as described above.

Fixed-locus knockout screens. TF genes were defined as genes with motifs on YeTFaSCo (de Boer and Hughes, 2012) that are not annotated as being part of a complex. Genes with no

non-systematic name in the GFF file from the Saccharomyces Genome Database (SGD; version R64.2.1) were excluded. In addition, we included knockouts of genes described as nuclear pore components in SGD, the wild type strain and knockouts of eight genes known to have minimal fitness consequences under multiple nutrient limitation conditions (Payen et al., 2016). The selected gene knockout strains from the MATa yeast deletion collection carrying the synthetic genetic array (SGA) reporter (Tong et al., 2001) (excluding those failing quality control or known to grow slowly) were grown in separate wells of deep 96-well plates, and transformed in 96-well format (**Supplementary file 3**) with a PCR construct similar to that used in the in-gene knockout screen, but with a unique 12 bp barcode added upstream of the pairing sequence (primer sequences in **Supplementary file 1**), and homology to the *YDR535C* coding sequence. Colonies were picked and verified by PCR, and one successful clone from each strain was pooled together and then diluted and grown overnight in YPD. Colonies from positive and negative control strains were repooled with new strains for subsequent fixed-locus *trans* MAP-C experiments.

Interactor knockout screen. The selected gene knockout strains from the MATa yeast deletion collection with the SGA reporter (Tong et al., 2001) were each transformed with a cocktail of constructs (derived from fixed-locus screen construct) with 14 different barcodes each, and then at least 8 colonies were Sanger sequenced, and three colonies from each strain carrying different barcodes were pooled and processed as described above, with positive and negative control strains included.

Rgt1 deletion scan. The *S. cerevisiae RGT1* gene was amplified with 537 bp of upstream promoter sequence and 221 bp of downstream terminator sequence, and cloned downstream of *URA3* in the same vector used for the fixed-locus knockout screens. We attempted to create all

10 amino acid (30 bp in-frame) deletions between amino acids 91 and 1040; we successfully created 87 of the 96 attempted deletions as plasmids. The wild-type *RGT1*-containing plasmid was amplified in two pieces, each using one primer in the ampicillin resistance gene and one within *RGT1* so that the resulting pieces each contain homology for the other. These pieces were combined using NEBuilder HiFi, and then transformed into *E. coli*, extracted, and verified for the deletion junction by Sanger sequencing. Each deletion plasmid was amplified with the same primers used to create the fixed-locus knockout constructs, with a different chosen barcode for each deletion. These PCR products were pooled and transformed *en masse* into the *rgt1* deletion strain from the MATa yeast deletion collection with the SGA reporter, in two replicates. Each pool was selected on synthetic complete medium without uracil including G418 and processed as described above.

Rgt1 domain deletion and phosphorylation site mutations. Tested mutations included deletion of the zinc finger domain (amino acids 32-90), Q/N-rich domain (amino acids 293-459), C-terminal domain (amino acids 461-1163), and S88A (TCG->GCG) and S758A (TCC->GCC). Mutant versions of the wild-type *RGT1*-containing plasmid were amplified and cloned as for the 10 amino acid deletion scan of Rgt1, and then transformed using mixtures of 14 different barcodes per desired mutation and Sanger sequenced as in the Rgt1 interactor knockout screen. Five distinct barcoded versions of each mutant were pooled and processed as above, with positive and negative control strains included.

3.5.3 *3C*

Cells were crosslinked by addition of 37% formaldehyde to a final concentration of 1% (v/v) and incubation at room temperature for 20 min, quenched by addition of 2.5M glycine to a final concentration of 150 mM and incubation at room temperature for 5 min, and then washed in

1x Tris-buffered saline (TBS) and stored as a pellet at -80C in aliquots of 50-100 µl dry pellets. Cells were lysed by vortexing in lysis buffer (TBS + 1% Triton X-100 with Pierce EDTA-free protease inhibitor tablet (Thermo Fisher Scientific, Waltham, MA)) with 500 µm acid-washed glass beads for 6 cycles of 2 min, with 2 min on ice between cycles. The lysate was collected by puncturing the bottom of each tube and then centrifuging the tube, stacked on top of an empty tube. The lysate was then washed in lysis buffer, then TBS, and finally resuspended in 10 mM Tris pH 8.0 to a volume of ~200 µl per 25 µl of starting dry pellet volume. A single 200 µl aliquot of lysate was then precleared by addition of 0.2% SDS and incubation at 65C for 10 min, cooled on ice, quenched by addition of 1% Triton X-100 (v/v), and then digested overnight with at least 100U of restriction enzyme (200U DpnII for *cis* experiments and galactose induction, 400U EcoRI-HF for *trans* experiments, and 100U NlaIII for estradiol inductions). The restriction digest was heat-inactivated at 65C for 20 min in the presence of 1.3% SDS, and then chilled on ice and added to a dilute ligation reaction in 4 ml volume with 1% Triton X-100, 1x T4 DNA Ligase Buffer (NEB), and 10,000U of T4 DNA ligase (NEB) and incubated at room temperature for 4 hours. The ligation products were reverse-crosslinked with proteinase K at 65C overnight, and then purified by phenol-chloroform extraction followed by clean-up on a Zymo DNA Clean & Concentrator-5 column (Zymo Research, Irvine, CA). The resulting 3C DNA was quantified using a Qubit. Each technical replicate was processed separately beginning with cell lysis.

3.5.4 *MAP-C library preparation and sequencing*

3C libraries were prepared by amplification of up to 8 reactions of 50 ng 3C DNA per replicate, using primer pairs specific to the chromosomal contact of interest (for pairing library) or a control off-target chromosomal contact (for off-target library), for 24-32 cycles. Genomic libraries were prepared by amplification of up to 4 reactions of 50 ng of either 3C DNA or

genomic DNA using primers flanking the targeted mutations or barcodes, for 17-22 cycles. Reactions for each replicate were pooled, purified by Ampure XP beads (Beckman Coulter Life Sciences, Brea, CA), and then re-amplified with primers flanking the mutagenized or barcode region and including sequencing adapter sequences for 5-8 cycles, and then again with primers adding sample indices and Illumina flow-cell adapters for 6-9 cycles. All reactions were prepared with KAPA HiFi HotStart ReadyMix (Roche, Basel, Switzerland) with recommended thermocycling conditions, and included 0.5x SYBR Green I to monitor amplification by quantitative PCR and minimize the number of PCR cycles. The final libraries were sequenced on an Illumina MiSeq or Nextseq 500 (Illumina, San Diego, CA) using paired-end sequencing. See **Supplementary file 2** for detailed information on each library.

3.5.5 *MAP-C sequencing analysis*

Paired-end reads were merged and adapter-trimmed using PEAR (Zhang et al., 2014), except for *trans* knockout experiments, in which only read 1 was used. These reads were then trimmed of the first 4 bp (corresponding to a randomized region for Illumina clustering purposes) and mapped using Bowtie 2 (Langmead and Salzberg, 2012) Reads were mapped to the *S. cerevisiae HAS1pr-TDA1pr* region, and then the read coverage was calculated using bedtools (Quinlan and Hall, 2010).

Error-prone PCR. Reads were mapped to the wild-type sequence of the mutagenized region. The resulting alignments were scored for number of substitutions, insertions, and deletions, and the fraction of reads with a substitution at each position were calculated.

3 bp substitutions. Reads were mapped to the wild-type sequence of the mutagenized region. The resulting alignments were scored for number of substitutions, insertions, and deletions, and the fraction of reads with a substitution at each position were calculated, excluding

reads with fewer than 3 substitutions (which correspond to PCR or sequencing errors of the wild-type sequence).

Trans knockout screens, Rgt1 deletion scan, and Rgt1 mutant screen. Reads were mapped to all 192 possible barcode sequences, and the normalized fraction of reads mapping to a given barcode with a MAPQ ≥ 20 in the 3C library compared to the genomic control was calculated for each replicate.

3.5.6 *3C qPCR*

3C DNA was amplified using the same conditions as MAP-C libraries, in three replicates per primer pair. The pairing 3C products were normalized to the off-target (intrachromosomal) 3C products, assuming 2-fold amplification per cycle.

3.5.7 *Chromatin immunoprecipitation*

Either 50 ml of exponentially growing (OD_{600} = 0.7-0.8) or 12 ml of saturated cultures of TAP-tagged strains (Ghaemmaghami et al., 2003) were crosslinked and lysed as for 3C, but for 15 minutes in FA lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1x Pierce EDTA-free protease inhibitor tablet),. Lysates were pelleted and resuspended in FA lysis buffer, sonicated using a Diagenode Bioruptor (Diagenode, Liege, Belgium) for 3 cycles of 10 min on the HIGH power setting, with 30 s cycles on and 30 s on, and cleared by centrifugation at 20000g for 10 min. An aliquot of 50 µl supernatant was saved for input, and the remaining sample was incubated with 10 µl of Dynabeads Pan Mouse IgG magnetic beads (pre-washed with FA lysis buffer) rotating overnight at 4C. The immunoprecipitations were washed twice with FA lysis buffer, once with high salt (500 mM NaCl) FA lysis buffer, twice with RIPA buffer (10 mM Tris-HCl pH 8, 250 mM LiCl,

0.5% Igepal CA-630, 0.5% sodium deoxycholate, 1 mM EDTA), and once with TE (50 mM Tris-HCl pH 8, 1 mM EDTA), and then eluted first with 100 µl TE + 1% SDS at 65C for 15 min, and a second time with 150 µl TE + 0.67% SDS. Inputs were diluted with 200 µl of TE + 1% SDS, and all samples were treated with 50 mg RNase A at 37C for 10 min and 100 mg proteinase K at 42C for 1 hr, and then reverse crosslinked overnight at 65C. DNA was purified with a Zymo ChIP DNA Clean & Concentrator-5 kit and eluted in 15 µl of 10 mM Tris-HCl pH 8.

qPCRs were performed using 5 µl of either 1:20 dilution of IP samples or 1:800 dilution of input samples, in 25 µl reactions with KAPA Robust 2G HotStart ReadyMix, using 0.5x SYBR Green I and standard cycling conditions for 40 cycles on a Bio-Rad C1000 Touch thermal cycler with a CFX96 Real-Time System (Bio-Rad Laboratories, Hercules, CA). Cq values were calculated using the Bio-Rad CFX Manager 3.1 software, using the single threshold mode for Cq calculation and baseline subtracted curve fitting. Primer efficiencies were calculated using a 5-fold dilution series of genomic DNA from *S. cerevisiae* BY4741 starting from 20 ng. Primer sequences are included in **Supplementary file 1**.

ChIP-seq libraries were prepared using Swift Accel-NGS 2S Plus (Swift Biosciences, Ann Arbor, MI) dual-indexed kits using 10 µl of IP samples or 1 ng of input samples, with 9 cycles of PCR for input samples and 12-15 cycles for IP samples. Libraries were sequenced to ~2.5-6 million read pairs per sample using 2 x 37 bp reads on an Illumina NextSeq 500.

3.5.8 *ChIP-seq analysis*

Sequencing reads were first pre-processed using cutadapt (Martin, 2011): reads were quality-trimmed (option -q 20), trimmed of adapter sequences, excluding any read pairs in which either read was shorter than 28 bp after trimming (option -m 28). Read pairs were mapped to the

sacCer3 *S. cerevisiae* reference genome using Bowtie 2 (Langmead and Salzberg, 2012) with the --very-sensitive parameter set, requiring the reads in each pair to be with 2000 bp of each other (option -X 2000). Read pairs in which both reads had a mapping quality score of at least 30 were deduplicated using samtools rmdup. Replicates were merged prior to calling peaks and generating fold enrichment tracks using MACS2 (https://github.com/taoliu/MACS) (Zhang et al., 2008). Fold enrichment tracks were visualized using the UCSC Genome Browser (Karolchik et al., 2003).

3.5.9 *RNA sequencing*

Yeast strains were grown overnight in YPD in biological triplicate (independent colonies, or for newly generated transformants, independent transformants), pelleted and then stored at -80C. RNA was purified using acid phenol extraction, and then treated with Turbo DNase and purified with a Qiagen RNeasy Mini kit (Qiagen, Hilden, Germany). Illumina libraries were prepared from 800-900 ng of total RNA, using the Illumina Truseq RNA Library Prep kit v2 (for TF binding site mutants) or the Illumina Truseq Stranded mRNA Library Prep kit (for TF knockouts). Libraries were sequenced to ~11-13 million read pairs per sample using 2 x 37 bp reads on an Illumina NextSeq 500.

3.5.10 *RNA-seq analysis*

Reads were preprocessed and mapped as with the ChIP-seq libraries, but with the -X 500 option for Bowtie 2. Read pairs in which both reads had a mapping quality score of at least 30 were overlapped with annotated genes using HTSeq (Anders et al., 2015). Global fold-change analyses were using DESeq2 (Love et al., 2014a).

3.5.11 *RT-qPCR*

Yeast were grown as described above in **Yeast strains and culture**, and then pelleted and stored at -80C. RNA was purified using acid phenol extraction, and then treated with Turbo DNase and purified with a Qiagen RNeasy Mini kit. For each sample, 1 ug of total RNA was annealed to oligo$(dT)_{20}$ (13 µl reaction with 1 µl of 50 µM oligo(dT), 1 µl 10 mM dNTP mix) by incubating at 65C for 5 min and then on ice for 1 min. Reverse transcription was performed with SuperScript IV by adding 4 µl of 5x SuperScript IV buffer, 1 µl SUPERase In RNase inhibitor, 1 µl 100 mM DTT, and 1 µl of SuperScript IV enzyme, and then incubating at 50C for 10 min and then 80C for 10 min. For qPCRs, 2.5 µl of each reverse transcription reaction was used for each 25 µl PCR, using KAPA Robust 2G HotStart ReadyMix with standard cycling conditions (except annealed at 55C) for 40 cycles on a Bio-Rad C1000 Touch thermal cycler with a CFX96 Real-Time System. Cq values were calculated using the Bio-Rad CFX Manager 3.1 software, using the regression mode for Cq calculation and baseline subtracted curve fitting. Primer efficiencies were calculated using a 5-fold dilution series of genomic DNA from *S. cerevisiae* BY4741 starting from 10 ng. Primers were specific to *S. cerevisiae* (at least 2 substitutions to *S. uvarum*). See **Supplementary file 1** for primer sequences.

3.5.12 *Rgt1 protein annotations*

Predicted intrinsic disorder was calculated using IUPred2 long disorder (Mészáros et al., 2018). Predicted secondary structure was calculated using Jpred4 (Drozdetskiy et al., 2015) separately on the first 400 amino acids and on the remaining 770 amino acids, as submissions are capped at 800 amino acids.

3.5.13 *Motif analysis*

Systematic scans of motifs were performed using YeTFaSCo (de Boer and Hughes, 2012) with the expert-curated no dubious motif set. Sequence logos were generated using ggseqlogo (Wagih, 2017).

3.5.14 *Motif cluster analysis*

Motif clusters were identified using MCAST (Grant et al., 2015) with a motif p-value threshold of 0.001, a maximum gap threshold of 200 bp, and an E-value threshold of 2, using the high-confidence motifs for Leu3 (#781), Rgt1 (#2227), and Sdd4 (#588) from YeTFaSCo (de Boer and Hughes, 2012). Individual motif occurrences for each TF were scored for the *S. cerevisiae* (version R64.2.1) and *S. uvarum* genomes (as revised in (Kim et al., 2017)) using FIMO (Grant et al., 2011) with a p-value threshold of 0.001 and the option —max-strand. Motif clusters were named based on the nearest downstream gene (on - strand if coordinate of gene < coordinate of motif cluster, and on + strand if coordinate of gene > coordinate of motif cluster). Motif clusters were defined to be homologous if they were upstream of homologous genes.

3.5.15 *De novo motif discovery*

De novo motif discovery was performed using MEME version 4.12.0 (Bailey et al., 2006). For all analyses, 100 bp centered at each ChIP-seq peak or tRNA gene was used. Motifs were allowed to be between 6 bp and either 10 bp or 20 bp, and the top 3 motifs were analyzed, with otherwise default settings.

3.5.16 *Hi-C*

Hi-C was performed and analyzed as in (Kim et al., 2017) using the restriction enzyme Sau3AI.

3.5.17 *Code availability*

Code used to analyze data and generate figures are available at https://github.com/shendurelab/MAP-C.

3.5.18 *Data availability*

All sequencing data have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE118118. Hi-C data from **Figure 3** and **Figure 3.14, Figure 3.15, and Figure 3.16** and RNA-seq data from **Figure 3.17** are from GEO accession number GSE88952. Processed microarray data of gene expression in TF deletions under exponential growth from **Figure 3.21 and Figure 3.22** are from GEO accession number GSE4654 (Hu et al., 2007).

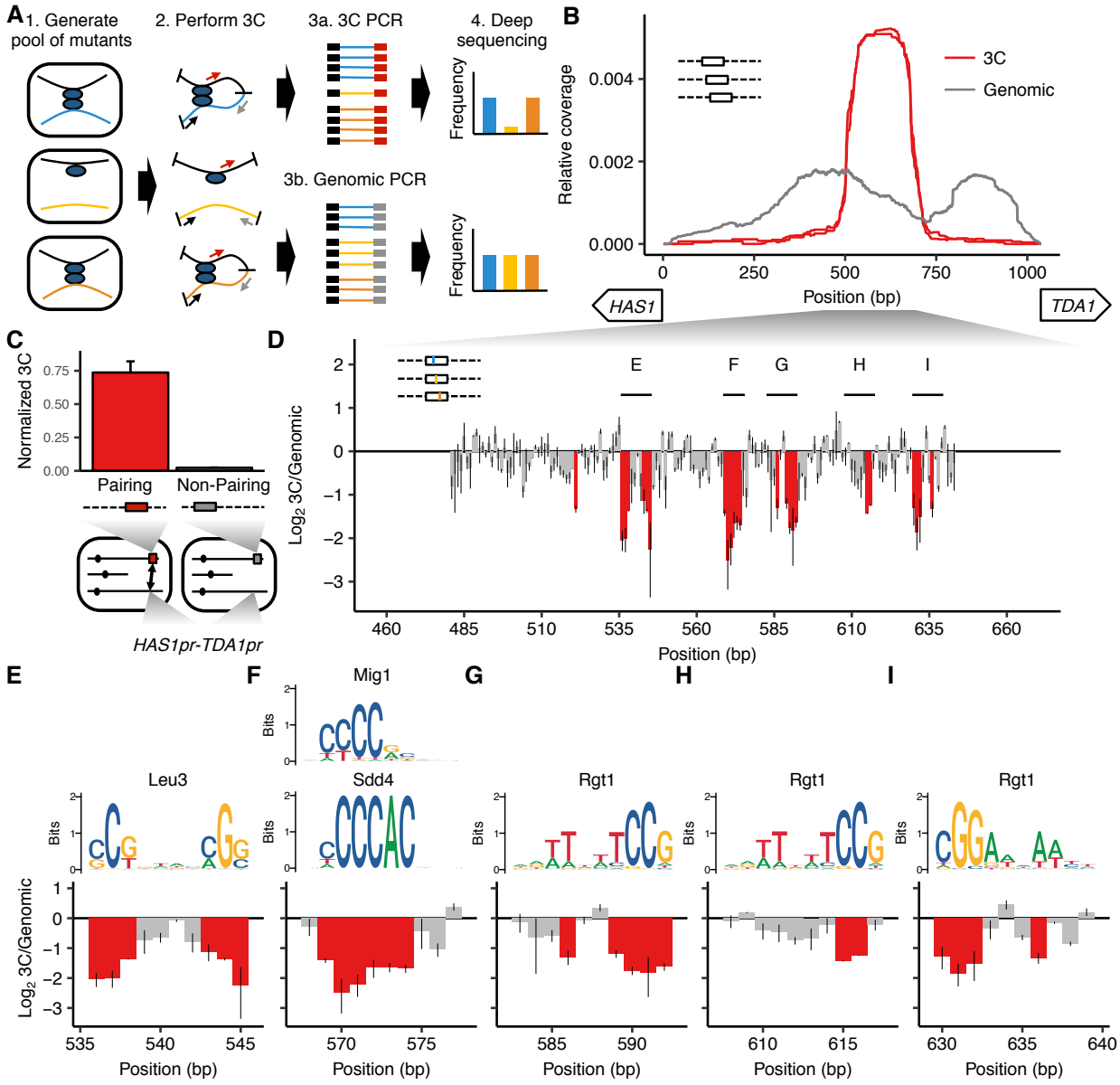

Figure 3.1. MAP-C identifies DNA sequences necessary and sufficient for inducible pairing between *HAS1pr-TDA1pr* alleles.

(A) In the *cis* MAP-C method, mutations in a ~250 bp segment of the genome are assessed for their effect on a specific 3D contact of that segment. Colored lines indicate mutant DNA sequences, and thin arrows indicate primers. (B) A ~150 bp region is sufficient for interchromosomal pairing. *cis* MAP-C was used to test 178 bp subsequences from the *S. cerevisiae HAS1pr-TDA1pr* region for pairing with the *S. uvarum HAS1pr-TDA1pr*. Shown are read coverage of the 3C (red) and genomic (gray) libraries, normalized to sum to 1. The two lines for each color represent technical replicates. Start positions and orientations of *HAS1* and *TDA1* coding sequences are shown on x-axis. (C) A minimal pairing region is sufficient for

ectopic pairing. Shown are contact frequencies between *HAS1pr-TDA1pr* and a pairing (red) or non-pairing (gray) sequence (coordinates shown in **Figure 3.2C**) integrated at the *FIT1* locus in haploid *S. cerevisiae*, as measured by 3C, normalized to contacts between *FIT1* and *HLR1* 10 kb away. Bars indicate mean ± s.d. of technical triplicates. (D) Base-pairs necessary for pairing, shown as ratio of the total substitution frequency at each position in the 3C library compared to the genomic library. Error bars indicate the two technical replicates. Positions most strongly required for pairing ($\log_2$ 3C/Genomic < -1.1) are highlighted in red. (E-I) Selected regions from panel D are highlighted, with sequence logos for matching transcription factor motifs. See **Figure 3.3 and Figure 3.4** for full set of overlapping motifs.

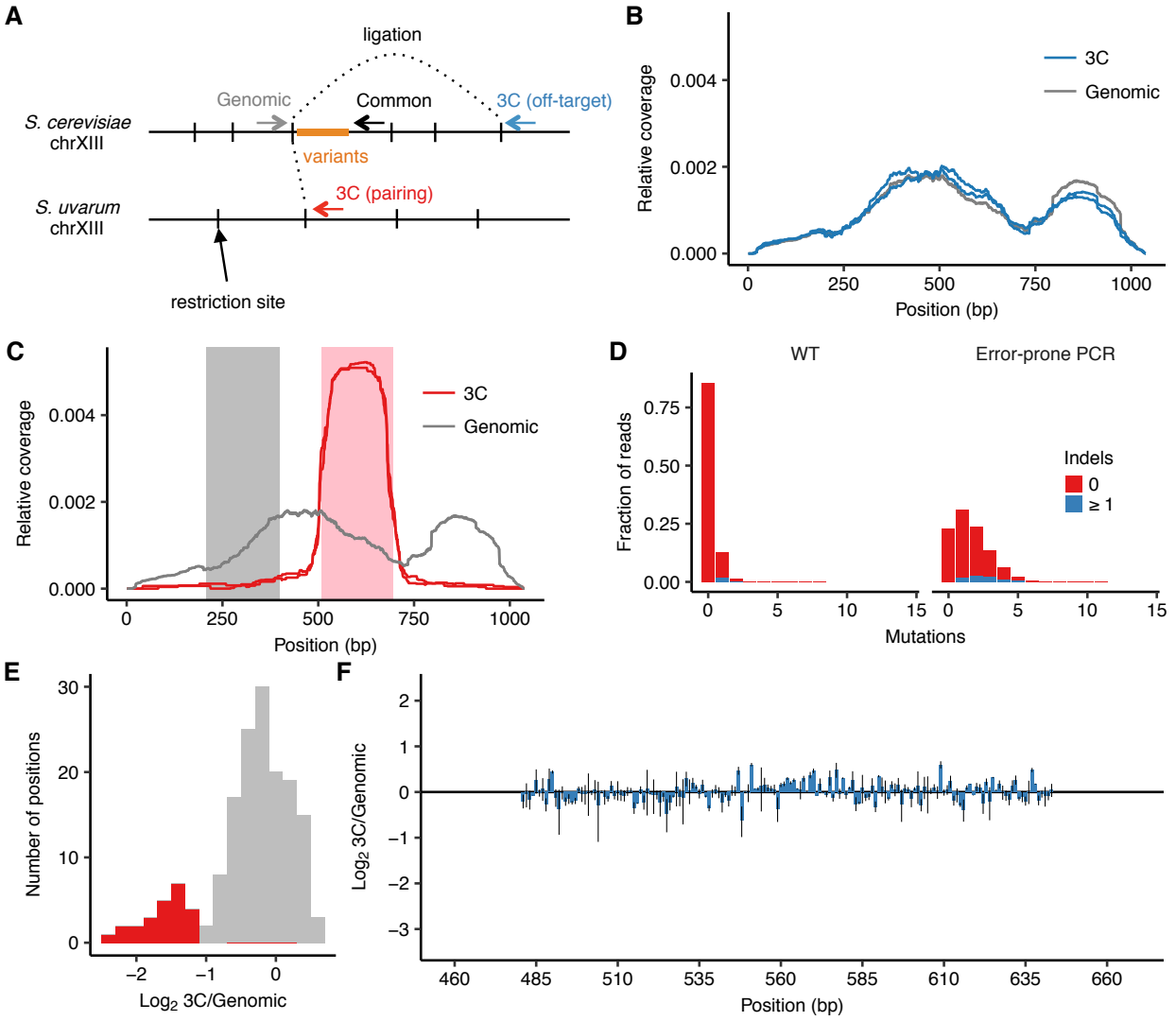

Figure 3.2. Design and controls for using *cis* MAP-C to dissect *HAS1pr-TDA1pr* pairing.

(A) Schematic (not to scale) of primer locations for *cis* MAP-C experiments. (B) Off-target control results for 178 bp subsequences of *S. cerevisiae HAS1pr-TDA1pr*, shown as read coverage for 3C and genomic libraries. Each line represents a technical replicate. (C) Regions of *HAS1pr-TDA1pr* used for testing ectopic pairing. Pairing sequence shown in pink, and non-pairing control shown in gray. (D) Distribution of the number of mutations in the wild type and error-prone PCR genomic libraries. (E) Distribution of relative mutation frequency in 3C libraries. Positions most strongly required for pairing and highlighted in red in **Figure 3.1D-I** are also shown here in red; other positions are in grey. (F) Off-target control results for error-prone PCR mutagenesis of central *S. cerevisiae HAS1pr-TDA1pr*, shown as ratio of the total substitution frequency at each position in the 3C library compared to the genomic library. Error bars indicate the two technical replicates.

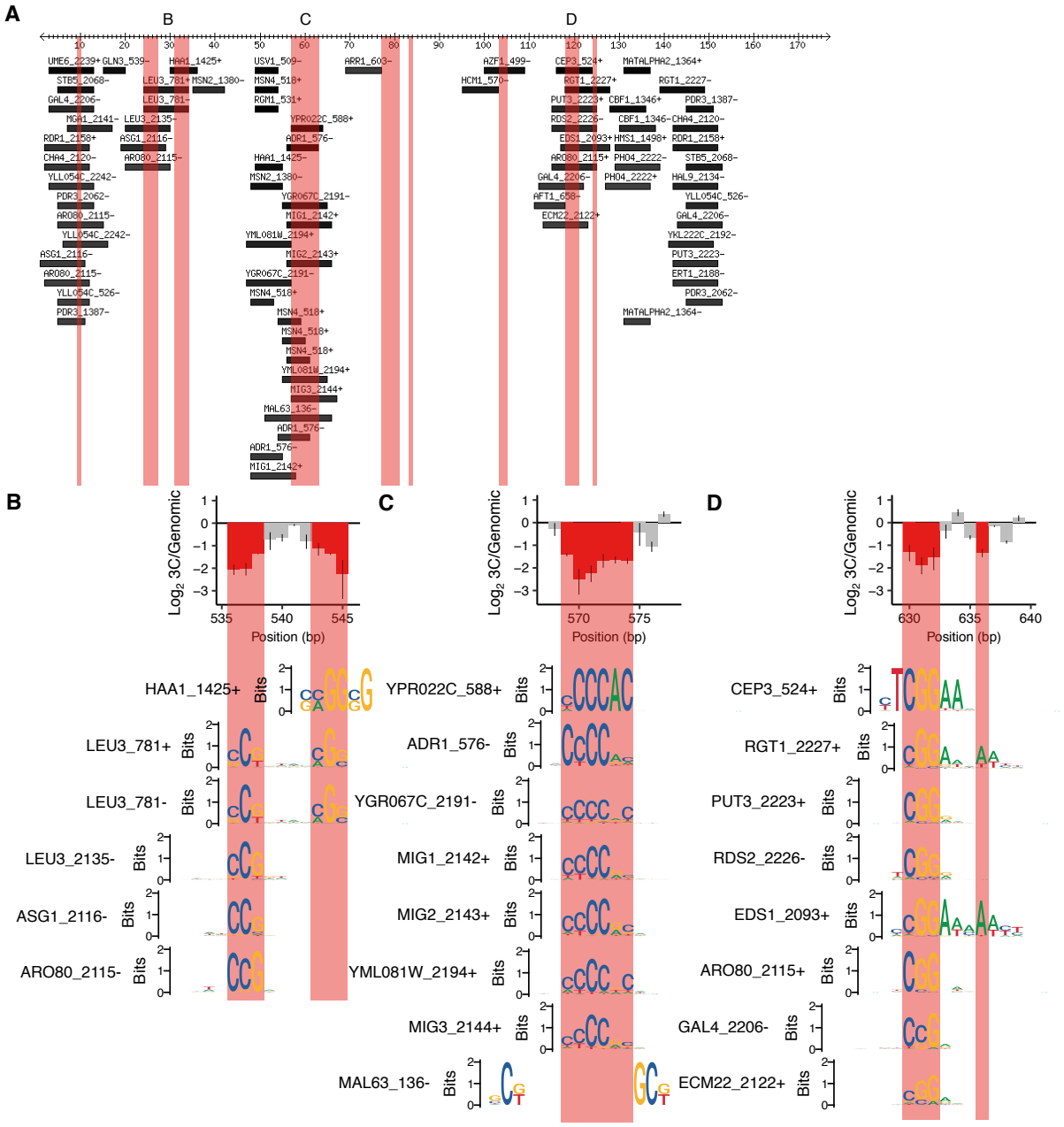

Figure 3.3. Motifs overlapping positions required for *HAS1pr-TDA1pr* pairing.

(A) Motifs overlapping region of *HAS1pr-TDA1pr* containing positions required for pairing. Coordinates are relative to 511 bp upstream of the *HAS1* coding sequence in *S. cerevisiae* (sacCer3). All "expert-curated no dubious" motifs from YeTFaSCo (de Boer and Hughes, 2012) scoring at least 75% of the maximum possible score are shown. B, C, and D indicate regions highlighted in other panels. Pink regions indicate positions required for *HAS1pr-TDA1pr* pairing (same as in **Figure 3.1D**). (B-D) Sequence logos of motifs spanning the cluster(s) of positions required for pairing are shown aligned relative to the error-prone PCR *cis* MAP-C data from **Figure 3.1E, F, and I**, in the same vertical order as in A.

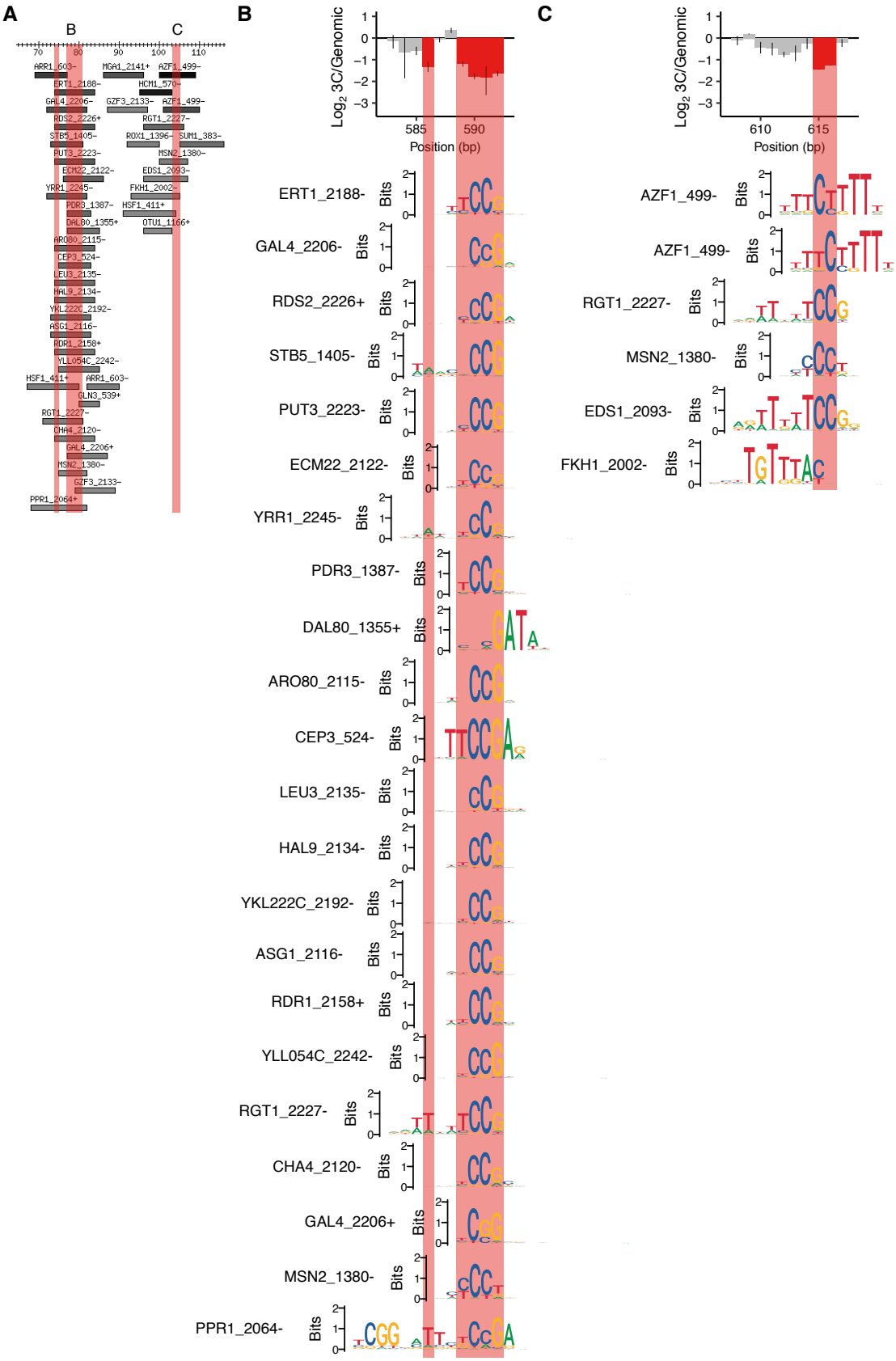
A
B
C
B
C
Log2 3C/Genomic
Position (bp)
585
590
610
615
Bits
ERT1_2188-
GAL4_2206-
RDS2_2226+
STB5_1405-
PUT3_2223-
ECM22_2122-
YRR1_2245-
PDR3_1387-
DAL80_1355+
ARO80_2115-
CEP3_524-
LEU3_2135-
HAL9_2134-
YKL222C_2192-
ASG1_2116-
RDR1_2158+
YLL054C_2242-
RGT1_2227-
CHA4_2120-
GAL4_2206+
MSN2_1380-
PPR1_2064-
AZF1_499-
AZF1_499-
RGT1_2227-
MSN2_1380-
EDS1_2093-
FKH1_2002-

Figure 3.4. Lower-scoring motifs matching positions required for *HAS1pr-TDA1pr* pairing.

(A) Motifs overlapping region of *HAS1pr-TDA1pr* containing positions required for pairing, zooming into the clusters of positions not shown in **Figure 3.3**. Coordinates are relative to 511 bp upstream of the *HAS1* coding sequence in *S. cerevisiae* (sacCer3). All "expert-curated no dubious" motifs from YeTFaSCo (de Boer and Hughes, 2012) scoring at least 45% of the maximum possible score are shown. B and C indicate regions highlighted in other panels. Pink regions indicate positions required for *HAS1pr-TDA1pr* pairing (same as in **Figure 3.1D**). (B-C) Sequence logos of motifs spanning the cluster of positions required for pairing are shown aligned relative to the error-prone PCR *cis* MAP-C data from **Figure 3.1G and H**, in the same vertical order as in panel A.

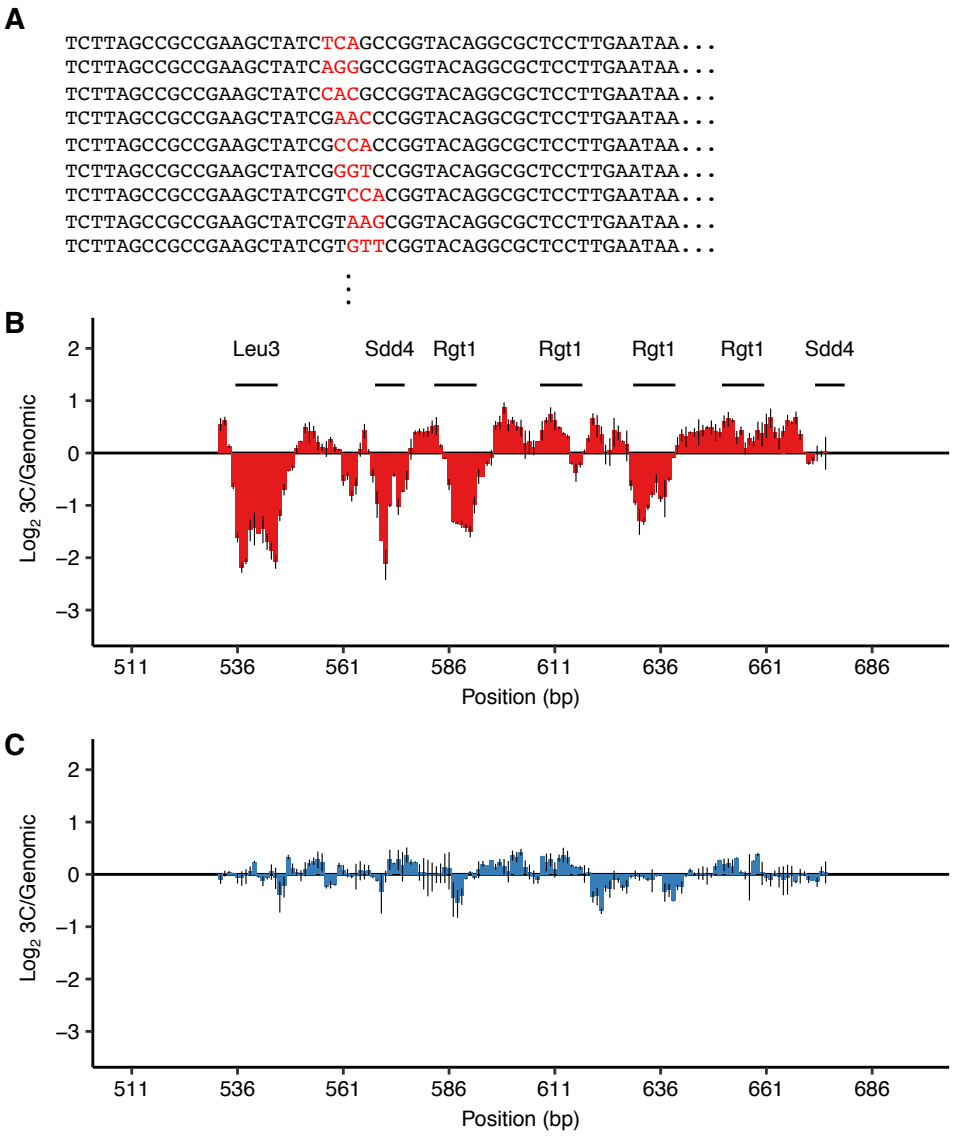

Figure 3.5. Validation of TF motifs required for *HAS1pr-TDA1pr* pairing with a 3 bp substitution mutant library.

(A) Library design. Examples of mutant sequences, with mutations in red. Each set of 3 bp was replaced with three random 3 bp sequences so that every nucleotide was switched to every other nucleotide. (B-C) Ratio of the total substitution frequency at each position in the 3C library compared to the genomic library, for contacts between *S. cerevisiae HAS1pr-TDA1pr* and the *S. uvarum* homolog (B) or *S. cerevisiae LCB1*, an off-target control (C). Horizontal line segments indicate the positions of TF motif matches.

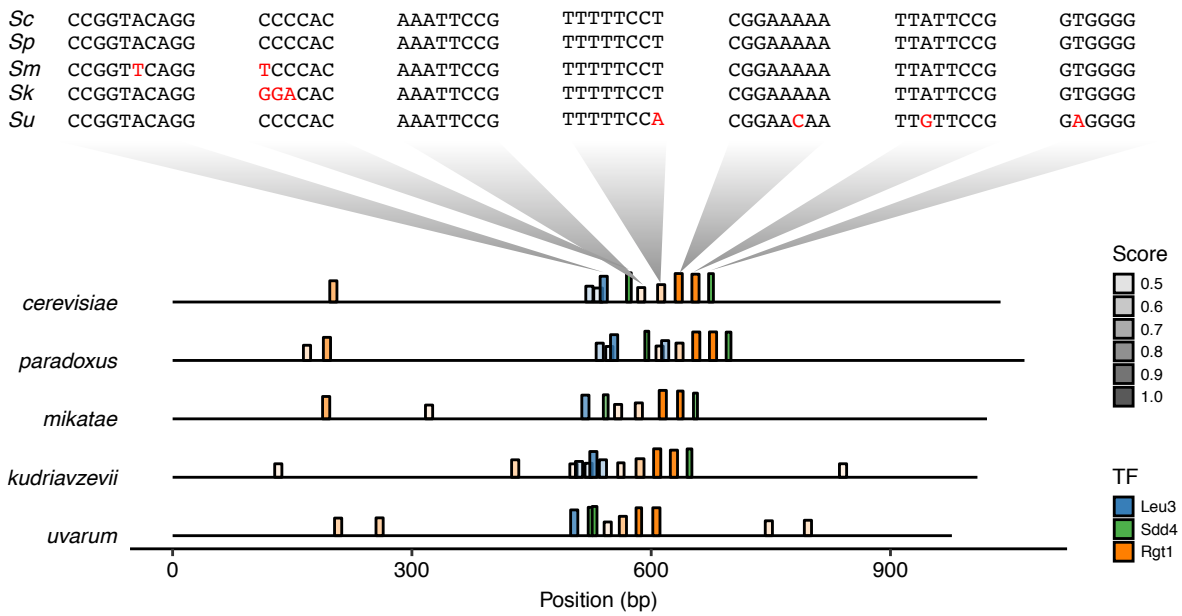

Figure 3.6. Conservation of TF motifs in *HAS1pr-TDA1pr*.

Position of Leu3, Sdd4, and Rgt1 motifs (at least 45% of maximum possible score, using high score position weight matrix from YeTFaSCo (de Boer and Hughes, 2012)) in the intergenic region between *HAS1* and *TDA1* in *Saccharomyces sensu stricto* yeasts *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum* (order of divergence from *S. cerevisiae*). Bar height and transparency both represent motif score (as fraction of maximum possible score). Above, the sequences of the regions aligning to the *S. cerevisiae* motifs are shown in the same vertical order, with mismatches in red.

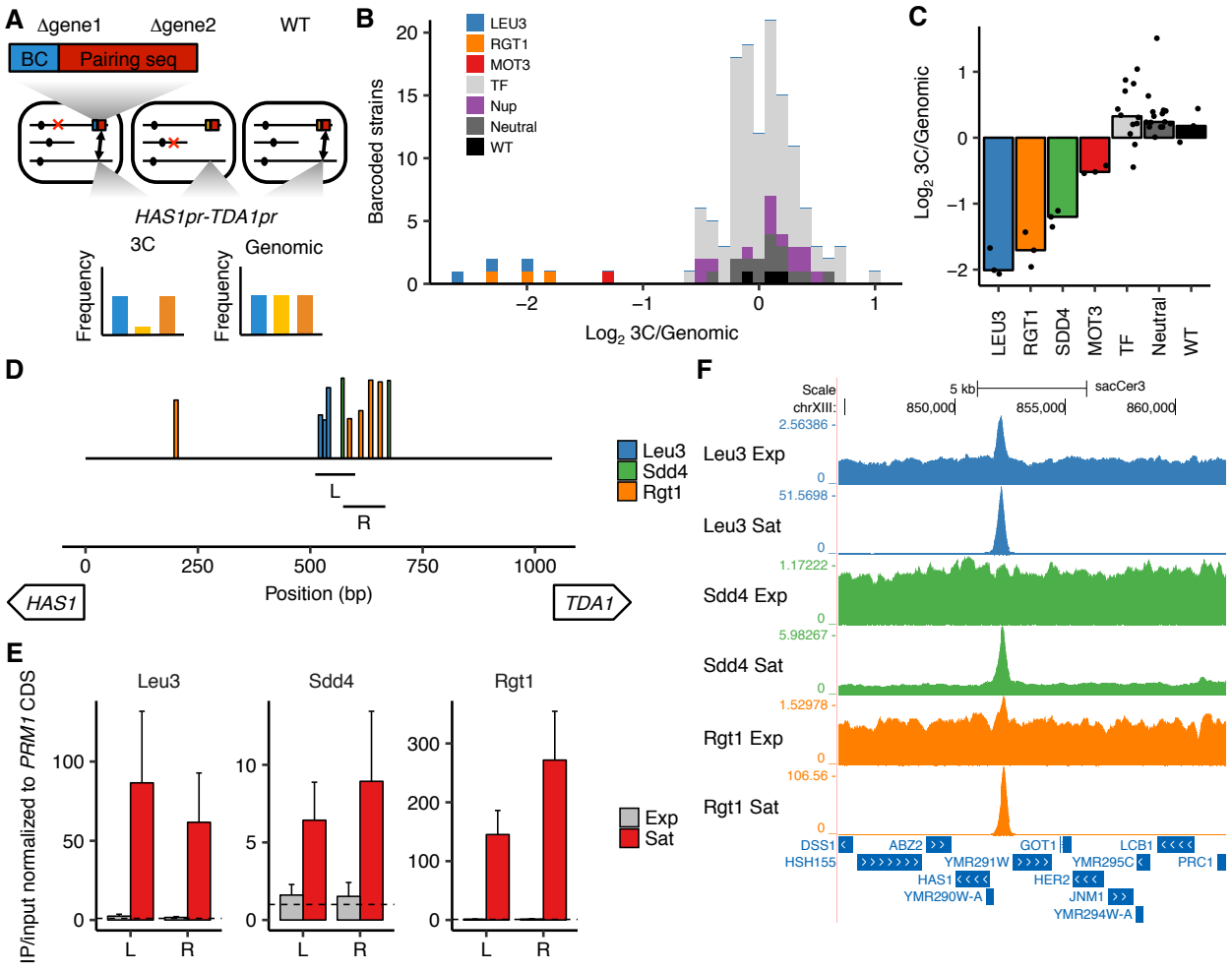

Figure 3.7. Transcription factors Leu3, Sdd4, and Rgt1 mediate *HAS1pr-TDA1pr* pairing.

(A) In *trans* MAP-C, the effect of TF knockouts or variants are assessed for their effect on a specific 3D contact by association with barcodes. Barcoded versions of the minimal pairing sequence were ectopically integrated and assessed for pairing with the native *HAS1pr-TDA1pr* in haploid *S. cerevisiae*. Red Xs indicate gene knockouts; red boxes indicate pairing sequence (**Figure 3.2C**); BC indicates barcode; double-headed arrows indicate the presence of chromosomal contacts. (B) Full TF knockout screen identifies Leu3 and Rgt1 as *trans* requirements for pairing. Histogram of relative abundance of each barcoded gene knockout strain in 3C library compared to the genomic library, excluding strains below a frequency of 0.3% of the pool. Barcode replicates are shown as separate squares in histogram. *LEU3*, *RGT1*, and *MOT3* are highlighted individually; TF indicates other transcription factors; Nup indicates nuclear pore complex components; Neutral indicates fitness-neutral negative controls (**Figure 3.9A**). (C) Validation TF knockout screen confirms that Leu3, Sdd4, and Rgt1 are required for pairing. Bar plot of median relative abundance in 3C library compared to the genomic library, with overlaid scatter plot of individual barcoded strains. TF includes *MIG1*, *VHR1*, *CBF1*, and

YGR067C. (D) Regions tested by ChIP-qPCR for TF binding. Bars show Leu3, Sdd4, and Rgt1 motif matches in *S. cerevisiae HAS1pr-TDA1pr*, and lines indicate regions (L and R) used for qPCRs in panel E. Bar heights indicate motif score as fraction of maximum possible score; all motifs with a score of at least 0.45 shown. (E) TFs bind *HAS1pr-TDA1pr* more strongly in saturated conditions. Chromatin immunoprecipitation qPCR results in exponentially growing (Exp) and saturated (Sat) cultures, normalized by input and to a negative control locus in the *PRM1* coding sequence. L and R indicate two primer sets as shown in panel D. Bars indicate mean ± s.e.m. of biological triplicates. Dashed lines indicate a value of 1 (background enrichment). (F). Chromatin immunoprecipitation sequencing data near the *HAS1pr-TDA1pr* locus (coordinates chrXIII:844,705-862,314 in sacCer3 reference), shown as fold enrichment in IPs over input.

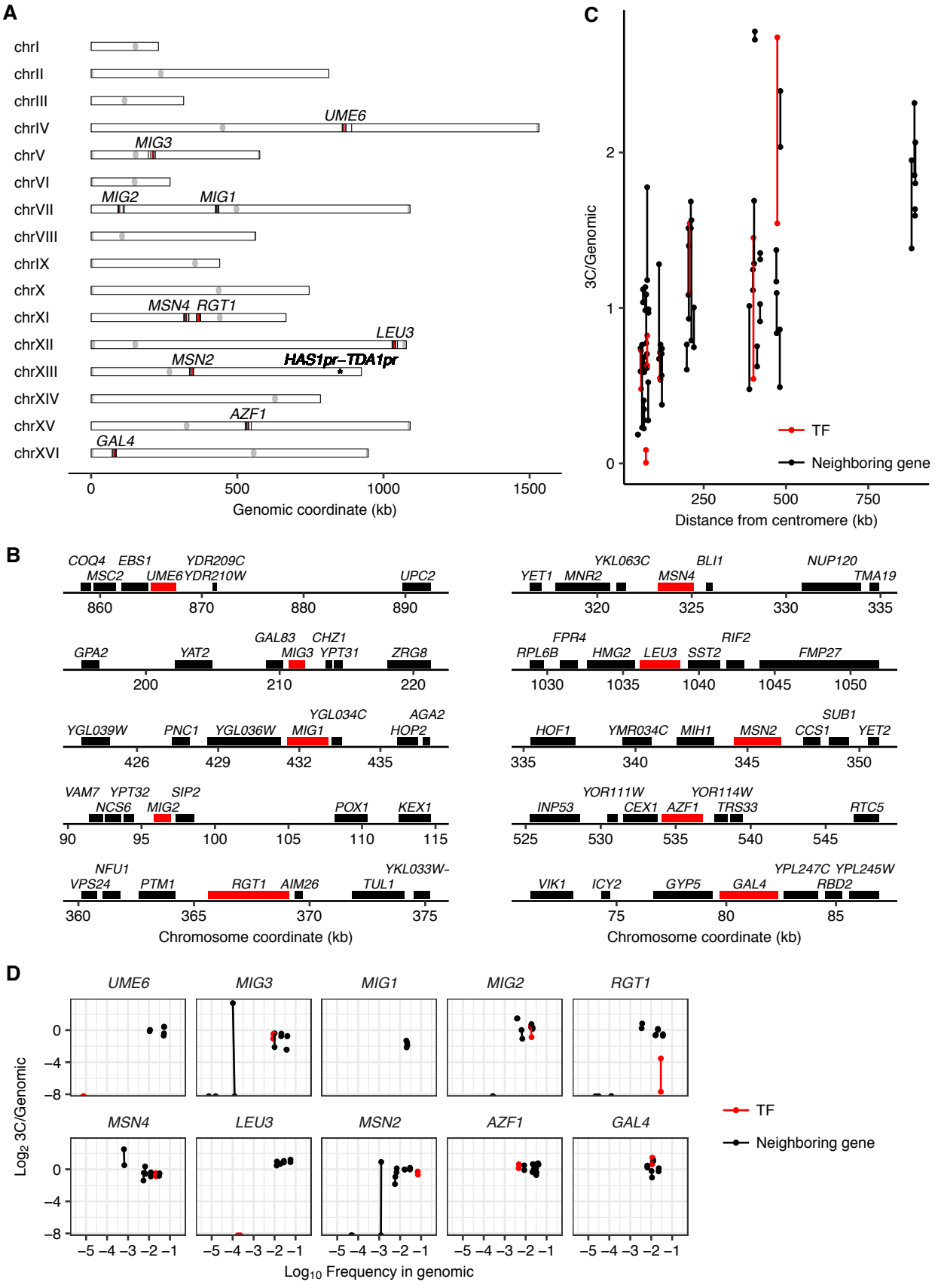
A
chrI
chrII
chrIII
chrIV
chrV
chrVI
chrVII
chrVIII
chrIX
chrX
chrXI
chrXII
chrXIII
chrXIV
chrXV
chrXVI
UME6
MIG3
MIG2
MIG1
MSN4
RGT1
LEU3
MSN2
HAS1pr–TDA1pr
AZF1
GAL4
0
500
1000
1500
Genomic coordinate (kb)
C
3C/Genomic
0
1
2
250
500
750
Distance from centromere (kb)
TF
Neighboring gene
B
COQ4
EBS1
YDR209C
MSC2
UME6
YDR210W
UPC2
860
870
880
890
GAL83
CHZ1
GPA2
YAT2
MIG3
YPT31
ZRG8
200
210
220
YGL034C
AGA2
YGL039W
PNC1
YGL036W
MIG1
HOP2
426
429
432
435
VAM7
YPT32
SIP2
NCS6
MIG2
POX1
KEX1
90
95
100
105
110
115
NFU1
YKL033W-
VPS24
PTM1
RGT1
AIM26
TUL1
360
365
370
375
Chromosome coordinate (kb)
YKL063C
BLI1
NUP120
YET1
MNR2
MSN4
TMA19
320
325
330
335
FPR4
RIF2
RPL6B
HMG2
LEU3
SST2
FMP27
1030
1035
1040
1045
1050
SUB1
HOF1
YMR034C
MIH1
MSN2
CCS1
YET2
335
340
345
350
YOR111W
YOR114W
INP53
CEX1
AZF1
TRS33
RTC5
525
530
535
540
545
YPL247C
YPL245W
VIK1
ICY2
GYP5
GAL4
RBD2
75
80
85
Chromosome coordinate (kb)
D
UME6
MIG3
MIG1
MIG2
RGT1
MSN4
LEU3
MSN2
AZF1
GAL4
Log2 3C/Genomic
0
−4
−8
−5 −4 −3 −2 −1
Log10 Frequency in genomic
TF
Neighboring gene

Figure 3.8. A pilot TF gene knockout screen for *HAS1pr-TDA1pr* pairing.

(A) Genomic locations of the tested TF gene knockouts (in red) and tested genomic neighbors (in black). (B) Zoom-in to tested TF gene knockouts (in red) and tested genomic neighbors (in black). (C) Effect of centromeric distance on ectopic *HAS1pr-TDA1pr* pairing in haploid *S. cerevisiae*. Each pair of connected dots represent two technical replicates. (D) Pilot TF knockout screen results, with each TF grouped with its genomic neighbors. Scatter plot of barcoded strain abundance in 3C library normalized to genomic library (y-axis) vs. abundance in genomic library (x-axis) for each of 10 tested TF gene knockouts (in red, separate panel per TF), along with genomic neighbors (in black). As in panel C, each pair of connected dots represent two technical replicates.

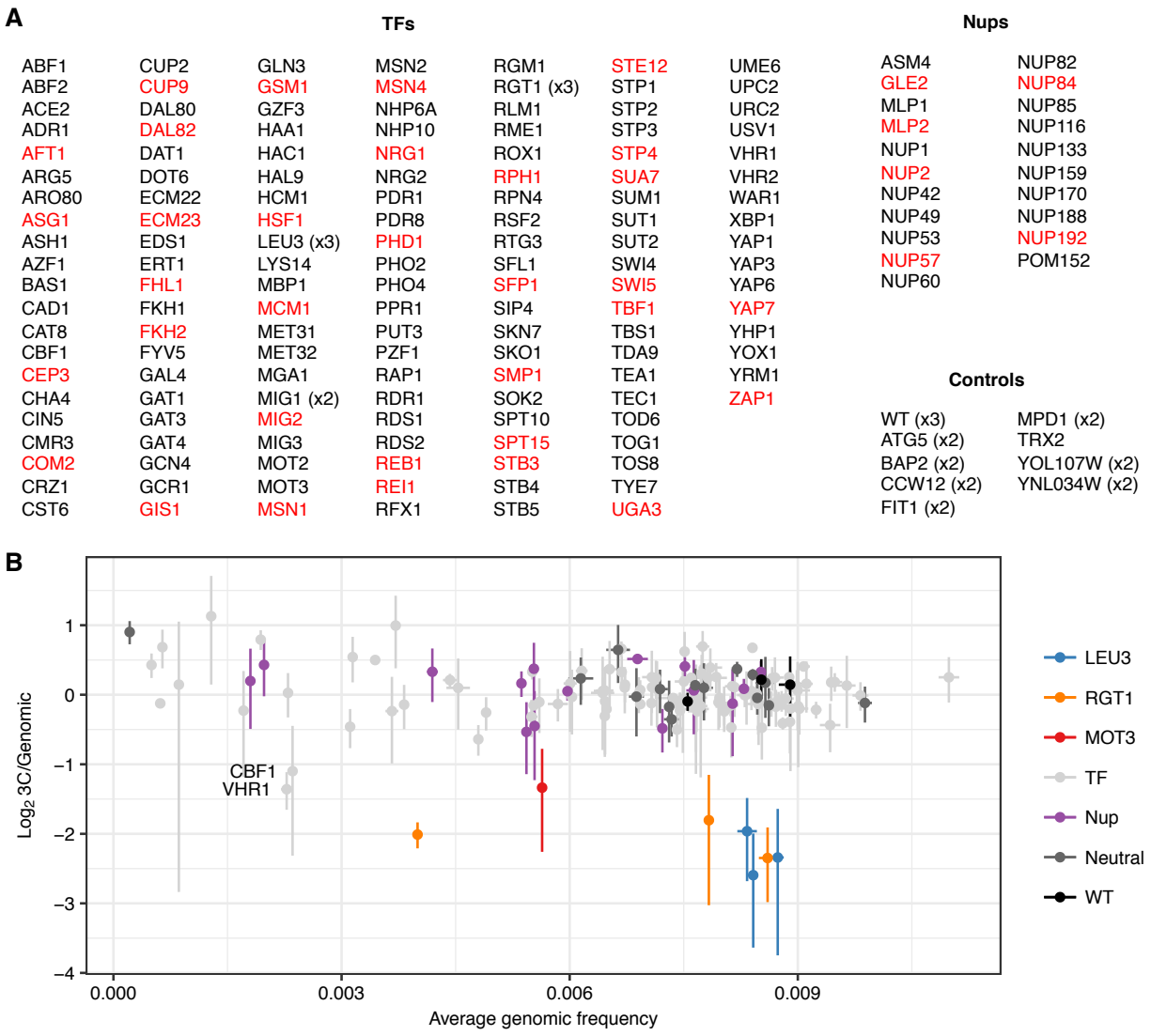

Figure 3.9. An expanded *trans* knockout screen for *HAS1pr-TDA1pr* pairing.

(A) List of tested gene knockouts. Strains that dropped out of the pool during library construction are shown in red. (B) Scatter plot of each barcoded strain's abundance in the genomic library vs. ratio of abundance in the 3C library compared to the genomic library. Center values indicate mean; error bars indicate s.d. of three technical replicates.

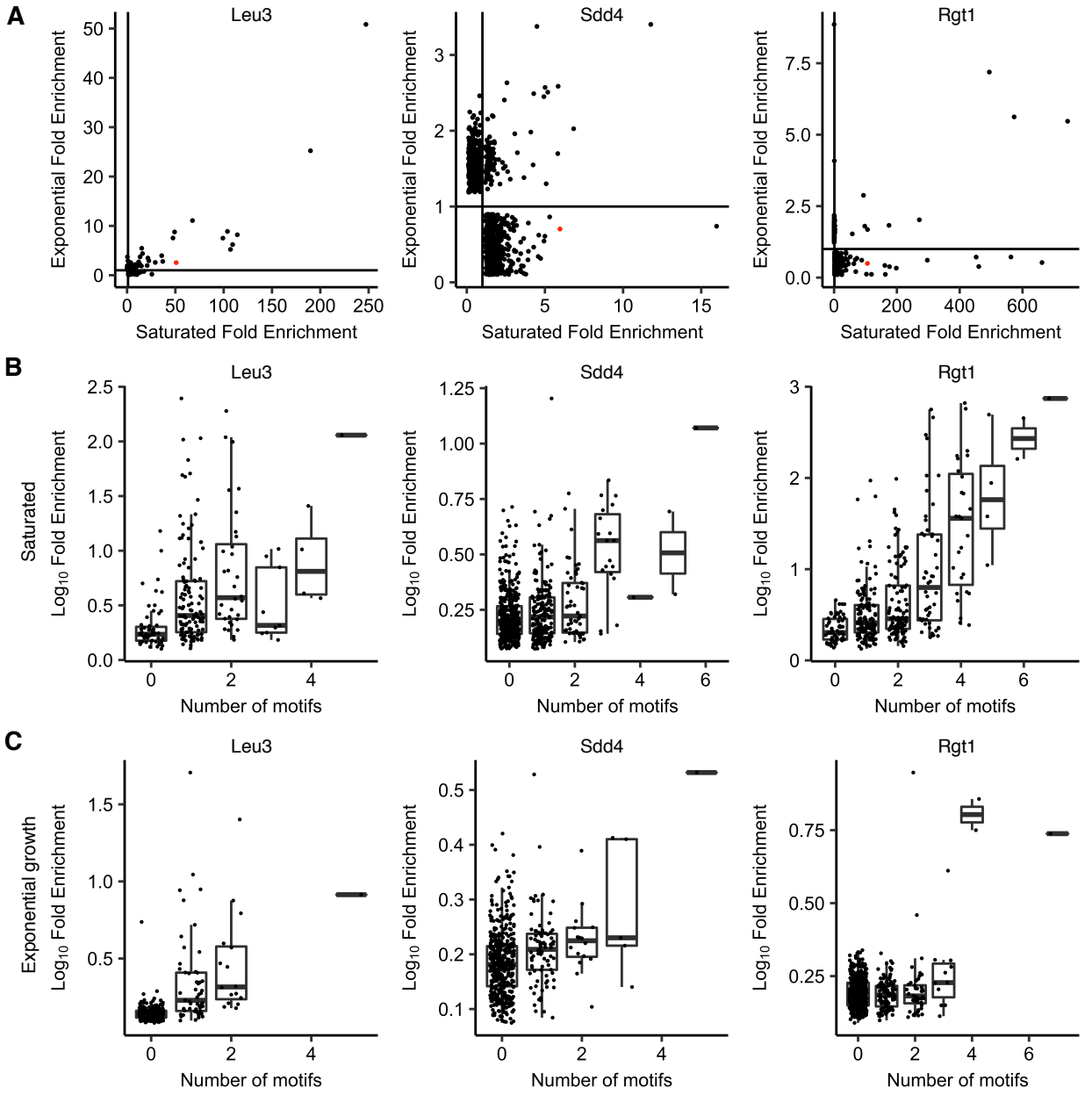

Figure 3.10. ChIP-seq of Leu3, Sdd4, and Rgt1 shows stronger motif-driven binding in saturated conditions.

(A) Scatter plot of ChIP-seq peak fold enrichment in saturated cultures vs. exponentially growing cultures, with *HAS1pr-TDA1pr* shown in red. Peaks with no overlapping significant peak in the other condition are shown with a random fold enrichment of 0.1-0.9, to prevent overplotting. Lines indicate fold enrichment of 1. (B) Boxplot of ChIP-seq peak $\log_{10}$ fold enrichment in saturated conditions, by number of motifs (p-value < 0.001 in FIMO) within 100 bp of the peak summit. Boxplot indicates median and interquartile range. (C) Same as panel B for exponential growth conditions.

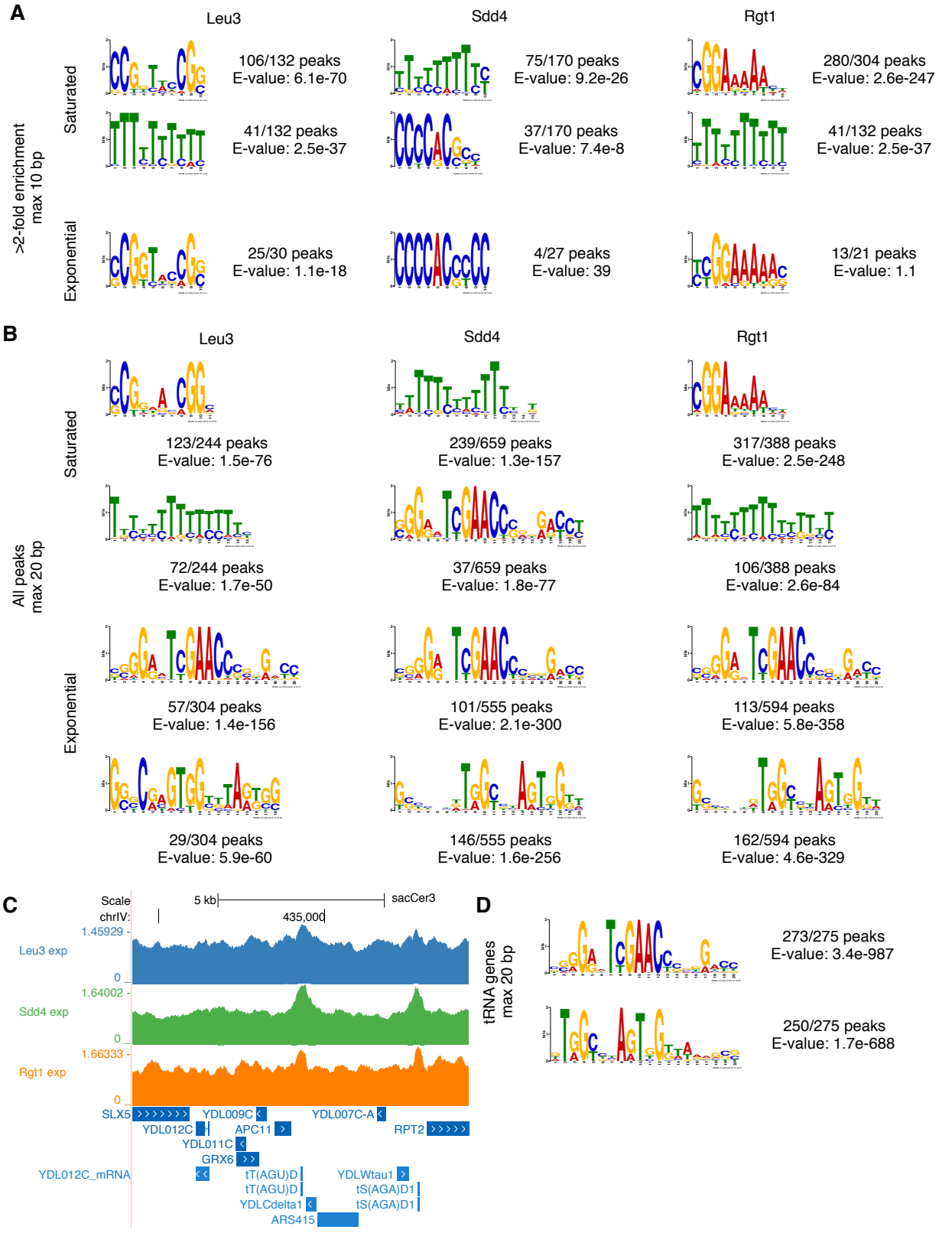
A
Leu3
Sdd4
Rgt1
>2-fold enrichment
max 10 bp
Saturated
Exponential
106/132 peaks
E-value: 6.1e-70
75/170 peaks
E-value: 9.2e-26
280/304 peaks
E-value: 2.6e-247
41/132 peaks
E-value: 2.5e-37
37/170 peaks
E-value: 7.4e-8
41/132 peaks
E-value: 2.5e-37
25/30 peaks
E-value: 1.1e-18
4/27 peaks
E-value: 39
13/21 peaks
E-value: 1.1
B
Leu3
Sdd4
Rgt1
All peaks
max 20 bp
Saturated
Exponential
123/244 peaks
E-value: 1.5e-76
239/659 peaks
E-value: 1.3e-157
317/388 peaks
E-value: 2.5e-248
72/244 peaks
E-value: 1.7e-50
37/659 peaks
E-value: 1.8e-77
106/388 peaks
E-value: 2.6e-84
57/304 peaks
E-value: 1.4e-156
101/555 peaks
E-value: 2.1e-300
113/594 peaks
E-value: 5.8e-358
29/304 peaks
E-value: 5.9e-60
146/555 peaks
E-value: 1.6e-256
162/594 peaks
E-value: 4.6e-329
C
Scale
chrIV:
5 kb
435,000
sacCer3
1.45929
Leu3 exp
0
1.64002
Sdd4 exp
0
1.66333
Rgt1 exp
0
SLX5
YDL009C
YDL007C-A
YDL012C
APC11
RPT2
YDL011C
GRX6
YDL012C_mRNA
tT(AGU)D
YDLWtau1
tT(AGU)D
tS(AGA)D1
YDLCdelta1
tS(AGA)D1
ARS415
D
tRNA genes
max 20 bp
273/275 peaks
E-value: 3.4e-987
250/275 peaks
E-value: 1.7e-688

Figure 3.11. *De novo* motif discovery reveals ChIP-seq enrichment of known motifs, poly-T tracts enriched in promoters, and tRNA genes.

(A) Expected TF motifs and poly(dA-dT) tracts, which are common in yeast promoters and activate transcription via their intrinsic nucleosome depletion (Iyer and Struhl, 1995; Struhl, 1985), are enriched in top ChIP-seq peaks. Motifs were discovered using MEME (E-value < 0.1, or top motif if none below threshold) (Bailey et al., 2006) on all ChIP-seq peaks with at least 2-fold enrichment, allowing motifs of up to 10 bp. Motifs are shown in the orientations that match known motifs. (B) Low-enrichment ChIP-seq peaks contain a long unexpected motif. As in panel A, motifs discovered using MEME (top two motifs) but on all ChIP-seq peaks, allowing motifs of up to 20 bp. (C) Low-enrichment ChIP-seq peaks at tRNA genes. Example of exponential growth ChIP-seq data at two tRNA gene loci. Coordinates (sacCer3 reference): chrIV:434220-434320. (D) tRNA genes contain motifs matching those from low-enrichment ChIP-seq peaks. Motifs discovered using MEME (top two motifs) on tRNA genes.

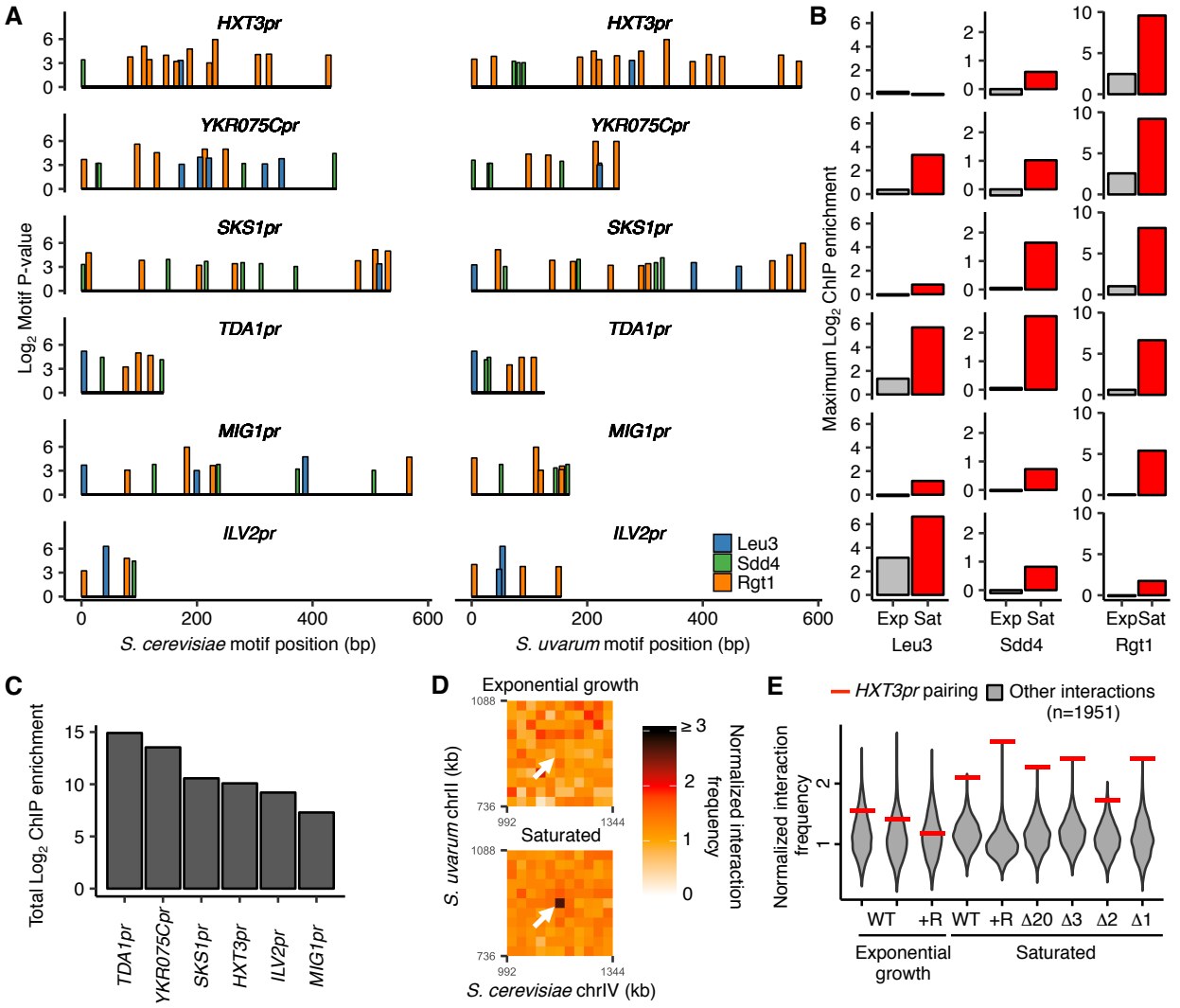

Figure 3.12. Combinatorial TF binding specifies inducible homolog pairing.

(A) Top clusters of Leu3, Sdd4, and Rgt1 motifs in homologous loci of both *S. cerevisiae* and *S. uvarum* genomes containing all three motifs in *S. cerevisiae*, in order of lowest *P*-value from top to bottom. (B) Not all motif clusters are bound by all three TFs. ChIP-seq data for exponentially growing (Exp) and saturated (Sat) *S. cerevisiae* corresponding to motif clusters in panel A. (C) Combined TF binding is strongest at *TDA1pr*. Shown are the sum of the ChIP $\log_2$ fold enrichments over inputs in saturated conditions for motif clusters in panel A. (D) *HXT3pr* also exhibits inducible homolog pairing. Hi-C contact maps of interactions between regions centered on *S. cerevisiae HXT3* and *S. uvarum HXT3* at 32 kb resolution in exponentially growing and saturated cultures of a *S. cerevisiae* x *S. uvarum* hybrid strain (YMD3920). White arrows indicate interactions between homologous *HXT3* promoters. (E) Strength of *HXT3* promoter pairing across conditions and strain backgrounds, at 32 kb resolution (red lines) compared to similar interactions (grey violin plots; i.e. interactions between an *S. cerevisiae* locus and an *S. uvarum* locus, where both loci are ≥15 genomic bins from a centromere and ≥2 bins from a

telomere, ≥2 bins from *HAS1pr-TDA1pr*, and not both on chrXII). WT represents strain ILY456 (YMD3259). +R indicates a restriction site added upstream of *HAS1* (YMD3920), Δ20 indicates a 20 kb deletion centered at *S. cerevisiae HAS1* (YMD3266), Δ3 indicates a 3 kb deletion centered at *S. cerevisiae HAS1* (YMD3267), Δ2 indicates a 2 kb deletion of the *S. cerevisiae HAS1* coding sequence (YMD3268), and Δ1 indicates a 1 kb deletion of the *S. cerevisiae HAS1pr-TDA1pr* intergenic region (YMD3269).

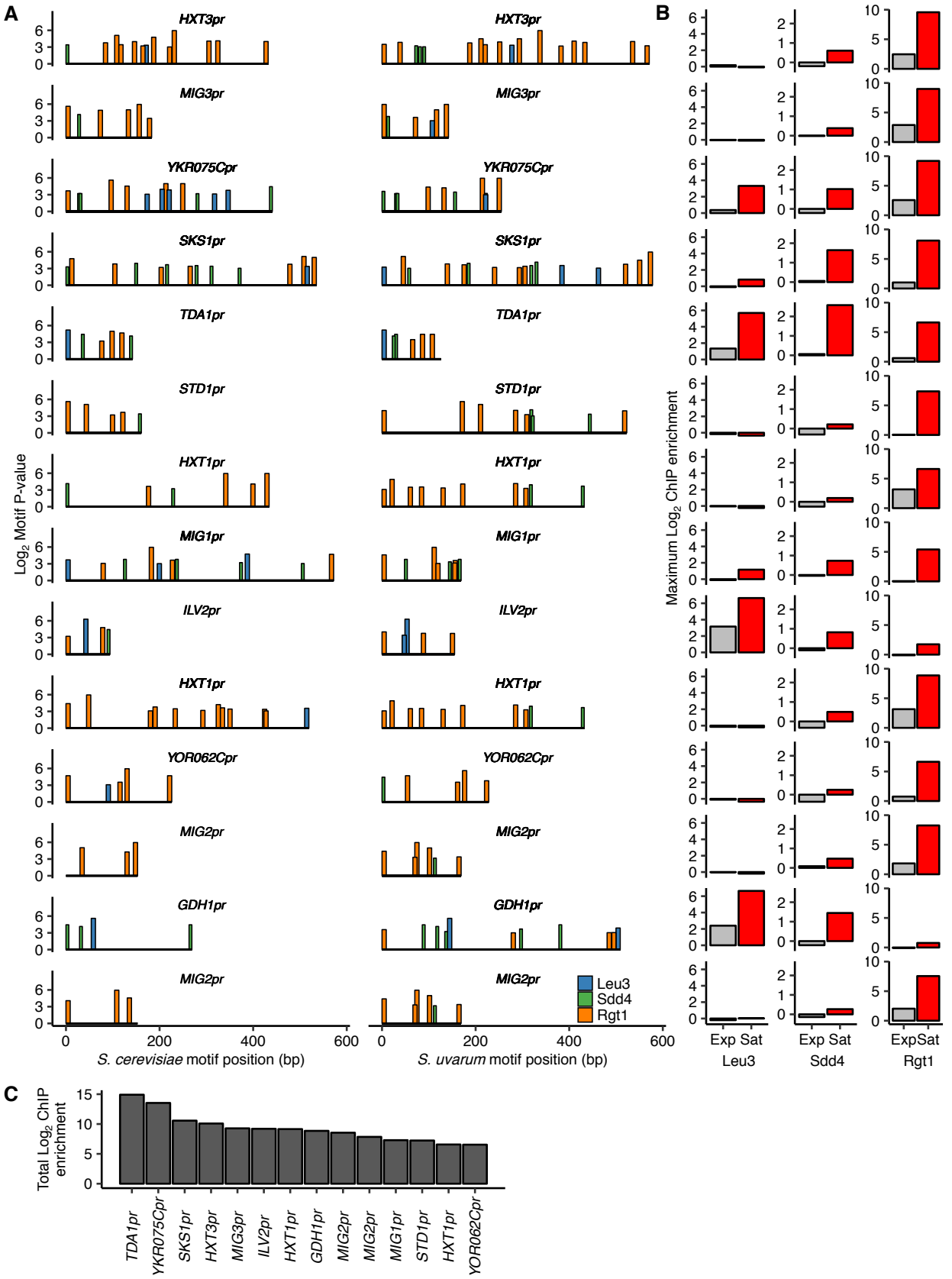
A
B
C
HXT3pr
MIG3pr
YKR075Cpr
SKS1pr
TDA1pr
STD1pr
HXT1pr
MIG1pr
ILV2pr
HXT1pr
YOR062Cpr
MIG2pr
GDH1pr
MIG2pr
Log2 Motif P-value
S. cerevisiae motif position (bp)
S. uvarum motif position (bp)
Leu3
Sdd4
Rgt1
Maximum Log2 ChIP enrichment
Exp Sat
Leu3
Sdd4
Rgt1
Total Log2 ChIP enrichment
TDA1pr
YKR075Cpr
SKS1pr
HXT3pr
MIG3pr
ILV2pr
HXT1pr
GDH1pr
MIG2pr
MIG2pr
MIG1pr
STD1pr
HXT1pr
YOR062Cpr

Figure 3.13. *HAS1pr-TDA1pr* exhibits uniquely strong combinatorial binding of Leu3, Sdd4, and Rgt1.

(A) Top clusters of Leu3, Sdd4, and Rgt1 motifs in homologous loci containing Leu3 and Rgt1 motifs in both *S. cerevisiae* and *S. uvarum* genomes as in **Figure 3.12,** but without requiring all three motifs in *S. cerevisiae*, in order of lowest *P*-value from top to bottom. (B) ChIP-seq data for exponentially growing (Exp) and saturated (Sat) *S. cerevisiae* corresponding to motif clusters in panel A. (C) The sum of the $\log_2$ fold enrichments in saturated conditions for motif clusters in panel A.

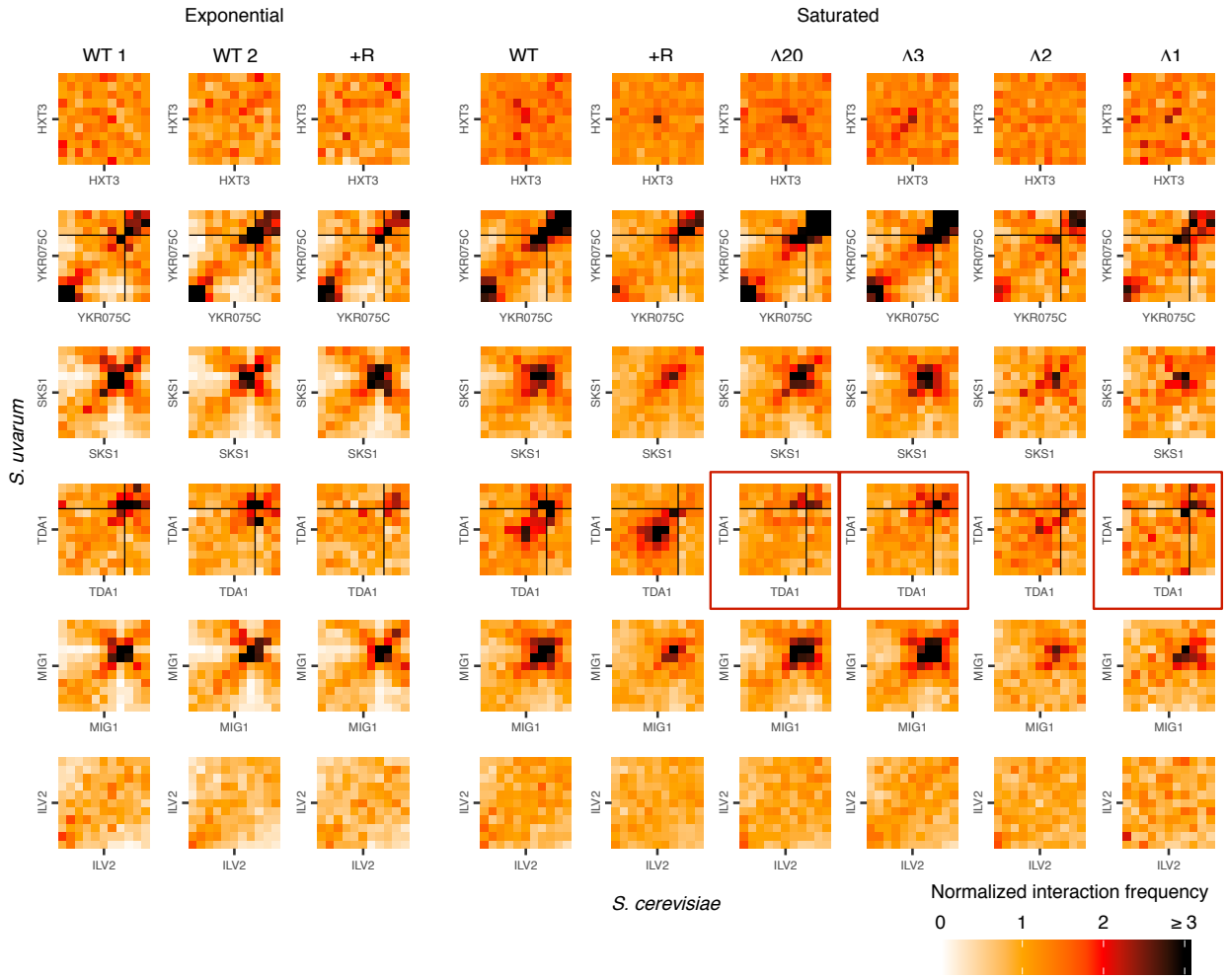

Figure 3.14. Hi-C evidence of inducible homolog pairing at *TDA1* and *HXT3*.

Hi-C contact maps of interactions between homologous regions centered on motif clusters in **Figure 3.12A**, at 32 kb resolution in exponentially growing and saturated cultures. WT represents strain ILY456 (YMD3259). +R indicates a restriction site added upstream of *HAS1* (YMD3920), Δ20 indicates a 20 kb deletion centered at *S. cerevisiae HAS1* (YMD3266), Δ3 indicates a 3 kb deletion centered at *S. cerevisiae HAS1* (YMD3267), Δ2 indicates a 2 kb deletion of the *S. cerevisiae HAS1* coding sequence (YMD3268), and Δ1 indicates a 1 kb deletion of the *S. cerevisiae HAS1pr-TDA1pr* intergenic region (YMD3269). Red boxes indicate the deletion strains where pairing between *TDA1* alleles should not occur. Black lines indicate chromosome starts/ends.

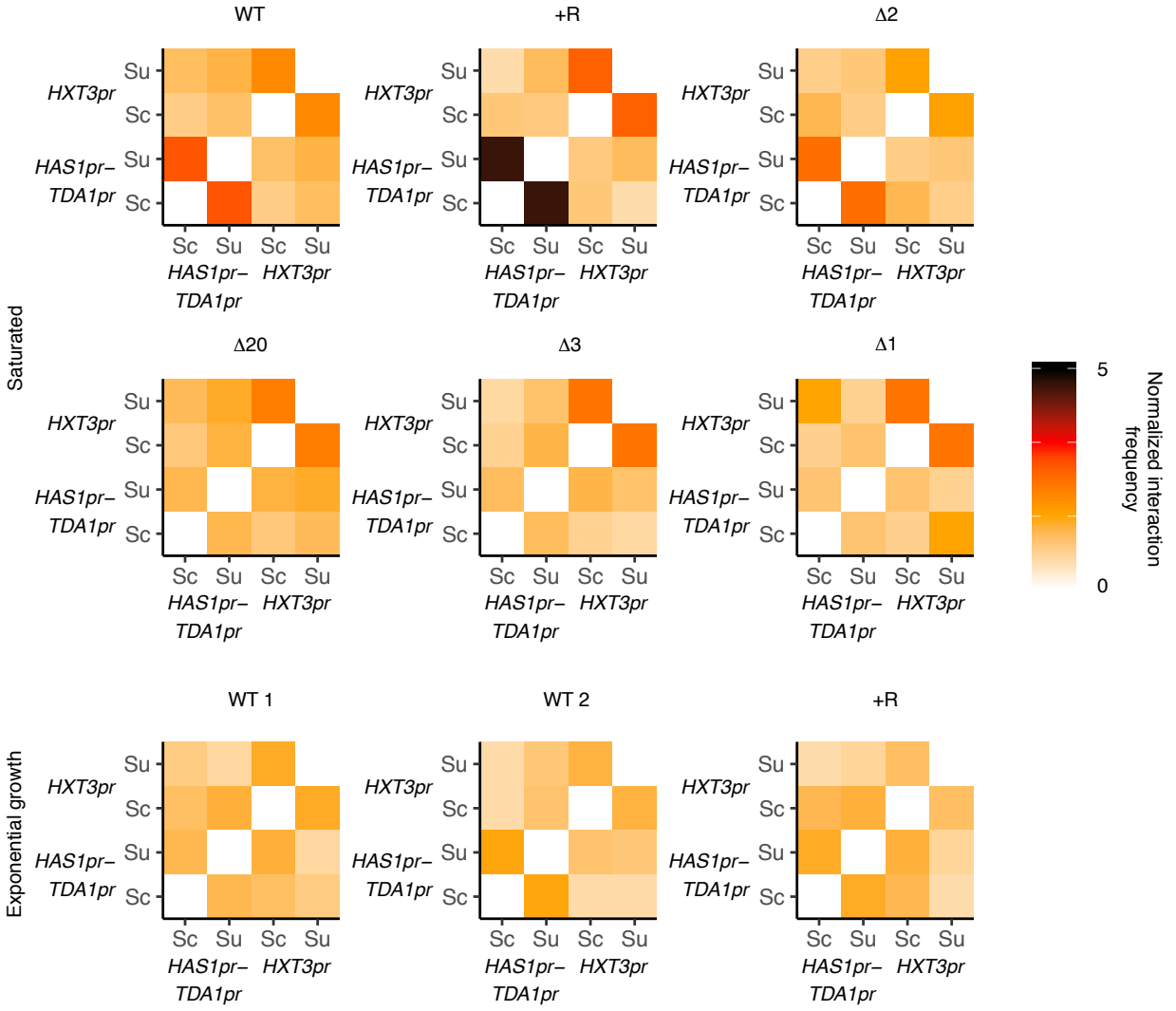

Figure 3.15. *HAS1pr-TDA1pr* and *HXT3pr* pairing are independent.

Hi-C interaction frequencies among the *S. cerevisiae* and *S. uvarum* copies of *HAS1pr-TDA1pr* and *HXT3pr*, at 32 kb resolution, across strains in pairing (saturated) and non-pairing (exponential growth) conditions. Sc indicates *S. cerevisiae*, and Su indicates *S. uvarum*. WT represents strain ILY456 (YMD3259). +R indicates a restriction site added upstream of *HAS1* (YMD3920), Δ20 indicates a 20 kb deletion centered at *S. cerevisiae HAS1* (YMD3266), Δ3 indicates a 3 kb deletion centered at *S. cerevisiae HAS1* (YMD3267), Δ2 indicates a 2 kb deletion of the *S. cerevisiae HAS1* coding sequence (YMD3268), and Δ1 indicates a 1 kb deletion of the *S. cerevisiae HAS1pr-TDA1pr* intergenic region (YMD3269).

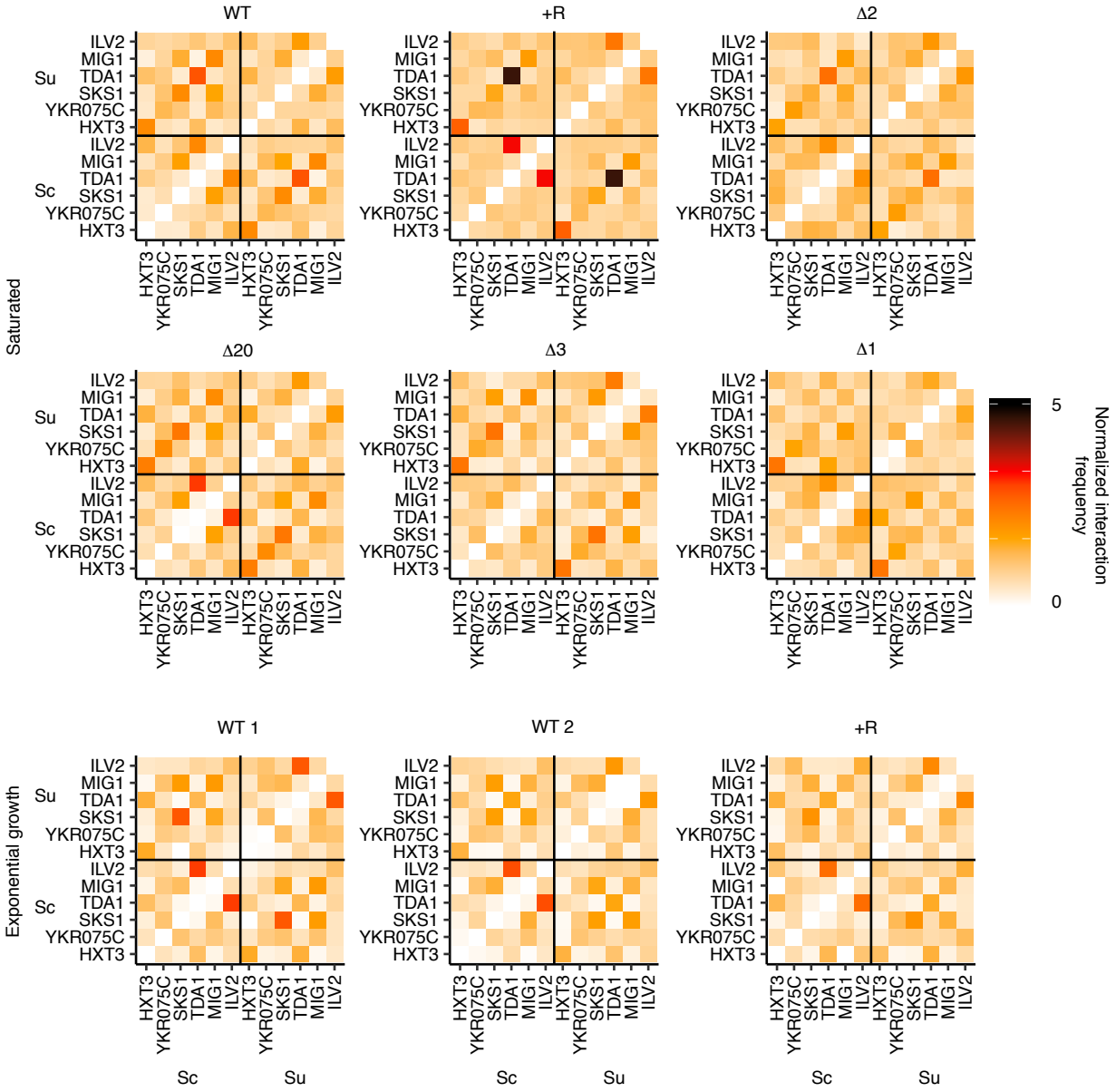

Figure 3.16. Lack of nonhomologous pairing between clusters of Leu3, Sdd4, and Rgt1 motifs.

Hi-C interaction frequencies among the *S. cerevisiae* and *S. uvarum* copies of motif clusters in **Figure 3.12,** at 32 kb resolution, across strains in pairing (saturated) and non-pairing (exponential growth) conditions. Sc indicates *S. cerevisiae*, and Su indicates *S. uvarum*. WT represents strain ILY456 (YMD3259). +R indicates a restriction site added upstream of *HAS1* (YMD3920), Δ20 indicates a 20 kb deletion centered at *S. cerevisiae HAS1* (YMD3266), Δ3 indicates a 3 kb deletion centered at *S. cerevisiae HAS1* (YMD3267), Δ2 indicates a 2 kb deletion of the *S. cerevisiae HAS1* coding sequence (YMD3268), and Δ1 indicates a 1 kb

deletion of the *S. cerevisiae HAS1pr-TDA1pr* intergenic region (YMD3269). Note that *ILV2* and *TDA1* are both on chrXIII, and are the only pair of motif clusters on the same chromosome.

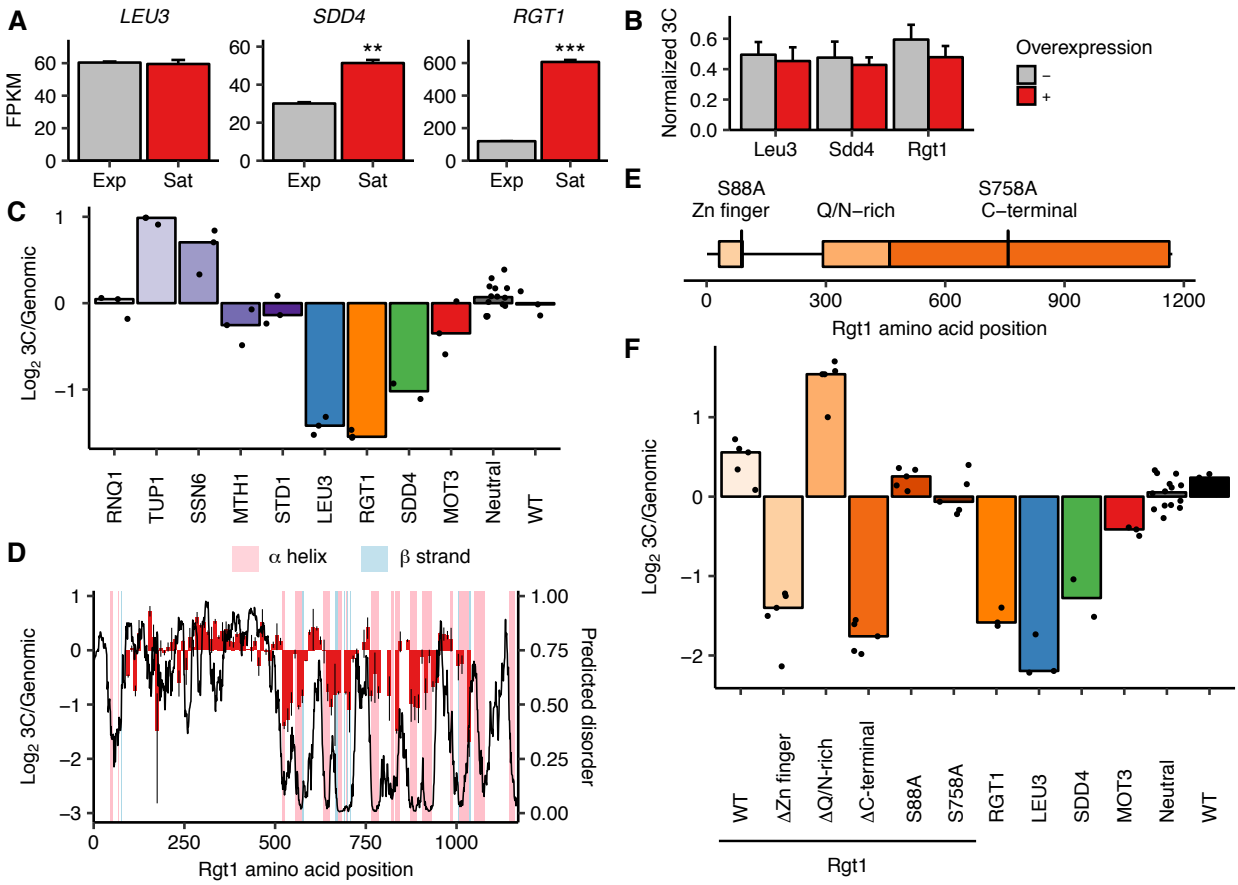

Figure 3.17. Rgt1 expression, interaction partners, and competing domains regulate *HAS1pr-TDA1pr* pairing.

(A) *RGT1* and *SDD4* are upregulated in saturated cultures. RNA-seq expression levels for *LEU3*, *SDD4*, *RGT1* in exponentially growing (Exp) and saturated (Sat) cultures. FPKM = fragments per kilobase per million read pairs. Bars indicate mean ± s.e.m. of biological triplicates. $^{**}P < 0.01$, $^{***}P < 0.001$ (Student's t-test). (B) Individual TF overexpression is insufficient for pairing. Shown are *HAS1pr-TDA1pr* homolog pairing frequencies with and without estradiol-induced overexpression of *LEU3*, *SDD4*, *RGT1* in *S. cerevisiae* x *S. uvarum* hybrids during exponential growth, measured by 3C and normalized to contacts between *HAS1pr-TDA1pr* and *LCB1* on *S. cerevisiae* chrXIII. Bars indicate mean ± s.d. of technical triplicates. (C) Effects of Rgt1 interaction partner deletions on ectopic *HAS1pr-TDA1pr* pairing in saturated conditions. Bar plot of median relative abundance in 3C library compared to the genomic library, with overlaid scatter plot of individual barcoded deletion strains. Retested controls are shown in same colors as in **Figure 3.7**; potential interaction partners are shown in shades of purple. (D) Regions of Rgt1 necessary for pairing in saturated culture conditions, with overlaid predicted disorder (black line, right y-axis) and secondary structure (pink and blue highlights). Each bar represents the ratio of the frequency of each 10 amino acid deletion in the 3C library compared to the genomic library. Error bars indicate the two biological replicates. (E) Domain deletions and phosphorylation site mutations tested in panel F. (F) Effects of Rgt1 domain deletions and

phosphorylation site mutations on ectopic *HAS1pr-TDA1pr* pairing in saturated cultures, plotted as in panel C. Rgt1 indicates the strains with an ectopic wild-type (WT) or mutant copy of Rgt1 in addition to a deletion of the endogenous *RGT1* locus.

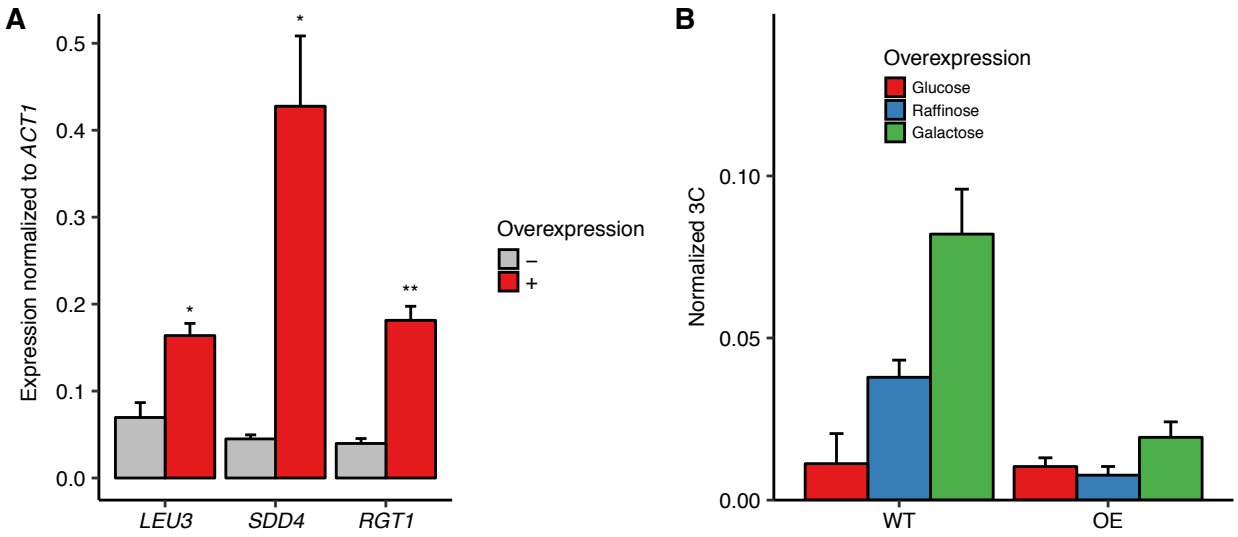

Figure 3.18. Overexpression of *RGT1* is not sufficient for *HAS1pr-TDA1pr* pairing.

(A) Estradiol-induced overexpression of TF genes. Bars show RNA levels of *S. cerevisiae LEU3*, *SDD4*, and *RGT1* with (red) and without (grey) estradiol-induced overexpression in *S. cerevisiae* x *S. uvarum* hybrids, measured by reverse transcription qPCR (RT-qPCR) and normalized to *S. cerevisiae ACT1* expression. Bars indicate mean ± s.e.m. of technical triplicates. **P* < 0.05, ***P* < 0.01 (Student's t-test). (B) *RGT1* overexpression does not result in increased pairing. Frequency of contacts between *HAS1pr-TDA1pr* alleles as measured by 3C, normalized to the frequency of contacts between *S. cerevisiae HAS1pr-TDA1pr* and *S. cerevisiae LCB1* in *S. cerevisiae* x *S. uvarum* hybrids (strain YMD3920) with (OE) or without (WT) a plasmid carrying a copy of epitope-tagged *RGT1-6xHis-HA-3C-pA* under a *GAL1* promoter, grown in glucose (red), raffinose (blue), or galactose (green). Bars indicate mean ± s.d. of technical triplicates.

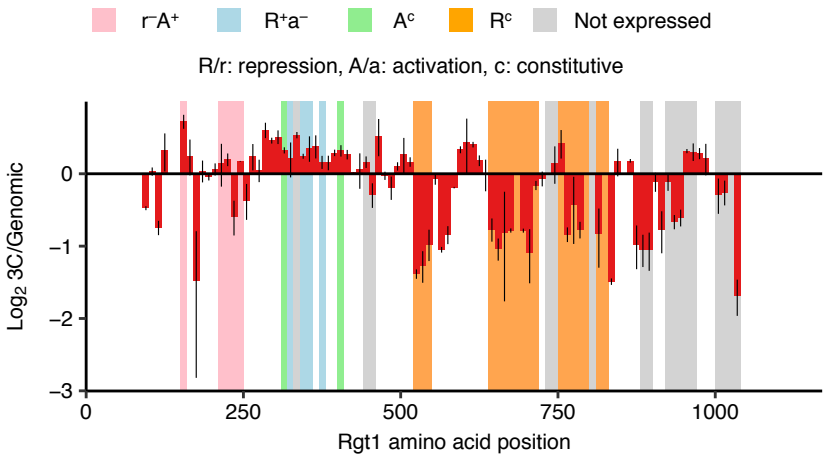

Figure 3.19. Regions of Rgt1 required for pairing correspond to regulatory domains.

Regions of Rgt1 necessary for *HAS1pr-TDA1pr* pairing in saturated conditions, as in **Figure 3.17D**, shown as ratio of the frequency of each 10 amino acid deletion in the 3C library compared to the genomic library. Error bars indicate the two biological replicates. Highlighted regions indicate the previously mapped functional domains of Rgt1 (Polish et al., 2005).

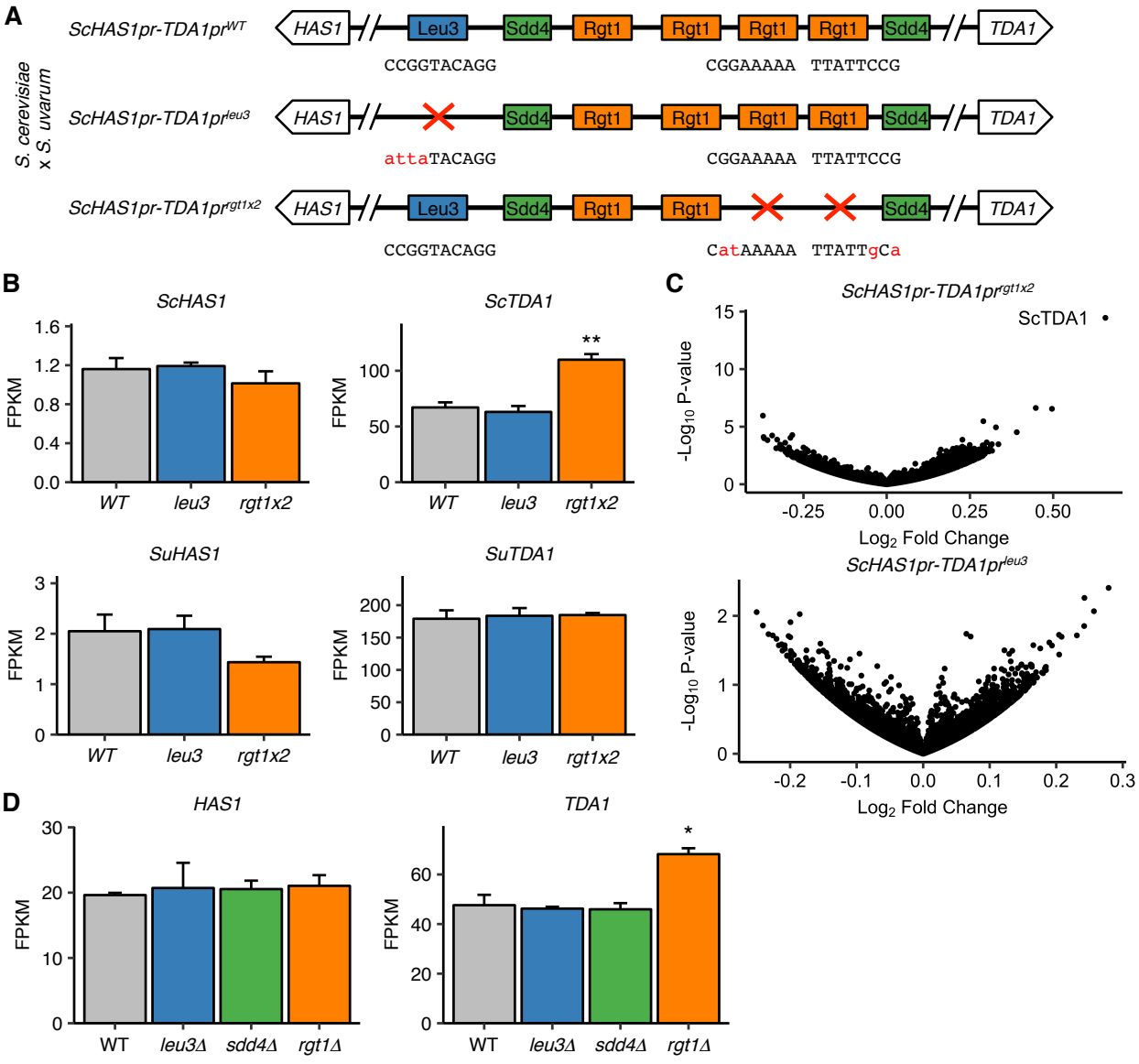

Figure 3.20. Pairing TFs play a minimal role in transcriptional regulation of the *HAS1-TDA1* locus.

(A) Mutations in *S. cerevisiae HAS1pr-TDA1pr* tested for transcriptional effects in *S. cerevisiae* x *S. uvarum* hybrids. Colored boxes indicate TF motifs. Wild type and mutated sequences are shown below boxes, with mutated bases in lowercase, red letters. (B) Rgt1 regulates *TDA1* expression in *cis*. RNA-seq expression levels for *S. cerevisiae* (*Sc*) and *S. uvarum* (*Su*) copies of *HAS1* and *TDA1* in saturated cultures of strains shown in panel A. FPKM = fragments per kilobase per million read pairs. Bars indicate mean ± s.e.m. of biological triplicates. $^{**}P < 0.01$ (Student's t-test). (C) Mutations in pairing region cause few transcriptional changes. Volcano plot of fold change vs. p-value in mutant strains compared to wild type, shown in (A). (D) RNA-seq expression levels for *HAS1* and *TDA1* in saturated cultures of wild type haploid *S. cerevisiae*

and gene deletion strains. Bars indicate mean ± s.e.m. of biological triplicates. *$P < 0.05$ (Student's t-test).

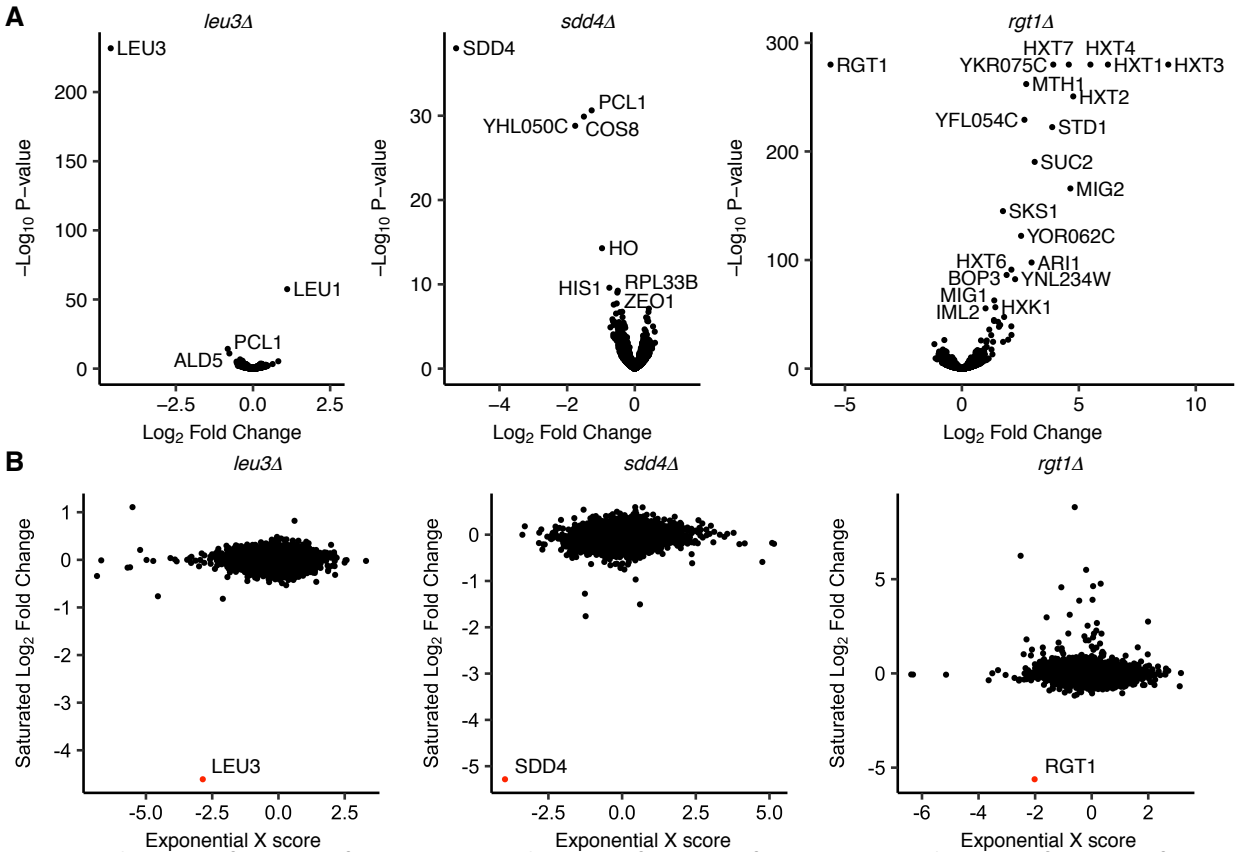

Figure 3.21. Transcriptional effects of TF deletion in saturated culture and exponential growth.

(A) Volcano plots of differentially expressed genes in TF deletion strains compared to a wild-type control in saturated culture conditions, by RNA-seq. (B) Comparison of $\log_2$ fold change from RNA-seq in saturated culture conditions to gene expression microarray "X scores" (normalized, confidence-weighted log ratios) in exponential growth (Hu et al., 2007), with deleted gene highlighted in red.

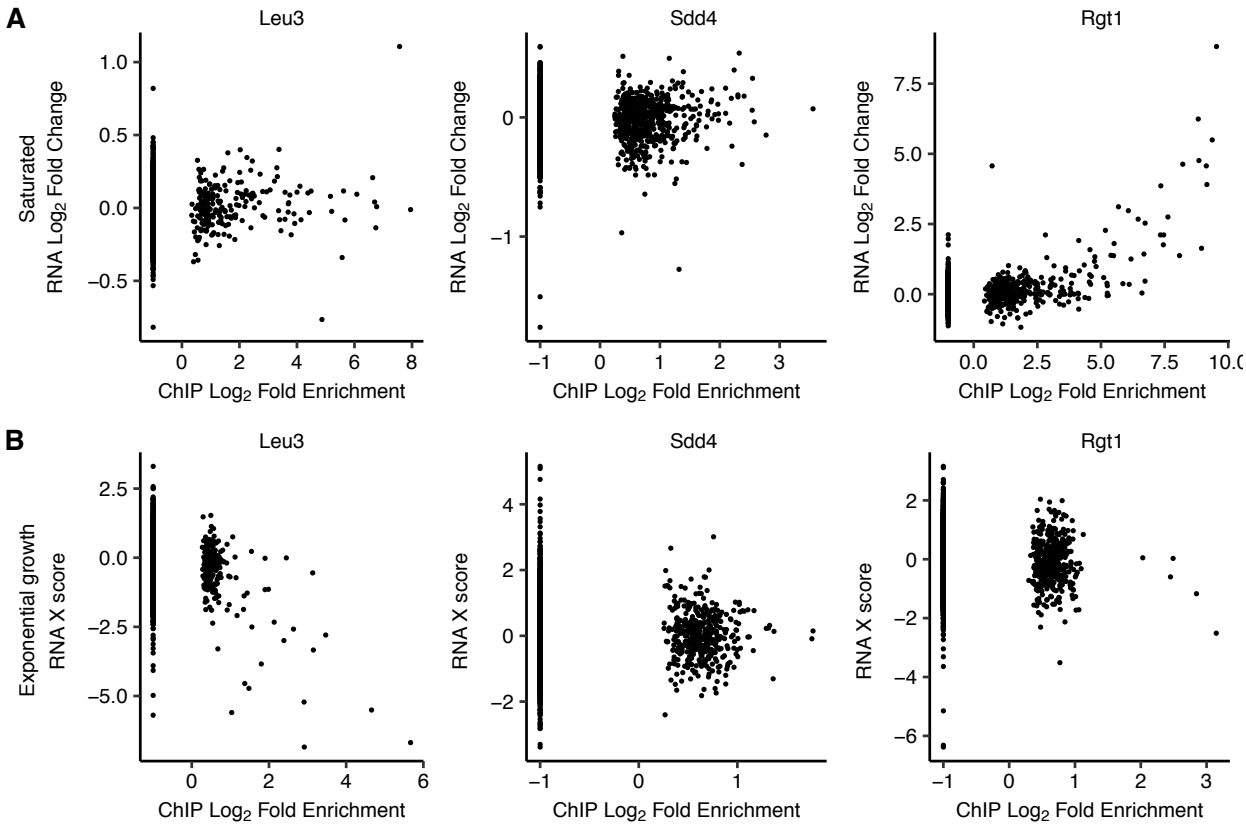

Figure 3.22. Comparison of TF binding targets and differentially expressed genes upon TF deletion.

(A) Genes with upstream Rgt1 ChIP-seq peaks in saturated cultures are upregulated upon *RGT1* deletion. Scatter plots of $\log_2$ ChIP-seq peak fold enrichment vs. $\log_2$ fold change in RNA expression for downstream genes, excluding the deleted TF gene, in saturated culture. (B) Genes with upstream Leu3 ChIP-seq peaks in exponential growth are downregulated upon *LEU3* deletion. Same as panel A but for exponential growth, using microarray "X scores" (normalized, confidence-weighted log ratios) for differential gene expression (Hu et al., 2007).

Chapter 4. OUTLOOK AND FUTURE DIRECTIONS

Note: Chapters 1 and 4 were adapted from a manuscript currently in preparation.

The relationship between 3D organization of the genome and the activity of transcription factors is complex, with interdependencies and several modes of action. Each of the biochemical activities of TFs—binding DNA, opening chromatin, recruiting cofactors, and possibly oligomerizing—may impact 3D genome organization. Conversely, chromosome conformation impacts TF function and the pairing of enhancers with promoters. Nevertheless, we are optimistic that the continuing development and improvement of technologies for measuring and perturbing both genome architecture and TFs offer hope of teasing apart these connections.

The first step is to identify which TFs play a role in 3D genome organization. In well-studied cell lines, existing ChIP-seq datasets can be compared to identify candidates that might shape 3D genome organization (Dixon et al., 2012). However, as attention extends beyond the key TFs that play a role across the genome and most cell types, these analyses will become less feasible. This gap can be complemented by high-throughput screening tools that test perturbations of either DNA sequences or TFs. For example, MAP-C couples saturation mutagenesis and existing mutant libraries to the 3C assay to systematically identify either loop anchor regions and TFs required for a chromosomal contact of interest (Kim et al., 2018a).

However, neither perturbations to TF binding sites nor TF gene/protein expression can determine *how* a TF impacts 3D genome organization. Instead, mutations of protein domains or specific catalytic activities can reveal specific TF functions that are required for genome structuring. For example, deletion of the CTCF RNA-binding domain (Hansen et al., 2018) provides insights into how RNA-binding impacts 3D genome organization.

To determine which TF functions (e.g. cofactor recruitment) are sufficient for structuring the genome, TF bypass experiments can be used. The canonical example is a study in which a zinc finger protein fused to the Ldb1 self-association domain was sufficient to induce looping and activation of globin genes (Deng et al., 2012, 2014). Numerous other studies have used similar TF bypass experiments but did not examine 3D genome organization. Stark and colleagues have used fusions between the Gal4 DNA-binding domain and TFs or cofactors to study TF and cofactor function within enhancers (Stampfel et al., 2015), while others have used dCas9 fusion proteins to recruit histone acetyltransferase (Hilton et al., 2015), DNA methyltransferase (Pflueger et al., 2018), and other chromatin remodelers. Catalytic or domain mutations can be combined with TF bypass experiments to further narrow down specific cofactor functions, such as SUV39H1 binding to H3K9me3, necessary for restructuring the genome (Wijchers et al., 2016). In addition, dCas9 and other DNA binding domains have been used to perturb specific aspects of 3D genome organization (Finlan et al., 2008; Morgan et al., 2017; Pollex and Heard, 2019; Reddy et al., 2008; Wang et al., 2018; Wijchers et al., 2016), and could be used to study their effects on TF function.

An exciting alternative approach for dissecting the mechanisms shaping 3D genome organization is dCas9 proteomics. Studies of the proteins binding to DNA generally require *a priori* knowledge of proteins of interest, as they must be labeled via genomic addition of epitope tags or with specific antibodies. Recent studies have coupled mass spectrometry with enrichment for the proteins in the proximity of a specific genomic locus, by targeting dCas9 either fused with to a biotinylated tag (Liu et al., 2017) or APEX2 (Gao et al., 2018; Myers et al., 2018), which can create radicals that fuse added biotin-phenol to nearby proteins. These methods will aid the systematic search for proteins involved in chromosome conformation.

In addition to perturbations, spatiotemporal resolution is another means of clarifying mechanism. Although we have focused on 3C-based technologies for studying 3D genome organization, microscopy methods are dramatically improving in resolution, throughput, and scalability (Bintu et al., 2018; Mateo et al., 2019). Live imaging methods in particular allow the tracking of individual cells, genomic loci, or even molecules, over the course of minutes or hours (Gu et al., 2018; Mir et al., 2018). As the capacity to simultaneously label features grows, we may eventually be able to see how different TFs work in the context of the 3D genome.

REFERENCES

Ahmed, S., Brickner, D.G., Light, W.H., Cajigas, I., McDonough, M., Froyshteter, A.B., Volpe, T., and Brickner, J.H. (2010). DNA zip codes control an ancient mechanism for gene targeting to the nuclear periphery. Nat. Cell Biol. *12*, 111–118.

Akhtar, A., and Gasser, S.M. (2007). The nuclear envelope and transcriptional control. Nat. Rev. Genet. *8*, 507–517.

Alberti, S. (2017). Phase separation in biology. Curr. Biol. *27*, R1097–R1102.

Alipour, E., and Marko, J.F. (2012). Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. *40*, 11202–11212.

Allen, B.L., and Taatjes, D.J. (2015). The Mediator complex: a central integrator of transcription. Nat. Rev. Mol. Cell Biol. *16*, 155–166.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andrulis, E.D., Neiman, A.M., Zappulla, D.C., and Sternglanz, R. (1998). Perinuclear localization of chromatin facilitates transcriptional silencing. Nature *394*, 592–595.

Apostolou, E., and Thanos, D. (2008). Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. Cell *134*, 85–96.

Badis, G., Chan, E.T., van Bakel, H., Pena-Castillo, L., Tillo, D., Tsui, K., Carlson, C.D., Gossett, A.J., Hasinoff, M.J., Warren, C.L., et al. (2008). A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. Mol. Cell *32*, 878–887.

Bailey, T.L., Williams, N., Misleh, C., and Li, W.W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. *34*, W369–W373.

Banani, S.F., Lee, H.O., Hyman, A.A., and Rosen, M.K. (2017). Biomolecular condensates: organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. *18*, 285–298.

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J., and Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. EMBO J. *28*, 3785–3798.

Bartman, C.R., Hsu, S.C., Hsiung, C.C.-S., Raj, A., and Blobel, G.A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. Mol. Cell *62*, 237–247.

Bartman, C.R., Hamagami, N., Keller, C.A., Giardine, B., Hardison, R.C., Blobel, G.A., and Raj, A. (2019). Transcriptional Burst Initiation and Polymerase Pause Release Are Key Control Points of Transcriptional Regulation. Mol. Cell *73*, 519–532.e4.

Barzel, A., and Kupiec, M. (2008). Finding a match: how do homologous sequences get together for recombination? Nat. Rev. Genet. *9*, 27–37.

Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., and Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science *362*.

Boehning, M., Dugast-Darzacq, C., Rankovic, M., Hansen, A.S., Yu, T., Marie-Nelly, H., McSwiggen, D.T., Kokic, G., Dailey, G.M., Cramer, P., et al. (2018). RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat. Struct. Mol. Biol. *25*, 833–840.

de Boer, C.G., and Hughes, T.R. (2012). YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. Nucleic Acids Res. *40*, D169–D179.

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., et al. (2018). Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell *175*, 1842–1855.e16.

Bonev, B., and Cavalli, G. (2016). Organization and function of the 3D genome. Nat. Rev. Genet. *17*, 661–678.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557–572.e24.

Braun, B.R., and Johnson, A.D. (1997). Control of filament formation in Candida albicans by the transcriptional repressor TUP1. Science *277*, 105–109.

Brickner, D.G., and Brickner, J.H. (2010). Cdk phosphorylation of a nucleoporin controls localization of active genes through the cell cycle. Mol. Biol. Cell *21*, 3421–3432.

Brickner, J.H., and Walter, P. (2004). Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol. *2*, e342.

Brickner, D.G., Ahmed, S., Meldi, L., Thompson, A., Light, W., Young, M., Hickman, T.L., Chu, F., Fabre, E., and Brickner, J.H. (2012). Transcription factor binding to a DNA zip code controls interchromosomal clustering at the nuclear periphery. Dev. Cell *22*, 1234–1246.

Brickner, D.G., Coukos, R., and Brickner, J.H. (2015). INO1 transcriptional memory leads to DNA zip code-dependent interchromosomal clustering. Microb Cell *2*, 481–490.

Brickner, D.G., Sood, V., Tutucci, E., Coukos, R., Viets, K., Singer, R.H., and Brickner, J.H. (2016). Subnuclear positioning and interchromosomal clustering of the GAL1-10 locus are controlled by separable, interdependent mechanisms. Mol. Biol. Cell *27*, 2980–2993.

Burgess, S.M., and Kleckner, N. (1999). Collisions between yeast chromosomal loci in vivo are governed by three layers of organization. Genes Dev. *13*, 1871–1883.

Burgess, S.M., Kleckner, N., and Weiner, B.M. (1999). Somatic pairing of homologs in budding yeast: existence and modulation. Genes Dev. *13*, 1627–1641.

Burton, J.N., Liachko, I., Dunham, M.J., and Shendure, J. (2014). Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. G3 *4*, 1339–1346.

Casolari, J.M., Brown, C.R., Komili, S., West, J., Hieronymus, H., and Silver, P.A. (2004). Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. Cell *117*, 427–439.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018a). Dynamic interplay between enhancer–promoter topology and gene activity. Nat. Genet. *50*, 1296–1303.

Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., van Steensel, B., Ma, J., and Belmont, A.S. (2018b). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J. Cell Biol. *217*, 4025–4048.

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y.-W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., et al. (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. Mol. Cell. Biol. *27*, 1631–1648.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods *13*, 1050–1054.

Cho, W.-K., Jayanth, N., English, B.P., Inoue, T., Andrews, J.O., Conway, W., Grimm, J.B., Spille, J.-H., Lavis, L.D., Lionnet, T., et al. (2016). RNA Polymerase II cluster dynamics predict mRNA output in living cells. Elife *5*.

Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415.

Chong, S., Dugast-Darzacq, C., Liu, Z., Dong, P., Dailey, G.M., Cattoglio, C., Heckert, A., Banala, S., Lavis, L., Darzacq, X., et al. (2018). Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science *361*.

Chowdhary, S., Kainth, A.S., and Gross, D.S. (2017). Heat Shock Protein Genes Undergo Dynamic Alteration in Their Three-Dimensional Structure and Genome Organization in Response to Thermal Stress. Mol. Cell. Biol. *37*.

Chowdhary, S., Kainth, A.S., Pincus, D., and Gross, D.S. (2019). Heat Shock Factor 1 Drives Intergenic Association of Its Target Gene Loci upon Heat Shock. Cell Rep. *26*, 18–28.e5.

Chu, D.B., Gromova, T., Newman, T.A.C., and Burgess, S.M. (2017). The Nucleoporin Nup2 Contains a Meiotic-Autonomous Region that Promotes the Dynamic Chromosome Events of Meiosis. Genetics *206*, 1319–1337.

Cremer, T., and Cremer, M. (2010). Chromosome territories. Cold Spring Harb. Perspect. Biol. *2*, a003889.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A.P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. Cell *160*, 191–203.

Danilowicz, C., Lee, C.H., Kim, K., Hatch, K., Coljee, V.W., Kleckner, N., and Prentiss, M. (2009). Single molecule detection of direct, homologous, DNA/DNA pairing. Proc. Natl. Acad. Sci. U. S. A. *106*, 19824–19829.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., et al. (2014). Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. Cell *158*, 849–860.

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

Derkatch, I.L., Uptain, S.M., Outeiro, T.F., Krishnan, R., Lindquist, S.L., and Liebman, S.W. (2004). Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI+] prion in yeast and aggregation of Sup35 in vitro. Proc. Natl. Acad. Sci. U. S. A. *101*, 12934–12939.

Dieppois, G., Iglesias, N., and Stutz, F. (2006). Cotranscriptional recruitment to the mRNA export receptor Mex67p contributes to nuclear pore anchoring of activated genes. Mol. Cell. Biol. *26*, 7858–7870.

D'Ippolito, A.M., McDowell, I.C., Barrera, A., Hong, L.K., Leichter, S.M., Bartelt, L.C., Vockley, C.M., Majoros, W.H., Safi, A., Song, L., et al. (2018). Pre-established Chromatin Interactions Mediate the Genomic Response to Glucocorticoids. Cell Syst.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dodd, I.B., Shearwin, K.E., Perkins, A.J., Burr, T., Hochschild, A., and Egan, J.B. (2004). Cooperativity in long-range gene regulation by the lambda CI repressor. Genes Dev. *18*, 344–354.

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. Nature *460*, 128.

Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: a protein secondary structure prediction server. Nucleic Acids Res. *43*, W389–W394.

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y.J., Lee, C., Shendure, J., Fields, S., Blau, C.A., and Noble, W.S. (2010). A three-dimensional model of the yeast genome. Nature *465*, 363–367.

Dultz, E., Tjong, H., Weider, E., Herzog, M., Young, B., Brune, C., Müllner, D., Loewen, C., Alber, F., and Weis, K. (2016). Global reorganization of budding yeast chromosome conformation in different physiological conditions. J. Cell Biol. *212*, 321–334.

D'Urso, A., and Brickner, J.H. (2017). Epigenetic transcriptional memory. Curr. Genet. *63*, 435–439.

D'Urso, A., Takahashi, Y.-H., Xiong, B., Marone, J., Coukos, R., Randise-Hinchliff, C., Wang, J.-P., Shilatifard, A., and Brickner, J.H. (2016). Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. Elife *5*, 16691.

Egecioglu, D.E., D'Urso, A., Brickner, D.G., Light, W.H., and Brickner, J.H. (2014). Chapter 21 - Approaches to Studying Subnuclear Organization and Gene–Nuclear Pore Interactions. In Methods in Cell Biology, V. Doye, ed. (Academic Press), pp. 463–485.

Fang, R., Yu, M., Li, G., Chee, S., Liu, T., Schmitt, A.D., and Ren, B. (2016). Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. Cell Res. *26*, 1345–1348.

Farley, E.K., Olson, K.M., Zhang, W., Brandt, A.J., Rokhsar, D.S., and Levine, M.S. (2015). Suboptimization of developmental enhancers. Science *350*, 325–328.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S. (2016). Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. Proc. Natl. Acad. Sci. U. S. A. *113*, 6508–6513.

Finlan, L.E., Sproul, D., Thomson, I., Boyle, S., Kerr, E., Perry, P., Ylstra, B., Chubb, J.R., and Bickmore, W.A. (2008). Recruitment to the nuclear periphery can alter expression of genes in human cells. PLoS Genet. *4*, e1000039.

Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. (2000). Chromosomal evolution in Saccharomyces. Nature *405*, 451–454.

Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110–114.

Fong, K.-W., Li, Y., Wang, W., Ma, W., Li, K., Qi, R.Z., Liu, D., Songyang, Z., and Chen, J. (2013). Whole-genome screening identifies proteins localized to distinct nuclear bodies. J. Cell Biol. *203*, 149–164.

Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Rep. *15*, 2038–2049.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Fulco, C.P., Munschauer, M., Anyoha, R., Munson, G., Grossman, S.R., Perez, E.M., Kane, M., Cleary, B., Lander, E.S., and Engreitz, J.M. (2016). Systematic mapping of functional enhancer-promoter connections with CRISPR interference. Science *354*, 769–773.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Patwardhan, T.A., Nguyen, T.H., et al. (2019). Activity-by-Contact model of enhancer specificity from thousands of CRISPR perturbations. bioRxiv 529990.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y.B., Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Furlong, E.E.M., and Levine, M. (2018). Developmental enhancers and chromosome topology. Science *361*, 1341–1345.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102–105.

Gao, X.D., Tu, L.-C., Mir, A., Rodriguez, T., Ding, Y., Leszyk, J., Dekker, J., Shaffer, S.A., Zhu, L.J., Wolfe, S.A., et al. (2018). C-BERST: defining subnuclear proteomic landscapes at genomic elements with dCas9-APEX2. Nat. Methods *15*, 433–436.

Gasperini, M., Starita, L., and Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. Nat. Protoc. *11*, 1782–1787.

Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377–390.e19.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

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

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Gladyshev, E., and Kleckner, N. (2014). Direct recognition of homology between double helices of DNA in Neurospora crassa. Nat. Commun. *5*, 3509.

Gladyshev, E., and Kleckner, N. (2017). DNA sequence homology induces cytosine-to-thymine mutation by a heterochromatin-related pathway in Neurospora. Nat. Genet. *49*, 887–894.

González, S.S., Barrio, E., Gafner, J., and Querol, A. (2006). Natural hybrids from Saccharomyces cerevisiae, Saccharomyces bayanus and Saccharomyces kudriavzevii in wine fermentations. FEMS Yeast Res. *6*, 1221–1234.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Grant, C.E., Johnson, J., Bailey, T.L., and Noble, W.S. (2015). MCAST: scanning forcis-regulatory motif clusters. Bioinformatics *32*, 1217–1219.

Griffith, J., Hochschild, A., and Ptashne, M. (1986). DNA loops induced by cooperative binding of lambda repressor. Nature *322*, 750–752.

Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050–1055.

Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., et al. (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948–951.

Guertin, M.J., and Lis, J.T. (2010). Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. *6*, e1001114.

Guidi, M., Ruault, M., Marbouty, M., Loïodice, I., Cournac, A., Billaudeau, C., Hocher, A., Mozziconacci, J., Koszul, R., and Taddei, A. (2015). Spatial reorganization of telomeres in long-lived quiescent cells. Genome Biol. *16*, 206.

Güldener, U., Heck, S., Fielder, T., Beinhauer, J., and Hegemann, J.H. (1996). A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. *24*, 2519–2524.

Guo, Y., Monahan, K., Wu, H., Gertz, J., Varley, K.E., Li, W., Myers, R.M., Maniatis, T., and Wu, Q. (2012). CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. Proc. Natl. Acad. Sci. U. S. A. *109*, 21081–21086.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell *169*, 693–707.e14.

Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Darzacq, X., and Tjian, R. (2018). An RNA-binding region regulates CTCF clustering and chromatin looping.

Henikoff, S., and Dreesen, T.D. (1989). Trans-inactivation of the Drosophila brown gene: evidence for transcriptional repression and somatic pairing dependence. Proc. Natl. Acad. Sci. U. S. A. *86*, 6704–6708.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Hogan, M.S., Parfitt, D.-E., Zepeda-Mendoza, C.J., Shen, M.M., and Spector, D.L. (2015). Transient pairing of homologous Oct4 alleles accompanies the onset of embryonic stem cell differentiation. Cell Stem Cell *16*, 275–288.

Hu, Z., Killion, P.J., and Iyer, V.R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. Nat. Genet. *39*, 683–687.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. EMBO J. *14*, 2570–2579.

Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell *167*, 1369–1384.e19.

Jeon, Y., and Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Jin, Q., Trelles-Sticken, E., Scherthan, H., and Loidl, J. (1998). Yeast nuclei display prominent centromere clustering that is reduced in nondividing cells and in meiotic prophase. J. Cell Biol. *141*, 21–29.

Jolma, A., Yin, Y., Nitta, K.R., Dave, K., Popov, A., Taipale, M., Enge, M., Kivioja, T., Morgunova, E., and Taipale, J. (2015). DNA-dependent formation of transcription factor pairs alters their binding specificity. Nature *527*, 384–388.

Jorgensen, P., Edgington, N.P., Schneider, B.L., Rupes, I., Tyers, M., and Futcher, B. (2007). The size of the nucleus increases as yeast cells grow. Mol. Biol. Cell *18*, 3523–3532.

Joyce, E.F., Williams, B.R., Xie, T., and -ting Wu, C. (2012). Identification of Genes That Promote or Antagonize Somatic Homolog Pairing Using a High-Throughput FISH–Based Screen. PLoS Genet. *8*, e1002667.

Joyce, E.F., Erceg, J., and Wu, C.-T. (2016). Pairing and anti-pairing: a balancing act in the diploid genome. Curr. Opin. Genet. Dev. *37*, 119–128.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kaps, S., Kettner, K., Migotti, R., Kanashova, T., Krause, U., Rödel, G., Dittmar, G., and Kriegel, T.M. (2015). Protein kinase Ymr291w/Tda1 is essential for glucose signaling in Saccharomyces cerevisiae on the level of hexokinase isoenzyme ScHxk2 phosphorylation. J. Biol. Chem.

Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. (2003). The UCSC Genome Browser Database. Nucleic Acids Res. *31*, 51–54.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241–254.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Kettner, K., Krause, U., Mosler, S., Bodenstein, C., Kriegel, T.M., and Rödel, G. (2012). Saccharomyces cerevisiae gene YMR291W/TDA1 mediates the in vivo phosphorylation of hexokinase isoenzyme 2 at serine-15. FEBS Lett. *586*, 455–458.

Kim, J.-H., Polish, J., and Johnston, M. (2003). Specificity and regulation of DNA binding by the yeast glucose transporter gene repressor Rgt1. Mol. Cell. Biol. *23*, 5208–5216.

Kim, S., Liachko, I., Brickner, D.G., Cook, K., Noble, W.S., Brickner, J.H., Shendure, J., and Dunham, M. (2017). The dynamic three-dimensional organization of the diploid yeast genome. eLife *6*, 23623.

Kim, S., Dunham, M., and Shendure, J. (2018a). A combination of transcription factors mediates inducible interchromosomal pairing. bioRxiv 385047.

Kim, Y.H., Marhon, S.A., Zhang, Y., Steger, D.J., Won, K.-J., and Lazar, M.A. (2018b). Rev-erbα dynamically modulates chromatin looping to control circadian gene transcription. Science *359*, 1274–1277.

Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat. Cell Biol. 1.

Kung, J.T., Kesner, B., An, J.Y., Ahn, J.Y., Cifuentes-Rojas, C., Colognori, D., Jeon, Y., Szanto, A., del Rosario, B.C., Pinter, S.F., et al. (2015). Locus-specific targeting to the X chromosome revealed by the RNA interactome of CTCF. Mol. Cell *57*, 361–375.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larson, A.G., Elnatan, D., Keenen, M.M., Trnka, M.J., Johnston, J.B., Burlingame, A.L., Agard, D.A., Redding, S., and Narlikar, G.J. (2017). Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. Nature *547*, 236–240.

Le, D.D., Shimko, T.C., Aditham, A.K., Keys, A.M., Longwell, S.A., Orenstein, Y., and Fordyce, P.M. (2018). Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding. Proc. Natl. Acad. Sci. U. S. A. *115*, E3702–E3711.

Levo, M., Zalckvar, E., Sharon, E., Dantas Machado, A.C., Kalma, Y., Lotam-Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., and Segal, E. (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Res. *25*, 1018–1029.

Levy, S.F., Blundell, J.R., Venkataram, S., Petrov, D.A., Fisher, D.S., and Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. Nature *519*, 181–186.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Light, W.H., Brickner, D.G., Brand, V.R., and Brickner, J.H. (2010). Interaction of a DNA zip code with the nuclear pore complex promotes H2A.Z incorporation and INO1 transcriptional memory. Mol. Cell *40*, 112–125.

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., et al. (2009). Population genomics of domestic and wild yeasts. Nature *458*, 337–341.

Liu, J., Perumal, N.B., Oldfield, C.J., Su, E.W., Uversky, V.N., and Dunker, A.K. (2006a). Intrinsic disorder in transcription factors. Biochemistry *45*, 6873–6888.

Liu, X., Lee, C.-K., Granek, J.A., Clarke, N.D., and Lieb, J.D. (2006b). Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. Genome Res. *16*, 1517–1528.

Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Cao, H., Ni, M., Liu, Y., Gu, Z., et al. (2017). In Situ Capture of Chromatin Interactions by Biotinylated dCas9. Cell *170*, 1028–1043.e19.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403–413.

Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Lorenz, A., Fuchs, J., Bürger, R., and Loidl, J. (2003). Chromosome pairing does not contribute to nuclear architecture in vegetative yeast cells. Eukaryot. Cell *2*, 856–866.

Love, M.I., Huber, W., and Anders, S. (2014a). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Love, P.E., Warzecha, C., and Li, L. (2014b). Ldb1 complexes: the new master regulators of erythroid gene transcription. Trends Genet. *30*, 1–9.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Luthra, R., Kerr, S.C., Harreman, M.T., Apponi, L.H., Fasken, M.B., Ramineni, S., Chaurasia, S., Valentini, S.R., and Corbett, A.H. (2007). Actively transcribed GAL genes can be physically linked to the nuclear pore by the SAGA chromatin modifying complex. J. Biol. Chem. *282*, 3042–3049.

Maass, P.G., Barutcu, A.R., Weiner, C.L., and Rinn, J.L. (2018). Inter-chromosomal Contact Properties in Live-Cell Imaging and in Hi-C. Mol. Cell *70*, 188–189.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Mateo, L.J., Murphy, S.E., Hafner, A., Cinquini, I.S., Walker, C.A., and Boettiger, A.N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. Nature *568*, 49–54.

McIsaac, R.S., Gibney, P.A., Chandran, S.S., Benjamin, K.R., and Botstein, D. (2014). Synthetic biology tools for programming gene expression without nutritional perturbations in Saccharomyces cerevisiae. Nucleic Acids Res. *42*, e48–e48.

Mertens, S., Steensels, J., Saels, V., De Rouck, G., Aerts, G., and Verstrepen, K.J. (2015). A large set of newly created interspecific Saccharomyces hybrids increases aromatic diversity in lager beers. Appl. Environ. Microbiol. *81*, 8202–8214.

Mészáros, B., Erdős, G., and Dosztányi, Z. (2018). IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. *46*, W329–W337.

Metz, C.W. (1916). Chromosome studies on the Diptera. II. The paired association of chromosomes in the Diptera, and its significance. J. Exp. Zool. *21*, 213–279.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35–45.

Miele, A., Bystricky, K., and Dekker, J. (2009). Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. PLoS Genet. *5*, e1000478.

Mifsud, B., Tavares-Cadete, F., Young, A.N., Sugar, R., Schoenfelder, S., Ferreira, L., Wingett, S.W., Andrews, S., Grey, W., Ewels, P.A., et al. (2015). Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat. Genet. *47*, 598–606.

Mir, M., Reimer, A., Haines, J.E., Li, X.-Y., Stadler, M., Garcia, H., Eisen, M.B., and Darzacq, X. (2017). Dense Bicoid hubs accentuate binding along the morphogen gradient. Genes Dev. *31*, 1784–1794.

Mir, M., Stadler, M.R., Ortiz, S.A., Hannon, C.E., Harrison, M.M., Darzacq, X., and Eisen, M.B. (2018). Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. Elife *7*.

Mirkin, E.V., Chang, F.S., and Kleckner, N. (2013). Dynamic trans interactions in yeast chromosomes. PLoS One *8*, e75895.

Mirkin, E.V., Chang, F.S., and Kleckner, N. (2014). Protein-Mediated Chromosome Pairing of Repetitive Arrays. J. Mol. Biol. *426*, 550–557.

Mitchell, J.A., and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. Genes Dev. *22*, 20–25.

Monahan, K., Horta, A., and Lomvardas, S. (2019). LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. Nature *565*, 448–453.

Moody, S.E., Perez, D., Pan, T.-C., Sarkisian, C.J., Portocarrero, C.P., Sterner, C.J., Notorfrancesco, K.L., Cardiff, R.D., and Chodosh, L.A. (2005). The transcriptional repressor Snail promotes mammary tumor recurrence. Cancer Cell *8*, 197–209.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.-F., et al. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat. Commun. *8*, 15993.

Morris, J.R., Geyer, P.K., and Wu, C.T. (1999). Core promoter elements can regulate transcription on a separate chromosome in trans. Genes Dev. *13*, 253–258.

Mortimer, R.K. (1958). Radiobiological and genetic studies on a polyploid series (haploid to hexaploid) of Saccharomyces cerevisiae. Radiat. Res. *9*, 312–326.

Muller, H., Scolari, V.F., Agier, N., Piazza, A., Thierry, A., Mercy, G., Descorps-Declere, S., Lazar-Stefanita, L., Espeli, O., Llorente, B., et al. (2018). Characterizing meiotic chromosomes' structure and pairing using a designer sequence optimized for Hi-C. Mol. Syst. Biol. *14*, e8293.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat. Methods *13*, 919–922.

Muñoz, S., Minamino, M., Casas-Delucchi, C.S., Patel, H., and Uhlmann, F. (2019). A Role for Chromatin Remodeling in Cohesin Loading onto Chromosomes. Mol. Cell.

Murrell, A., Heeson, S., and Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. *36*, 889–893.

Myers, S.A., Wright, J., Peckner, R., Kalish, B.T., Zhang, F., and Carr, S.A. (2018). Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. Nat. Methods *15*, 437–439.

Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annu. Rev. Biochem. *70*, 475–501.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471–481.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Nolis, I.K., McKay, D.J., Mantouvalou, E., Lomvardas, S., Merika, M., and Thanos, D. (2009). Transcription factors mediate long-range enhancer–promoter interactions. Proc. Natl. Acad. Sci. U. S. A. *106*, 20222–20227.

Noordermeer, D., de Wit, E., Klous, P., van de Werken, H., Simonis, M., Lopez-Jones, M., Eussen, B., de Klein, A., Singer, R.H., and de Laat, W. (2011). Variegated gene expression caused by cell-specific long-range DNA interactions. Nat. Cell Biol. *13*, 944–951.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.

Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N., and Mirny, L.A. (2018). Chromatin organization by an interplay of loop extrusion and compartmental segregation. Proc. Natl. Acad. Sci. U. S. A. *115*, E6697–E6706.

O'Malley, R.C., Huang, S.-S.C., Song, L., Lewsey, M.G., Bartlett, A., Nery, J.R., Galli, M., Gallavotti, A., and Ecker, J.R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell *166*, 1598.

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. Nat. Genet. *36*, 1065.

Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., and Boeke, J.D. (2004). A robust toolkit for functional profiling of the yeast genome. Mol. Cell *16*, 487–496.

Pant, V., Kurukuti, S., Pugacheva, E., Shamsuddin, S., Mariano, P., Renkawitz, R., Klenova, E., Lobanenkov, V., and Ohlsson, R. (2004). Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. Mol. Cell. Biol. *24*, 3497–3504.

Pascual-Garcia, P., Debo, B., Aleman, J.R., Talamas, J.A., Lan, Y., Nguyen, N.H., Won, K.J., and Capelson, M. (2017). Metazoan Nuclear Pores Provide a Scaffold for Poised Genes and Mediate Induced Enhancer-Promoter Contacts. Mol. Cell *66*, 63–76.e6.

Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'er, D., and Shendure, J. (2009). High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol. *27*, 1173–1175.

Payen, C., Sunshine, A.B., Ong, G.T., Pogachar, J.L., Zhao, W., and Dunham, M.J. (2016). High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations. PLoS Genet. *12*, e1006339.

Pflueger, C., Tan, D., Swain, T., Nguyen, T., Pflueger, J., Nefzger, C., Polo, J.M., Ford, E., and Lister, R. (2018). A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. Genome Res. *28*, 1193–1206.

Pherson, M., Misulovin, Z., Gause, M., and Dorsett, D. (2019). Cohesin occupancy and composition at enhancers and promoters are linked to DNA replication origin proximity in Drosophila. Genome Res. *29*, 602–612.

Polish, J.A., Kim, J.-H., and Johnston, M. (2005). How the Rgt1 transcription factor of Saccharomyces cerevisiae is regulated by glucose. Genetics *169*, 583–594.

Pollex, T., and Heard, E. (2019). Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation. Nat. Genet. *51*, 285–295.

Priest, D.G., Cui, L., Kumar, S., Dunlap, D.D., Dodd, I.B., and Shearwin, K.E. (2014). Quantitation of the DNA tethering effect in long-range DNA looping in vivo and in vitro using the Lac and λ repressors. Proc. Natl. Acad. Sci. U. S. A. *111*, 349–354.

Qian, Z., Dimitriadis, E.K., Edgar, R., Eswaramoorthy, P., and Adhya, S. (2012). Galactose repressor mediated intersegmental chromosomal connections in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *109*, 11336–11341.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744–757.e24.

Randise-Hinchliff, C., and Brickner, J.H. (2016). Transcription factors dynamically control the spatial organization of the yeast genome. Nucleus *7*, 369–374.

Randise-Hinchliff, C., Coukos, R., Sood, V., Sumner, M.C., Zdraljevic, S., Meldi Sholl, L., Garvey Brickner, D., Ahmed, S., Watchmaker, L., and Brickner, J.H. (2016). Strategies to regulate transcription factor-mediated gene positioning and interchromosomal clustering at the nuclear periphery. J. Cell Biol. *212*, 633–646.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–320.e24.

Reddy, K.L., Zullo, J.M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. Nature *452*, 243–247.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A.S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J. Cell Biol. *135*, 1685–1700.

Roop, J.I., Chang, K.C., and Brem, R.B. (2016). Polygenic evolution of a sugar specialization trade-off in yeast. Nature *530*, 336–339.

Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat. Rev. Genet. *19*, 789–800.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I.S.M., Hermetz, K., Wang, P., Ruan, Y., and Corces, V.G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol. Cell *67*, 837–852.e7.

Roy, A., Shin, Y.J., Cho, K.H., and Kim, J.-H. (2013). Mth1 regulates the interaction between the Rgt1 repressor and the Ssn6-Tup1 corepressor complex by modulating PKA-dependent phosphorylation of Rgt1. Mol. Biol. Cell *24*, 1493–1503.

Rutledge, M.T., Russo, M., Belton, J.-M., Dekker, J., and Broach, J.R. (2015). The yeast genome undergoes significant topological reorganization in quiescence. Nucleic Acids Res. *43*, 8299–8313.

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., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U. S. A. *112*, E6456–E6465.

Sankaran, V.G., Menne, T.F., Xu, J., Akie, T.E., Lettre, G., Van Handel, B., Mikkola, H.K.A., Hirschhorn, J.N., Cantor, A.B., and Orkin, S.H. (2008). Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A. Science *322*, 1839–1842.

Scannell, D.R., Zill, O.A., Rokas, A., Payen, C., Dunham, M.J., Eisen, M.B., Rine, J., Johnston, M., and Hittinger, C.T. (2011). The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3 *1*, 11–25.

Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., et al. (2016a). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. *17*, 2042–2059.

Schmitt, A.D., Hu, M., and Ren, B. (2016b). Genome-wide mapping and analysis of chromosome architecture. Nat. Rev. Mol. Cell Biol. *17*, 743–755.

Schober, H., Kalck, V., Vega-Palas, M.A., Van Houwe, G., Sage, D., Unser, M., Gartenberg, M.R., and Gasser, S.M. (2008). Controlled exchange of chromosomal arms reveals principles driving telomere interactions in yeast. Genome Res. *18*, 261–271.

Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N.F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J.A., Umlauf, D., Dimitrova, D.S., et al. (2009). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat. Genet. *42*, 53.

Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., Tavares-Cadete, F., Sugar, R., Javierre, B.-M., Nagano, T., Katsman, Y., Sakthidevi, M., Wingett, S.W., et al. (2015). The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res. *25*, 582–597.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C., Mirny, L., et al. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56.

Shalem, O., Sanjana, N.E., and Zhang, F. (2015). High-throughput functional genomics using CRISPR-Cas9. Nat. Rev. Genet. *16*, 299–311.

Sigova, A.A., Abraham, B.J., Ji, X., Molinie, B., Hannett, N.M., Guo, Y.E., Jangi, M., Giallourakis, C.C., Sharp, P.A., and Young, R.A. (2015). Transcription factor trapping by RNA in gene regulatory elements. Science *350*, 978–981.

Sima, J., Chakraborty, A., Dileep, V., Michalski, M., Klein, K.N., Holcomb, N.P., Turner, J.L., Paulsen, M.T., Rivera-Mulia, J.C., Trevilla-Garcia, C., et al. (2019). Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. Cell *176*, 816–830.e18.

Slattery, M., Zhou, T., Yang, L., Dantas Machado, A.C., Gordân, R., and Rohs, R. (2014). Absence of a simple code: how transcription factors read the genome. Trends Biochem. Sci. *39*, 381–399.

Song, S.-H., Hou, C., and Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. Mol. Cell *28*, 810–822.

Sood, V., Cajigas, I., D'Urso, A., Light, W.H., and Brickner, J.H. (2017). Epigenetic Transcriptional Memory of GAL Genes Depends on Growth in Glucose and the Tup1 Transcription Factor in Saccharomyces cerevisiae. Genetics *206*, 1895–1907.

Staby, L., O'Shea, C., Willemoës, M., Theisen, F., Kragelund, B.B., and Skriver, K. (2017). Eukaryotic transcription factors: paradigms of protein intrinsic disorder. Biochem. J *474*, 2509–2532.

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. *50*, 238–249.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151.

van Steensel, B., and Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. Nat. Rev. Mol. Cell Biol.

Straight, A.F., Belmont, A.S., Robinett, C.C., and Murray, A.W. (1996). GFP tagging of budding yeast chromosomes reveals that protein–protein interactions can mediate sister chromatid cohesion. Curr. Biol. *6*, 1599–1608.

Strom, A.R., Emelyanov, A.V., Mir, M., Fyodorov, D.V., Darzacq, X., and Karpen, G.H. (2017). Phase separation drives heterochromatin domain formation. Nature *547*, 241–245.

Struhl, K. (1985). Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc. Natl. Acad. Sci. U. S. A. *82*, 8419–8423.

Su, W., Jackson, S., Tjian, R., and Echols, H. (1991). DNA looping between sites for transcriptional activation: self-association of DNA-bound Sp1. Genes Dev. *5*, 820–826.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. Science *332*, 472–474.

Swygert, S.G., Kim, S., Wu, X., Fu, T., Hsieh, T.-H., Rando, O.J., Eisenman, R.N., Shendure, J., McKnight, J.N., and Tsukiyama, T. (2019). Condensin-Dependent Chromatin Compaction Represses Transcription Globally during Quiescence. Mol. Cell *73*, 533–546.e4.

Sze, J.Y., Woontner, M., Jaehning, J.A., and Kohlhaw, G.B. (1992). In vitro transcriptional activation by a metabolic intermediate: activation by Leu3 depends on alpha-isopropylmalate. Science *258*, 1143–1145.

Taddei, A. (2007). Active genes at the nuclear pore complex. Curr. Opin. Cell Biol. *19*, 305–310.

Taddei, A., Van Houwe, G., Hediger, F., Kalck, V., Cubizolles, F., Schober, H., and Gasser, S.M. (2006). Nuclear pore association confers optimal expression levels for an inducible yeast gene. Nature *441*, 774–778.

Taddei, A., Schober, H., and Gasser, S.M. (2010). The budding yeast nucleus. Cold Spring Harb. Perspect. Biol. *2*, a000612.

Tehranchi, A., Hie, B., Dacre, M., Kaplow, I., Pettie, K., Combs, P., and Fraser, H.B. (2019). Fine-mapping cis-regulatory variants in diverse human populations. Elife *8*.

Terakawa, T., Bisht, S., Eeftens, J.M., Dekker, C., Haering, C.H., and Greene, E.C. (2017). The condensin complex is a mechanochemical motor that translocates along DNA. Science *358*, 672–676.

Therizols, P., Duong, T., Dujon, B., Zimmer, C., and Fabre, E. (2010). Chromosome arm length and nuclear constraints determine the dynamic relationship of yeast subtelomeres. Proc. Natl. Acad. Sci. U. S. A. *107*, 2025–2030.

Tjong, H., Gong, K., Chen, L., and Alber, F. (2012). Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. Genome Res. *22*, 1295–1305.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tsai, A., Muthusamy, A.K., Alves, M.R., Lavis, L.D., Singer, R.H., Stern, D.L., and Crocker, J. (2017). Nuclear microenvironments modulate transcription from low-affinity enhancers. Elife *6*, 28975.

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., et al. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. Cell *175*, 292–294.

Vockley, C.M., D'Ippolito, A.M., McDowell, I.C., Majoros, W.H., Safi, A., Song, L., Crawford, G.E., and Reddy, T.E. (2016). Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. Cell *166*, 1269–1281.e19.

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647.

Wang, D., Hu, Y., Zheng, F., Zhou, K., and Kohlhaw, G.B. (1997). Evidence that intramolecular interactions are involved in masking the activation domain of transcriptional activator Leu3p. J. Biol. Chem. *272*, 19383–19392.

Wang, H., Xu, X., Nguyen, C.M., Liu, Y., Gao, Y., Lin, X., Daley, T., Kipniss, N.H., La Russa, M., and Qi, L.S. (2018). CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. Cell *175*, 1405–1417.e14.

Wang, S., Su, J.-H., Beliveau, B.J., Bintu, B., Moffitt, J.R., Wu, C.-T., and Zhuang, X. (2017). Spatial organization of chromatin domains and compartments in single chromosomes. Biophys. J. *112*, 217a.

Weiner, B.M., and Kleckner, N. (1994). Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. Cell *77*, 977–991.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

Wijchers, P.J., Krijger, P.H.L., Geeven, G., Zhu, Y., Denker, A., Verstegen, M.J.A.M., Valdes-Quezada, C., Vermeulen, C., Janssen, M., Teunissen, H., et al. (2016). Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. Mol. Cell *61*, 461–473.

de Wit, E., and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. Genes Dev. *26*, 11–24.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell *60*, 676–684.

Wong, H., Marie-Nelly, H., Herbert, S., Carrivain, P., Blanc, H., Koszul, R., Fabre, E., and Zimmer, C. (2012). A predictive computational model of the dynamic 3D interphase yeast nucleus. Curr. Biol. *22*, 1881–1890.

Xiao, H., Perisic, O., and Lis, J.T. (1991). Cooperative binding of Drosophila heat shock factor to arrays of a conserved 5 bp unit. Cell *64*, 585–593.

Xu, N., Tsai, C.-L., and Lee, J.T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. Science *311*, 1149–1152.

Xu, N., Donohoe, M.E., Silva, S.S., and Lee, J.T. (2007). Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. Nat. Genet. *39*, 1390.

Yang, C.H., Lambie, E.J., Hardin, J., Craft, J., and Snyder, M. (1989). Higher order structure is present in the yeast nucleus: autoantibody probes demonstrate that the nucleolus lies opposite the spindle pole body. Chromosoma *98*, 123–128.

Yang, F., Deng, X., Ma, W., Berletch, J.B., Rabaia, N., Wei, G., Moore, J.M., Filippova, G.N., Xu, J., Liu, Y., et al. (2015). The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. Genome Biol. *16*, 52.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*.

Yusufzai, T.M., Tagami, H., Nakatani, Y., and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol. Cell *13*, 291–298.

Zabidi, M.A., and Stark, A. (2016). Regulatory Enhancer–Core-Promoter Communication via Transcription Factors and Cofactors. Trends Genet. *32*, 801–814.

Zhang, D., and Bai, L. (2016). Interallelic interaction and gene regulation in budding yeast. Proc. Natl. Acad. Sci. U. S. A. *113*, 4428–4433.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhu, F., Farnung, L., Kaasinen, E., Sahu, B., Yin, Y., Wei, B., Dodonova, S.O., Nitta, K.R., Morgunova, E., Taipale, M., et al. (2018). The interaction landscape between transcription factors and the nucleosome. Nature *562*, 76–81.

VITA

Seungsoo Kim was born in South Korea, but grew up primarily in Vancouver, WA. He graduated *summa cum laude* from Harvard University with an A.B. in Chemical and Physical Biology. As an undergraduate, he worked in the lab of Roy Kishony, where he studied the rates and mechanisms of evolution of antibiotic resistance. He then entered the Genome Sciences Ph.D. program, where he pursued his interest in genomics, particularly 3D genome organization, in the lab of Jay Shendure. Upon graduation, he will become a postdoctoral scholar in the lab of Joanna Wysocka at Stanford.