Transcription Factor Binding to a DNA Zip Code Controls Interchromosomal Clustering at the Nuclear Periphery

Donna Garvey Brickner, 1 Sara Ahmed, 1 Lauren Meldi, 1,4 Abbey Thompson, 1,4 Will Light, 1,4 Matthew Young, 1 Taylor L. Hickman, 2 Feixia Chu, 2 Emmanuelle Fabre, 3,5 and Jason H. Brickner 1,*

1Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA
2Department of Molecular, Cellular and Biomedical Sciences, University of New Hampshire, Durham, NH 03824, USA
3Unité de Génétique Moléculaire des Levures, URA2171 CNRS, Institut Pasteur, rue du Dr Roux, 75015 Paris, France
4These authors contributed equally to this work
5Present address: Groupe Régulation Spatiale des Génomes, UMR 3525 CNRS, Institut Pasteur, rue du Dr Roux, 75015 Paris France

DOI 10.1016/j.devcel.2012.03.012

SUMMARY

Active genes in yeast can be targeted to the nuclear periphery through interaction of cis-acting “DNA zip codes” with the nuclear pore complex. We find that genes with identical zip codes cluster together. This clustering was specific; pairs of genes that were targeted to the nuclear periphery by different zip codes did not cluster together. Insertion of two different zip codes (GRS I or GRS III) at an ectopic site induced clustering with endogenous genes that have that zip code. Targeting to the nuclear periphery and interaction with the nuclear pore is a prerequisite for gene clustering, but clustering can be maintained in the nucleoplasm. Finally, we find that the Put3 transcription factor recognizes the GRS I zip code to mediate both targeting to the NPC and interchromosomal clustering. These results suggest that zip-code-mediated clustering of genes at the nuclear periphery influences the three-dimensional arrangement of the yeast genome.

INTRODUCTION

The nucleus is spatially organized. Chromosomes fold, occupy distinct “territories,” and interact with stable nuclear structures such as the nuclear lamina and the nuclear envelope (Meldi and Brickner, 2011). Furthermore, the position of individual genes within the nucleus can both reflect and impact their expression; coregulated genes can cluster together and physically interact to promote either expression or silencing (Brown et al., 2006, 2008; Schoenfelder et al., 2010). The colocalization of active, coregulated genes has been proposed to occur at “transcription factories” and to promote the efficient recruitment of factors involved in their expression (Iborra et al., 1996; Schoenfelder et al., 2010; Xu and Cook, 2008a). However, the molecular mechanisms controlling gene positioning and clustering are still unclear.

As a model for these phenomena, we have studied the targeting of genes to the nuclear periphery upon activation (Brickner, 2009; Egecioglu and Brickner, 2011; Taddei, 2007). In yeast, many inducible genes, including GAL1, GAL2, INO1, HSP104, and TSA2, are targeted from the nucleoplasm to the nuclear periphery upon activation (Ahmed et al., 2010; Brickner and Walter, 2004; Cabal et al., 2006; Casolari et al., 2004, 2005; Dieppois et al., 2006). Targeting is mediated by physical interaction of the promoters of these genes with the nuclear pore complex (NPC) and results in their constrained diffusion along the nuclear envelope (Ahmed et al., 2010; Cabal et al., 2006; Light et al., 2010; Luthra et al., 2007; Schmid et al., 2006). Targeting to the NPC promotes stronger transcription (Ahmed et al., 2010; Brickner et al., 2007; Brickner and Walter, 2004; Menon et al., 2005; Taddei et al., 2006). A similar phenomenon has been reported in Drosophila; interaction of nuclear pore proteins with genes promotes their transcription (Capelson et al., 2010; Kalverda et al., 2010). However, in Drosophila, many of these genes interact with nuclear pore proteins in the nucleoplasm, away from the NPC (Capelson et al., 2010; Kalverda et al., 2010).

Within the promoters of INO1 and TSA2, we have identified DNA elements that are necessary for targeting to the nuclear periphery and interaction with the NPC (Ahmed et al., 2010; Light et al., 2010). These elements are distinct from the known upstream activating sequences (UASs) that control transcription of these genes. Importantly, these elements function as DNA zip codes: they are sufficient, when introduced at an ectopic locus, to induce targeting to the nuclear periphery (Ahmed et al., 2010; Light et al., 2010). These results suggest that the spatial organization of the genome is, to some extent, encoded in the DNA and that the utilization of such positional information can be regulated (Ahmed and Brickner, 2010).

To explore the role of DNA zip codes and interaction with the NPC in affecting the spatial organization of the genome, we asked if zip codes could cause an ectopic locus to colocalize with the endogenous gene. Here we show that endogenous genes with identical DNA zip codes cluster together, whereas genes with different zip codes do not. Two zip codes from different promoters (GRS I or GRS III), when inserted at an ectopic location, induce colocalization with the endogenous genes having these zip codes. Finally, GRS I-mediated clustering requires interaction with the Put3 transcription factor and interaction with the NPC. These results suggest that DNA...
zip codes can induce gene clustering at the nuclear envelope that has global effects on the spatial organization of the genome.

**RESULTS**

**INO1 Is Targeted to a Restricted Portion of the Nuclear Envelope**

We first asked if the *INO1* gene was localized to the nuclear envelope generally or to a restricted portion of the nuclear envelope. Previous work localizing genes with respect to stable subnuclear structures found that genes in the yeast nucleus localized to restricted subnuclear territories (Berger et al., 2008). Using this high-resolution statistical mapping approach, we mapped the location of the *INO1* gene with respect to the nucleolus, the nuclear envelope and the center of the nucleus under repressing and activating conditions (Figures 1A–1C) (Berger et al., 2008). In cells grown in repressing conditions, *INO1* localized in the nucleoplasm with no obvious bias in its distribution (Figure 1D, left). In cells grown under activating conditions, *INO1* localized at the nuclear periphery preferentially to a position corresponding to $-75^\circ \pm 31^\circ$ acute angle between the line connecting the locus to the center of the nucleus and the axis connecting the center of the nucleus and the center of the nucleolus ($\alpha$; Figures 1B and 1C; Figure S1 and Table S1 available online). We also observed a population of cells in which *INO1* was localized in the nucleoplasm, near the nucleolus (Figure 1C). This population might correspond to cells that are in S-phase, a period of the cell cycle in which peripheral targeting is temporarily lost (Bricker and Brickner, 2010). A similar bimodal distribution has been observed for active GAL1 when it is targeted to the nuclear periphery (Berger et al., 2008). Regardless, when targeted to the nuclear periphery, *INO1* localized to a restricted portion of the envelope.

**Clustering of INO1 upon Activation**

If gene positioning is encoded in cis-acting DNA elements, it might be possible to either target an ectopic site to the same location as the endogenous gene or to induce clustering of genes. To test this idea, we integrated the *INO1* gene from chromosome X beside the *URA3* gene on chromosome V. Like the endogenous *INO1* gene, this hybrid locus (*URA3:INO1*) is targeted to the nuclear periphery upon activation of *INO1* (Ahmed et al., 2010; Light et al., 2010). This allowed us to compare the positions of the endogenous *INO1* gene and the ectopic *URA3:INO1* gene.

We compared the positions of these loci with respect to each other in a strain having an array of Lac repressor binding sites beside *URA3*, an array of Tet repressor binding sites beside *INO1* and expressing LacI-RFP and TetR-GFP (Figure 2A). We measured the distance between the center of the red spot and the center of the green spot for $\geq 100$ fixed cells in which the two dots were within the same confocal section (z depth $\sim 0.7 \, \mu m$) (see Experimental Procedures and Figure 2B). The distances were binned into 0.2 $\mu m$ classes to generate distributions of distances within the population, which we compared using a two-tailed t test. As a negative control, we determined the distribution of distances between active *INO1* (at the nuclear periphery) and *URA3* (in the nucleoplasm). We observed a normal distribution of distances between *INO1* and *URA3*, with a mean distance of $1.08 \pm 0.43 \, \mu m$ (Figure 2C). In contrast, the distances between *INO1* and *URA3:INO1* (under activating conditions) were clearly shifted to shorter distances, with a mean distance of $0.48 \pm 0.28 \, \mu m$ ($p < 0.0001$) (Figure 2C).

Therefore, the introduction of *INO1* at *URA3* caused *URA3* to localize to a similar portion of the nucleus as the endogenous *INO1* gene.

The change in distances between these loci was highlighted when we plotted the fraction of the spot pairs that were qualitatively “clustered” (defined here as a distance $< 0.55 \, \mu m$). The two dimensional area of a circle of diameter 0.55 $\mu m$ (0.24 $\mu m^2$) is $\sim 8\%$ of the two dimensional area of the typical haploid yeast nucleus (3.14 $\mu m^2$; diameter = 2 $\mu m$). Clustering increased from $11\%$ for *INO1* versus *URA3* to $66\%$ for *INO1* versus *URA3:INO1* ($p < 0.0001$, Fischer’s exact test; Figure 2C). Clustering of the two loci was dependent on activation; *INO1* did not cluster with *URA3:INO1* under repressing conditions.
Figure 2. INO1 Clustering

(A) Schematic of experimental strategy. An array of 128 Lac repressor-binding sites was integrated beside URA3 on chromosome V and an array of 112 Tet repressor-binding sites was integrated beside INO1 on chromosome X in a strain expressing GFP-LacI and RFP-TetR. The positions of GRS I and GRS II in the promoter of INO1 are indicated as I and II, respectively. To create URA3:INO1, the INO1 gene was integrated beside URA3.

(B) Example distances between two loci. Cells were fixed, stained with antibodies against GFP and RFP, visualized by line-scanning confocal microscopy and distances between the centers of the spots were measured using Zeiss LSM software. Scale bar represents 1 µm.

(C) Distribution of distances between active INO1 and URA3 or between active INO1 and URA3:INO1. (C–G) Left: plot of the distribution of the frequency of each distance between loci in the population (n ≥ 100 cells). Right: the fraction of cells in which the loci were <0.55 µm apart.

(D) Distribution of distances between either two alleles of URA3 or two alleles of INO1 in diploid cells in repressing conditions.

(E) Distribution of distances between two alleles of INO1 in diploid cells in repressing (same data as in E) and activating conditions.

(F) Distribution of distances between two alleles of INO1 in diploid cells in repressing (same data as in E) and activating conditions.

(G) Distribution of distances between INO1 and GAL1 in cells grown in activating conditions for INO1 and either repressing (glucose) or activating (galactose) conditions for GAL1.
(Figure 2D). Therefore, activation of \( \text{INO}1 \) on chromosome X led to clustering with \( \text{URA}3: \text{INO}1 \) on chromosome V.

The mean distance and clustering between the \( \text{INO}1 \) and \( \text{URA}3 \) (1.08 ± 0.43 μm, 11%) was significantly different than the mean distance and clustering between repressed \( \text{INO}1 \) and \( \text{URA}3: \text{INO}1 \) (0.85 ± 0.38 μm, 20%) (Figures 2C and 2D). This may reflect either the proximity of these two loci in the nucleoplasm when \( \text{INO}1 \) is repressed or a small amount of background expression of \( \text{INO}1 \) under repressing conditions. Consistent with this latter possibility, the localization of repressed \( \text{INO}1 \) or \( \text{URA}3: \text{INO}1 \) at the nuclear periphery is systematically higher than the localization of \( \text{URA}3 \) and the nuclear periphery (Ahmed et al., 2010; Brickner and Walter, 2004). For this reason, we compared against repressed \( \text{URA}3: \text{INO}1 \) for subsequent experiments.

To ask if the two endogenous alleles of \( \text{INO}1 \) in a diploid nucleus cluster upon activation, we used a diploid yeast strain having a Lac repressor array integrated beside one allele of \( \text{INO}1 \) and the Tet repressor array integrated beside the other allele. The mean distance and the clustering between the two copies of repressed \( \text{INO}1 \) were similar to that of two copies of \( \text{URA}3 \) (1.06 ± 0.38 μm versus 1.00 ± 0.47 μm; \( p = 0.2998 \)) (Figure 2E). In contrast, upon activation of \( \text{INO}1 \), the mean distance between the two copies of \( \text{INO}1 \) was significantly reduced (0.60 ± 0.33 μm; \( p < 0.0001 \)) and the clustering was significantly increased (20% versus 52%; \( p < 0.0001 \); Figure 2F). Thus, the clustering of active \( \text{INO}1 \) occurred both in haploid cells between the endogenous gene and an ectopic locus and in diploids between two alleles of the endogenous gene.

**Clustering Is Gene Specific**

To probe the specificity of gene clustering, we localized active \( \text{INO}1 \) with respect to \( \text{GAL}1 \), another gene that is targeted to the nuclear periphery upon activation (Berger et al., 2008; Cabal et al., 2006; Casolari et al., 2004; Schmid et al., 2006). The distributions of distances between \( \text{INO}1 \) and either active or repressed \( \text{GAL}1 \) were indistinguishable (Figure 2G) and were very similar to the distribution of distances between repressed \( \text{INO}1 \) and \( \text{URA}3: \text{INO}1 \) (\( p = 0.4635 \) for active \( \text{GAL}1 \)). Thus, targeting to the nuclear periphery is not sufficient to induce gene clustering.

We performed a series of pairwise comparisons between \( \text{HSP}104 \) and \( \text{GAL}1, \text{INO}1, \) and \( \text{GAL}2 \), genes that localize at the nuclear periphery upon activation. For these experiments, we used a different experimental strategy (Figure 3A). A “large” array of 256 Lac repressor-binding sites was integrated adjacent to \( \text{HSP}104 \) and a “small” array of 128 binding sites adjacent to \( \text{INO}1, \text{GAL}1, \) or \( \text{GAL}2 \). This resulted in strains with a discernibly larger green dot marking \( \text{HSP}104 \) and a small green dot marking \( \text{INO}1, \text{GAL}1, \) or \( \text{GAL}2 \) (Figure 3A). \( \text{HSP}104 \) is targeted to the nuclear periphery upon activation under heat shock or in the presence of 10% ethanol (Dieppois et al., 2006) and the \( \text{GAL}1 \) and \( \text{GAL}2 \) genes are targeted to the nuclear periphery upon activation by growth in galactose (Casolari et al., 2004). We measured the distance between the spots under both uninducing and inducing conditions for these genes. Because we used a different fixation method (methanol instead of formaldehyde; see Experimental Procedures), which causes the cells to shrink slightly (Brickner et al., 2010), the distances between the spots under uninducing conditions were slightly smaller in these experiments than in the experiments using two different fluorescent proteins. However, for all three of these pairwise comparisons, we observed neither a significant change in the distribution of distances between the genes nor significant clustering upon activation (Figures 3B–3D). This was due to the difference in fixation conditions because we were able to observe clustering using this fixation method (see below). Therefore, \( \text{INO}1, \text{GAL}1, \) and \( \text{GAL}2 \) do not cluster with \( \text{HSP}104 \).

\( \text{HSP}104 \) and \( \text{GAL}2 \) are ~290 kb apart and on left and right arms of chromosome XII, separated by ~107 genes (Figure 3D). \( \text{HSP}104 \) and \( \text{GAL}2 \) localize to different parts of the nucleus. Whereas \( \text{GAL}2 \) colocalized with the nucleolus (Berger et al., 2008; Brickner et al., 2010; Gard et al., 2009), presumably because of its proximity to the rDNA genes on the right arm of chromosome XII, \( \text{HSP}104 \) did not (Figures 3E and 3F). Therefore, within chromosome XII, the positioning of \( \text{HSP}104 \) and \( \text{GAL}2 \) at the nuclear periphery is distinct and is likely impacted by neighboring elements such as the centromere and the rDNA locus, which have strong, stable localization patterns (Duan et al., 2010).

**Targeting of \( \text{INO}1 \) to the Nuclear Periphery Is a Prerequisite for Clustering**

Targeting of genes to the nuclear periphery involves the interaction between the nuclear pore complex (NPC) and their promoters (Ahmed et al., 2010; Brickner et al., 2007; Cabal et al., 2006; Casolari et al., 2004, 2005; Light et al., 2010; Schmid et al., 2006; Taddei et al., 2006). Repressed \( \text{INO}1 \) colocalizes with the nuclear envelope in ~30% of the cells in the population and active \( \text{INO}1 \) colocalizes with the nuclear envelope in ~65% of the cells in the population (Brickner et al., 2010; Brickner and Walter, 2004). Similar localization frequencies are seen for other genes that are targeted to the nuclear periphery (Brickner et al., 2007; Casolari et al., 2004). Thus, targeting to the nuclear periphery does not result in localization of a gene to the nuclear periphery in 100% of the cells. This reflects both the dynamic nature of localization—even when localized at the nuclear periphery, genes continue to move (Cabal et al., 2006)—and the regulation of gene localization during the cell cycle; genes like \( \text{INO}1, \text{GAL}1, \) and \( \text{HSP}104 \) lose peripheral localization during S-phase (Brickner and Brickner, 2010, 2011). Therefore, clustering of active \( \text{INO}1 \) could occur at the nuclear periphery, in the nucleoplasm or both.

If targeting to the nuclear periphery were involved in gene clustering, then clustering would be disrupted by mutations in the nuclear pore that block targeting to the NPC. Nup2, part of the nucleoplasmic basket of the NPC, interacts with the active \( \text{INO}1 \) promoter by ChIP and is required for targeting of \( \text{INO}1 \) to the nuclear periphery (Ahmed et al., 2010; Brickner et al., 2007; Light et al., 2010). In cells lacking Nup2, we did not observe clustering of active \( \text{INO}1 \) and \( \text{URA}3: \text{INO}1 \) (Figure 4A). Therefore, interaction of genes with the NPC is required for both peripheral targeting and interchromosomal clustering.

We next asked if clustering of active \( \text{INO}1 \) was strictly correlated with localization at the nuclear periphery. We compared clustering of \( \text{INO}1 \) with \( \text{URA}3: \text{INO}1 \) in three classes of cells: cells in which both genes colocaled with the nuclear envelope (on/on), cells in which neither locus colocaled with the nuclear envelope (off/ off), and cells in which one gene was colocaled with the nuclear envelope and the other was not (off/on). When \( \text{INO}1 \) was active, \( \text{INO}1 \) and \( \text{URA}3: \text{INO}1 \) clustered in all three classes of cells; however, when \( \text{INO}1 \) was repressed, \( \text{INO}1 \) and \( \text{URA}3: \text{INO}1 \) did not cluster in any of the three classes of cells (on/on, off/off, or off/on). These results indicate that targeting of \( \text{INO}1 \) to the nuclear periphery is necessary but not sufficient for gene clustering.
Figure 3. Gene Clustering Is Specific

(A) Left: schematic of two green dot experimental strategy. An array of 256 Lac repressor-binding sites was integrated beside HSP104 and an array of 128 Lac repressor-binding sites was integrated beside other loci. Right: representative confocal micrographs of a strain having a large array and a small array, stained with anti-GFP and anti-myc (to stain myc-tagged Sec63 in the endoplasmic reticulum and nuclear envelope). Scale bar represents 1 μm.

(B) Distribution of distances between HSP104 and INO1, under conditions in which only INO1 is active (–inositol) or conditions in which both genes are active (+inositol +10% ethanol). (B, C, D) Top: chromosomal locations of genes. Left: plot of the distribution of the distances between loci in the population. Right: the fraction of cells in which the loci were <0.55 μm apart.

(C) Distribution of distances between HSP104 and GAL1, under uninducing (glucose) or inducing (galactose +10% ethanol) conditions.

(D) Distribution of distances between HSP104 and GAL2, grown in uninducing or inducing conditions.

(E) Representative images scored for colocalization of HSP104 or GAL2 with either the nuclear envelope (stained with anti-myc for Sec63-myc) or the nucleolus (stained with anti-Nop5/6). Scale bar represents 1 μm.

(F) The fraction of cells in which HSP104 or GAL2 colocalized with the nuclear envelope or the nucleolus (n = 3; 30–50 cells per biological replicate; error bars represent SEM).
before activating INO1 and URA3:INO1 were arrested with hydroxyurea either after activating (C) Cells having the Tet repressor array at envelope.

Clustering of the two loci was determined in cells of each of the three classes: both genes on the nuclear envelope (on/on), one gene on the nuclear envelope (on/off), or both genes off the nuclear envelope (off/off). Clustering was highest in the population of cells in which both loci colocalized with the nuclear envelope (72% clustering; mean distance = 0.40 ± 0.25 µm) and was not apparent in cells in which only one locus colocalized with the nuclear envelope (12.5% clustering; mean distance = 0.75 ± 0.33 µm; Figure 4B). In contrast, if we compared two genes that did not show clustering (INO1 and HSP104) in the cells in which both loci were at the nuclear periphery, they did not cluster (mean distance = 0.88 ± 0.24 µm; data not shown). Thus, for genes that cluster, we observe the highest level of clustering when both loci are at the nuclear periphery.

Surprisingly, we also observed significant clustering in the cells in which both loci localized in the nucleoplasm (59% clustering; mean distance = 0.49 ± 0.32 µm; Figure 4B). This was true when we compared active INO1 with URA3:INO1, but was not true when we quantified the clustering of repressed INO1 with URA3:INO1, URA3, or a second allele of INO1 in the nucleoplasm (Figure 2). Therefore, although clustering correlated with activation and occurred between loci that are targeted to the nuclear periphery by the same mechanism, clustering can also occur in the nucleoplasm.

We hypothesized that targeting to the nuclear periphery was a prerequisite step to establish clustering, which could be maintained in the nucleoplasm. To test this idea, we compared the clustering of INO1 in cells arrested during S phase before or after being targeted to the nuclear periphery. Treating cells with hydroxyurea traps cells in S phase, a period of the cell cycle when peripheral localization of INO1 is lost (Brickner and Brickner, 2010). In cells in which INO1 and URA3:INO1 were already active before treating with hydroxyurea, they remained clustered after arrest (Figure 4C). Therefore, clustering was maintained in the nucleoplasm during S-phase in cells in which the two loci had previously been targeted to the nuclear periphery.

If clustering in the nucleoplasm requires previous targeting to the nuclear periphery, then inducing INO1 after cells have been arrested in S-phase should not result in clustering. We found that INO1 and URA3:INO1 did not cluster in cells that were starved for inositol after they were arrested with hydroxyurea (Figure 4C). This suggests that targeting to the nuclear periphery is a prerequisite for gene clustering and that, once established, clustering can be maintained in the nucleoplasm.

DNA Zip Codes Control Gene Clustering

Given the importance of targeting to the nuclear periphery for gene clustering, we asked if DNA zip codes control the clustering of INO1 with URA3:INO1. Two redundant gene recruitment sequences (GRS I and II) in the promoter of INO1 are responsible for targeting active INO1 to the nuclear periphery (Ahmed et al., 2010). Because the construct used to create URA3:INO1 possesses only the GRS I element (Figure 2A; Ahmed et al., 2010), we hypothesized that the GRS I element controlled clustering of INO1 and URA3:INO1. To test this idea, we compared the localization of grsImutINO1 with URA3:INO1. This mutation disrupts the GRS I element at the endogenous INO1 locus, but does not block targeting of INO1 to the nuclear periphery because GRS II is still functional (Ahmed et al., 2010). Mutation of the GRS I element in the INO1 promoter led to dramatic

Figure 4. Clustering of INO1 Requires Targeting to the Nuclear Pore Complex

(A) Distribution of distances between INO1 and URA3:INO1 in NUP2 (data from Figure 2C) and nup2Δ cells grown under activating conditions (–inositol).

(B) Clustering of active INO1 and URA3:INO1 in cells in which the nuclear envelope was also stained. Clustering of the two loci was determined in ~30 cells of each of the three classes: both genes on the nuclear envelope (on/on), one gene on the nuclear envelope (on/off), or both genes off the nuclear envelope (off/off). In all cases where the locus was scored as peripheral, the center of the green dot was <0.25 µm from the cytoplasmic edge of the nuclear envelope.

(C) Cells having the Tet repressor array at INO1 and the Lac repressor array at URA3:INO1 were arrested with hydroxyurea either after activating INO1 or before activating INO1.
increase in the mean distance between \textit{INO1} and \textit{URA3:INO1} (1.03 \pm .29 \mu m) and abolished clustering of \textit{INO1} with \textit{URA3:INO1} (Ahmed et al., 2010) in cells grown under activating conditions (\textit{\textminus}inositol). The strategy described in Figure 2A was used for (A) and (B).

(B) Distribution of distances between active \textit{INO1} and either \textit{URA3:INO1} (same as Figure 2C) or \textit{URA3:GRS I} in cells grown under activating conditions (\textit{\textminus}inositol).

(C) Distribution of distances between active \textit{HSP104} and either \textit{URA3} or \textit{URA3:HSP104prom} in cells grown in the presence of 10\% ethanol for 1 hr. The strategy described in Figure 3A was used for (C) and (D).

(D) Distribution of distances between \textit{HSP104} and \textit{URA3:GRS III} in cells grown in the presence or absence of 10\% ethanol. See also Figures S2 and S3.

Figure 5. Gene Clustering Is Controlled by DNA Zip Codes

(A) Distribution of distances between \textit{URA3:INO1} and either wild type \textit{INO1} (same data as Figure 2C) or \textit{grsImut INO1} (Ahmed et al., 2010) in cells grown under activating conditions (\textit{\textminus}inositol). The strategy described in Figure 2A was used for (A) and (B).

(B) Distribution of distances between active \textit{INO1} and either \textit{URA3:INO1} (same as Figure 2C) or \textit{URA3:GRS I} in cells grown under activating conditions (\textit{\textminus}inositol).

(C) Distribution of distances between active \textit{HSP104} and either \textit{URA3} or \textit{URA3:HSP104prom} in cells grown in the presence of 10\% ethanol for 1 hr. The strategy described in Figure 3A was used for (C) and (D).

(D) Distribution of distances between \textit{HSP104} and \textit{URA3:GRS III} in cells grown in the presence or absence of 10\% ethanol.

We mapped the DNA zip code activity in the \textit{HSP104} promoter to a 30 base pair fragment (GRS III) within the \textit{HSP104} promoter.
that was sufficient to target URA3 to the nuclear periphery (Figure S3A). Targeting by GRS III was constitutive and independent of activation of HSP104 (Figure S3A), suggesting that, like the GRS elements in the INO1 promoter, the zip code activity of GRS III is negatively regulated in the context of the promoter (Ahmed et al., 2010). Under inducing conditions, HSP104 clustered with URA3:GRS III (Figure 5D; mean = 0.53 ± 0.29 μm; 60% clustering), but not under noninducing conditions (Figure 5D; mean = 0.72 ± 0.25 μm; 25% clustering; p < 0.0001). Therefore, two different DNA zip codes are sufficient to mediate clustering of URA3 with two different endogenous genes.

**Endogenous GRS I Genes Cluster**

Because the GRS I element was necessary and sufficient to induce clustering with active INO1, we tested if other genes with the GRS I element would also cluster with INO1. TSA2 on chromosome IV has a GRS I element in its promoter that is required for targeting TSA2 to the nuclear periphery (Ahmed et al., 2010). Therefore, we asked if TSA2 and INO1 cluster at the nuclear periphery when active. When INO1 was active and TSA2 was not (Figure 6A), the mean distance (0.83 ± 0.41 μm) and the clustering (25%) between the genes was similar to the mean distance and clustering between repressed INO1 and URA3:INO1 (mean = 0.85 ± 0.38 μm; 20% clustering). However, in cells in which both TSA2 and INO1 were active (Figure 6A), we observed a significant decrease in the mean distance between the genes (mean = 0.58 ± 0.38 μm; p < 0.0001) and a significant increase in the fraction of the population in which they are clustered (54%; p < 0.0001). Therefore, two endogenous GRS I-targeted genes on different chromosomes cluster together in the nucleus.

To confirm that clustering of active INO1 with active TSA2 is due to GRS I-mediated targeting to the nuclear periphery, we compared the localization of active grsImutINO1 with TSA2. The distribution of distances between active grsImutINO1 and TSA2 revealed that the mean distance (0.91 ± 0.42 μm) and the clustering (25%) were very similar to the mean distance and clustering of active INO1 with uninduced TSA2 (Figure 6B). Thus, clustering of INO1 and TSA2 requires the GRS I zip code.

The Put3 Transcription Factor Recognizes the GRS I Zip Code to Mediate Gene Targeting and Clustering

Given the importance of the GRS I zip code in controlling gene targeting to the nuclear periphery and clustering, we sought to identify protein(s) that recognize GRS I to mediate targeting and clustering. We observed two activities from yeast lysates in electrophoretic mobility shift assays that bound to a 4× GRS I probe (band A and band B; Figure 7A). These activities were sensitive to heat, SDS, and proteinase digestion, suggesting that they represent proteins (Figure S4A). Competition with unlabelled wild-type and mutant 1× GRS I demonstrated that binding of band A was specific (Figure S4B). Also, whereas band A was able to bind both the multimerized 4× GRS I and a single copy 1× GRS I, band B was able to bind to multimerized GRS I probe only, suggesting that band B was an in vitro artifact (Figure S4C). Therefore, we identified the protein responsible for band A by enriching for this activity using DNA affinity chromatography, followed by mass spectrometry (see Experimental Procedures). This experiment identified 50 candidate proteins that were enriched in the eluate from 4× GRS I beads relative to control beads (Table S2). Lysates from strains either expressing tagged versions of these proteins or lacking these proteins were tested using the electrophoretic mobility shift assay (EMSA) (see Experimental Procedures; data not shown). Only one of the 50 proteins affected band A activity; lysates from strains lacking the transcription factor Put3 exhibited band B activity, but not band A activity (Figure 7A, lanes 3 and 4). This defect was complemented by expression of GST-Put3 (Figure 7A, lane 5). In the put3Δ strain expressing GST-Put3, band A was super-shifted into the well in the presence of anti-GST antibody (Figure 7A, lane 6), suggesting that GST-Put3 is in complex with GRS I.

To test if Put3 plays a role in GRS I-dependent targeting to the nuclear periphery, we localized INO1, URA3:INO1, and TSA2 in strains lacking Put3 (Figures 7B and 7C). In strains lacking Put3, targeting of URA3:INO1 and TSA2 to the nuclear periphery was lost, but targeting of INO1 (which possesses GRS II) to the nuclear periphery was maintained (Figures 7B and 7C). Likewise, Put3 was also required for targeting of URA3:GRS I to the nuclear periphery (Figure S4E). Loss of Put3 also resulted in a defect in the expression URA3:INO1 very similar to the effect...
of mutation of the GRS I element, but did not affect expression of INO1 (Figure 7F). Therefore, Put3 is required for GRS I-, but not GRS II-mediated targeting to the nuclear periphery and transcription. Consistent with this conclusion, GRS II did not compete with GRS I for binding to Put3 in EMSA experiments (Figure S4D).

Put3 is a Zn$_2$Cys$_6$ zinc finger transcription factor that regulates expression of genes involved in proline metabolism.

**Figure 7. The Put3 Transcription Factor Mediates GRS I-Dependent Clustering**

(A) An electrophoretic mobility shift assay of yeast lysates incubated with radiolabeled 4× GRS probe. Lysates were prepared from either a wild-type strain (lanes 1 and 2) or a put3Δ mutant strain (lanes 3–6). The put3Δ strain was transformed with a plasmid expressing GST-Put3 under the control of the ADH1 promoter (lanes 5 and 6). Anti-GST antibody was added to reactions in lanes 2, 4, and 6.

(B) Localization of repressed and active INO1 and URA3:INO1 in PUT3 and put3Δ mutants with respect to the nuclear envelope. The dynamic range of this assay is from 20%–85% and the blue, hatched line represents the distribution of the URA3 gene with respect to the nuclear envelope (Brickner et al., 2010; Brickner and Walter, 2004).

(C) Localization of TSA2 in PUT3 and put3Δ cells before or after heat shock (30 min).

(D) ChIP against GST or GST-Put3 from cells grown in the presence or absence of inositol.

(E) ChIP against Nup2-TAP from PUT3 and put3Δ strains grown in the presence or absence of inositol. For (C)–(E), the immunoprecipitated DNA was quantified relative to input by real-time quantitative PCR.

(F) mRNA levels for INO1, URA3:INO1, or URA3:grsImut INO1 in PUT3 or put3Δ strains, quantified by RT-qPCR relative to ACT1 mRNA. For (B)–(F), error bars represent SEM.

(G) Left: distribution of distances between Tet repressor-marked INO1 and Lac repressor-marked URA3:INO1 in wild-type (same data as in Figure 2A) and put3Δ cells grown under activating conditions (-inositol). Right: fraction of cells in which the loci were <0.55 μm apart. See also Figures S4, S5, and S6 and Table S2.
Put3 binds to the UASput element in the promoters of PUT1 and PUT2 (Siddiqui and Brandriss, 1989). This element (CGG-N10-GCC) is not obviously related to the GRS I element defined by zip code activity at URAG3 (GGGTTGGA; Ahmed et al., 2010). To confirm that Put3 interacts with the GRS I element in vivo, we performed chromatin immunoprecipitation (ChIP) using GST-Put3. Whereas GST-Put3 interacted constitutively with the PUT2 promoter by ChIP (Figure 7D), it interacted with the INO1 and TSA2 promoters only under activating conditions (Figures 7D and S4F). GST-Put3 did not interact with RPA34, an intergenic locus ~4.5 kb upstream of the INO1 promoter (Figure 7D). Thus, Put3 specifically binds to active GRS I-containing promoters in vivo in a manner that correlates with targeting to the nuclear periphery.

Targeting to the nuclear periphery involves a physical interaction of genes with the NPC. The GRS I is sufficient to induce a ChIP interaction with the NPC (Ahmed et al., 2010). To ask if interaction of the NPC with the GRS I requires Put3, we performed ChIP with Nup2-TAP in the put3Δ mutant strain and quantified the recovery using primers flanking the GRS I element. Interaction of Nup2-TAP with the GRS I element in the INO1 promoter was lost in the put3Δ strain (Figure 7G). Therefore, Put3 is necessary for GRS I-mediated targeting to the nuclear periphery and interaction with the nuclear pore complex. Finally, loss of Put3 led to loss of clustering between INO1 and URA3:INO1 (Figure 7G; p < 0.0001 for both mean distance and clustering compared with the PUT3 strain). Therefore, Put3 physically binds to the GRS I in vivo and in vitro and is required for GRS I-mediated targeting and clustering at the nuclear periphery.

**DISCUSSION**

Here we demonstrate that genes that localize at the nuclear periphery through interaction with the NPC can cluster together with genes on other chromosomes. This property is controlled by DNA zip codes in their promoters that can be transplanted to an ectopic site on another chromosome. Clustering mediated by DNA zip codes requires trans-acting factors and interaction with the nuclear pore. This suggests that DNA zip code–encoded positioning of individual genes impacts the proximity of loci on different chromosomes within the nucleus and the folding of the genome as a whole.

Interchromosomal clustering of loci is a common theme and has been observed in many cell types and under many conditions. Telomeres cluster in most cells and this has been proposed to favor repression of subtelomeric genes (Bass et al., 1997; Cooper et al., 1998; Dernburg et al., 1996; Gotta and Gasser, 1996; Lanzuolo et al., 2007; Scherthan et al., 1996; Tolhuis et al., 2011). Likewise, Polycomb-repressed genes in Drosophila cluster at Polycomb bodies (Lanzuolo et al., 2007; Tolhuis et al., 2011). In budding yeast, many of the 274 tRNA genes cluster into distinct portions of the nuclear envelope. This suggests that Put3 mediates the GRS I–dependent effects on localization and clustering. Put3 is required for GRS I zip code activity both in the context of promoters where gene localization is regulated and when inserted at URA3, where it is not. This suggests that Put3 mediates the GRS I–dependent effects on localization and clustering.

A role for Put3 in controlling gene localization and clustering is unanticipated. Put3 is a nucleoplasmic protein with no obvious bias in its localization (unpublished data; Huh et al., 2003). Furthermore, the sequence to which Put3 binds in the context of the PUT1 and PUT2 promoters (the UAS PUT; CGG-N10-GCC) does not resemble the sequence of the core GRS I (GGGTTGGA), deduced by mapping zip code activity by insertion at URA3 (Ahmed et al., 2010). However, the binding specificity of Put3 is still incompletely understood. A genome-wide ChIP-chip study identified the Put3 motif as CGGAAGCC (Macsac et al., 2006). Two different studies using unbiased biochemical approaches found that the Put3 DNA binding domain interacts with the sequence TCCCGGG (Badis et al., 2008; Zhu et al., 2009). From a library of 32,897 8mers, the sequence CGGGGTTA, genome-wide chromosome conformation capture studies suggest that many genes of related function cluster in the nucleus (Tanizawa et al., 2010). Although clustering of these loci in some cases requires DNA binding proteins (Laroche et al., 1998; Schoenfelder et al., 2010) and is correlated with shared DNA elements (Gotta et al., 1996; Moinar and Kleckner, 2008; Schoenfelder et al., 2010; Tanizawa et al., 2010) or sequence homology (Moinar and Kleckner, 2008), it is unclear if these factors mediate clustering or regulate clustering. Here we find that small DNA zip codes in the promoters of genes that are targeted to the nuclear periphery are both necessary and sufficient to confer interchromosomal gene clustering.

Our observations support the notion that genomes code for positioning of genes to restricted territories within the nucleus (Berger et al., 2008). We have found that two different DNA zip codes from the INO1 and HSP104 promoters, when integrated adjacent to the URA3 gene, are sufficient to cause URA3 to cluster with the genes from which they came. Because INO1 and HSP104 do not localize to the same portion of the nuclear envelope (Figure 3), this suggests: 1) that URA3 is not highly constrained, 2) that these DNA zip codes are able to override any local positioning information to induce clustering with either INO1 or HSP104, and 3) that INO1 and HSP104 may be targeted to distinct portions of the nuclear envelope.

The transcription factor Put3 binds to the GRS I zip code and is necessary for GRS I–mediated targeting to the nuclear periphery and clustering. This suggests that transcription factors can affect both transcription and gene localization. Although physical interaction with the NPC has been correlated with transcription factor binding sites (Casolari et al., 2004; Schmid et al., 2006) and clustering of coregulated genes at transcription factories requires the transcription factor Klf1 (Schoenfelder et al., 2010), it is still unclear if these transcription factors regulate gene localization or mediate gene localization. Put3 is required for GRS I zip code activity both in the context of promoters where gene localization is regulated and when inserted at URA3, where it is not. This suggests that Put3 mediates the GRS I–dependent effects on localization and clustering.

...
of stress-inducible genes with the NPC. Consistent with this hypothesis, the \( \text{PUT1} \) or \( \text{PUT2} \) promoters do not interact with the NPC by ChIP (Casolari et al., 2004), despite the fact that Put3 is bound to these promoters constitutively (Figure 7; Axelrod et al., 1991). Furthermore, the \( \text{UAS}_{\text{PUT}} \), when inserted at \( \text{URA3} \), does not function as a DNA zip code (S.A. and J.H.B., unpublished data). Finally, a fragment of Put3 that includes the amino terminal DNA binding domain and dimerization domain, but lacks the carboxy terminal 853 amino acids, supports GRS I-mediated targeting to the nuclear periphery (S.A. and J.H.B., unpublished data). This fragment lacks the activation domain and does not support \( \text{PUT} \) gene expression (des Etages et al., 1996). Together, these observations suggest that the Put3 DNA binding domain binds to these two sequences in two distinct conformations, allowing it to interact with distinct “effectors.”

Our results raise the possibility that genes are targeted to restricted portions of the nuclear periphery. \( \text{INO1} \) localizes to a restricted band at the nuclear envelope (Figure 1). The zip code responsible for targeting \( \text{INO1} \) to the nuclear periphery is also sufficient to induce interchromosomal clustering of an ectopic locus with the endogenous \( \text{INO1} \) gene (Figure 5). This suggests that \( \text{INO1} \) and other GRS I genes are targeted to the same portion of the nuclear envelope. We do not propose that genes that share zip codes are targeted to the same NPC; our data are consistent with genes being targeted to a portion of the nuclear envelope that would include a number of NPCs. How might this work? It is possible that zip code adaptor proteins such as Put3 interact with proteins that are stably and heterogeneously distributed at the nuclear envelope. However, there are very few yeast proteins that are heterogeneously distributed on the nuclear envelope and these are generally associated with telomeres (Gotta et al., 1996; Huh et al., 2003). Therefore, it is possible that the active forms of these hypothetical proteins, perhaps controlled by posttranslational modifications, are heterogeneously distributed. Likewise, it is conceivable that the biochemical structure or arrangement of the subunits of NPC might be subtly different along different parts of the nuclear envelope, which would not be obvious from steady state concentrations of individual subunits. If so, it will be important to explain how the heterogeneous distribution of these activities is established or maintained.

Alternatively, perhaps gene targeting to the nuclear periphery is not precise. DNA zip codes could mediate both an interaction with the NPC and homotypic clustering between genes. Chromosomes are constrained by their size, folding and stable association with subnuclear structures. The subnuclear positioning of genes is a product of their position along these polymers. The NPC might serve as a stable surface on which homotypic interactions between genes could occur more efficiently. If so, then the apparently restricted localization of genes at the nuclear periphery might reflect the constraints associated with their position along the chromosome, rather than zip code-mediated positioning. Consistent with the possibility that clustering and interaction with the NPC can be uncoupled, previous targeting to the NPC is sufficient to maintain clustering during S-phase, a period when peripheral localization is lost (Figures 4B and 4C; Brickner and Brickner, 2010).

The clustering we have observed is specific. We have observed clustering between genes that have identical zip codes and not between genes with different zip codes (Figure 3). It remains to be seen if there are zip codes with overlapping distributions. But the available data suggest that, although many genes interact with the NPC, they may be targeted to distinct populations of NPCs by distinct mechanisms. Furthermore, because endogenous genes on different chromosomes that share zip codes cluster together (e.g., \( \text{INO1} \) with \( \text{INO1} \) and \( \text{INO1} \) with \( \text{TSA2} \)), it raises the fascinating possibility that genes that share a zip code also share factors important for their transcription or mRNA export. Clustering of genes with related functions might allow better coordination of their expression. If so, then clustering of genes might serve to couple spatial compartmentalization of the nucleus with functional compartmentalization of the genome.

**EXPERIMENTAL PROCEDURES**

**Strains and Plasmids**

Yeast strains and cloning are described in Supplemental Experimental Procedures.

**Statistical Mapping**

Statistical mapping was carried out as described (Berger et al., 2008), with modifications detailed in Supplemental Experimental Procedures.

**RFP-LacI + GFPTetR Two-Dot Assay**

Plasmid p15816-INO1 was digested with MscI and integrated downstream of \( \text{INO1} \) in strains transformed with GFP-Tet repressor (integrated at \( \text{LEU2} \); Grund et al., 2008). These strains were then transformed with plasmid pME08 expressing RFP-LacI under the control of the ADH2 promoter (Jiang et al., 2009) and LacO array plasmids integrated either at \( \text{TSA2} \) (pLacO-TSA2) or at \( \text{URA3} \) as follows: pLacO128 (URA3), pLacO128-INO1 (URA3/INO1), and pLacODGRS I 141-75 (URA3/GRS I; Ahmed et al., 2010; Brickner and Walter, 2004). Except for the experiments in Figures 2E and 2F, all experiments were performed with haploid cells. Cells were grown overnight in SDC-Trp (+//−inositol) diluted into fresh media containing 2% ethanol as a carbon source to derepress expression of RFP-LacI. Cells were fixed two times in 4% formaldehyde for 30 min and prepared for immunofluorescence using rabbit polyclonal anti-RFP and mouse monoclonal anti-GFP antibodies as described (Brickner et al., 2010). The distances between the centers of the dots was measured for ≥ 100 cells using Zeiss LSM software.

**GFP-LacI Large and Small Dots Assay**

Plasmid pSF3013 having 256 Lac repressor binding sites was integrated at \( \text{HSP104} \) as described (Diepo bypass et al., 2006) in a strain transformed with GFP-LacI plasmid pAFS144 (Robinett et al., 1996). This strain was then transformed with one of the following plasmids having 128 lac repressor binding sites: pLacO-INO1, pLacO-GAL1, or pLacO-GAL2, integrated at \( \text{INO1} \), \( \text{GAL1} \), or \( \text{GAL2} \), respectively as described (Brickner et al., 2007; Brickner and Walter, 2004; Diepo bypass et al., 2006). Cells were fixed and processed for immunofluorescence against GFP as described (Brickner et al., 2010) and the distance between the two dots was measured using Zeiss LSM software.

**EMSA**

Cells were grown and permeabilized as described (Brickner et al., 2001). After permeabilization of the cells, the supernatant was collected and 2 µl was added to 20 µl gel-shift reaction mix (0.5 mM DTT, 20% glycerol, 100 mM KCl, 20 mM HEPES-KOH, pH 6.8, 0.2 mM EDTA, 50 µg/ml Poly d(dC), 0.5 mM [\(^{32}\text{P}\)]-labeled 4× GRS I probe, 0.04% bromophenol blue), incubated 15 min and separated on 6% native polyacrylamide gels in 0.5× TBE [Tris/boric acid/EDTA]. Gels were dried on Whatman filter paper, exposed to PhosphorImager screen overnight, and imaged on a Typhoon PhosphorImager. Sequences of the DNAs used: 4× GRS I: TTAAG-[TCCGGGTGGATG]2-AGGT; 1× GRS I: GTTGGCGGTGATGCDCGC; 1× mutGRS I: GTTGCACAAAAACCATGCDCGC.
DNA Zip Codes Control Interchromosomal Clustering

On-Line Capillary LC-MS and LC-MS-MS Analysis

GRS I-binding proteins were enriched as described in Supplemental Experimental Procedures. LC-MS and LC-MS-MS were performed as described previously (Chu et al., 2006) and as detailed in Supplemental Experimental Procedures.

LCMSMS RAW data files were processed using PAVA (Guan et al., 2011). The centroided peak lists of the CID spectra were searched against a database that consisted of the Swiss-Prot protein database, to which a randomized version had been concatenated (Elias and Gygi, 2007), using Batch-Tag, a program in the University of California San Francisco Protein Prospector version 5.9.2. A precursor mass tolerance of 15 ppm and a fragment mass tolerance of 0.5 Da were used for protein database search. Protein hits are reported with a Protein Prospector protein score ≥22, protein discriminant score ≥0.0, and a peptide expectation value ≤0.01 (Chakley et al., 2005). This threshold of protein identification parameters did not return any substantial false positive protein hits from the randomized half of the concatenated database.

Chromatin Immunoprecipitation

ChIP of Nup2-TAP was carried out as described (Ahmed et al., 2010; Light et al., 2009) and as detailed in Supplemental Experimental Procedures.

RT-qPCR

INO1 and ACT1 mRNA levels were quantified by RT-qPCR as described (Brickner et al., 2007).

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.devcel.2012.03.012.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM080484 and a W.M. Keck Young Scholars in Biomedical Research Award (J.H.B.). S.A. and W.L. were each supported by a Rappaport Award for Research Excellence; A.T. was supported by a Northwestern Undergraduate Research Fellowship; M.Y. was supported by NIH training grant T32 GM008061; F.C. is supported by the New Hampshire Agricultural Experiment Station and NIFA/USDA H00567; T.L.H. was supported by a UNH Summer Undergraduate Research Fellowship; E.F. is supported by CNRS, Institut Pasteur, and ANR-PIRIBIO grant ANR-09-PIR-0024. The authors thank Dr. Françoise Stutz (University of Geneva), Dr. James Hopper (The Ohio State University), and Dr. Ghislain Cabal (Instituto de Medicina Molecular, Lisboa, Portugal) for sharing plasmids, Daryl Stave-ness for technical assistance, and Dr. Eric Weiss for help with the EMSA experiments.

Received: November 22, 2011
Revised: February 17, 2012
Accepted: March 27, 2012
Published online: May 10, 2012

REFERENCES


Berger, A.B., Cabal, G.G., Fabre, E., Duong, T., Buc, H., Nehruss, U., Olivo-


