DNA Research, 2016, 23(2), 155–169
doi: 10.1093/dnares/dsw005
Advance Access Publication Date: 29 February 2016
Full Paper

OXFORD

Full Paper

Spatially coordinated replication and minimization of expression noise constrain three-dimensional organization of yeast genome

Arashdeep Singh, Meenakshi Bagadia, and Kuljeet Singh Sandhu*

Department of Biological Sciences, Indian Institute of Science Education and Research (IISER)-Mohali, SAS Nagar 140306, India

*To whom correspondence should be addressed. Tel. +91 172-2293182. Fax. +91 172-2240266. E-mail: sandhuks@iisermohali.ac.in

Edited by Prof. Takashi Ito

Received 4 November 2015; Accepted 31 January 2016

Abstract

Despite recent advances, the underlying functional constraints that shape the three-dimensional organization of eukaryotic genome are not entirely clear. Through comprehensive multivariate analyses of genome-wide datasets, we show that *cis* and *trans* interactions in yeast genome have significantly distinct functional associations. In particular, (i) the *trans* interactions are constrained by coordinated replication and co-varying mutation rates of early replicating domains through interactions among early origins, while *cis* interactions are constrained by coordination of late replication through interactions among late origins; (ii) *cis* and *trans* interactions exhibit differential preference for nucleosome occupancy; (iii) *cis* interactions are also constrained by the essentiality and co-fitness of interacting genes. Essential gene clusters associate with high average interaction frequency, relatively short-range interactions of low variance, and exhibit less fluctuations in chromatin conformation, marking a physically restrained state of engaged loci that, we suggest, is important to mitigate the epigenetic errors by restricting the spatial mobility of loci. Indeed, the genes with lower expression noise associate with relatively short-range interactions of lower variance and exhibit relatively higher average interaction frequency, a property that is conserved across *Escherichia coli*, yeast, and mESCs. Altogether, our observations highlight the coordination of replication and the minimization of expression noise, not necessarily co-expression of genes, as potent evolutionary constraints shaping the spatial organization of yeast genome.

Key words: 3D genome organization, long-range chromatin interactions, replication, expression noise, evolutionary constraints, essential genes

1. Introduction

Eukaryotic genes and their regulatory elements communicate with each other through a complex wiring of long-range interactions.[1] It is now well established that distal enhancers can physically juxtapose to their cognate promoters for transcriptional regulation.[2–8] Interestingly, distant genes can also co-localize in nuclear space.[9,10] The prevailing view is that the genes spatially cluster at concentrated foci of RNA polymerase II, also known as transcription factories.[11–16] It is suggested that the spatial convergence of genes at transcription factories provide a topological basis of co-expression of engaged genes; however, such proposals have not been subjected to proper scrutiny. Recent advent of high throughput derivatives of Chromosome Conformation Capture (3C) has availed genome-wide quantitative data of long-range chromatin interactions across diverse spectra of model systems.[17–26] Briefly, in 3C-derived techniques, the chromatin is cross-linked with formaldehyde, restriction

© The Author 2016. Published by Oxford University Press on behalf of Kazusa DNA Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

digested, and the open ends of cross-linked products are ligated in diluted conditions to prefer intra-molecular ligation over inter-molecular. In HiC, the ligated junctions are then pulled down and sequenced using deep sequencing to unravel all-to-all chromatin interactions.[21] HiC has revealed large topologically associated domains (TADs) that exhibit high density of intra-connectivity of chromatin and are largely conserved across cell lineages.[27,28] TADs are tightly associated with the chromatin type and replication timing, and are marked by CTCF on boundaries.[27,29,30] Widespread enhancer-to-promoter interactions, that are mostly cell-type specific, have been uncovered across several systems.[28,31–33] Zhang et al.[31] have suggested differential usage of enhancers during embryonic stem cell differentiation. Some studies have also revealed promoter-to-terminator interactions commonly found for housekeeping genes,[34] possibly ascribing a circular template for recurrent transcription. Most interesting of all is the widespread promoter-to-promoter interactions among genes impinging from neighbouring regions to form discrete multi-gene complexes.[34,35] However, what functional and evolutionary constraints might have shaped the large-scale organization of promoter–promoter interactions is not entirely clear. Although the genes within multi-gene complexes are shown to be co-expressed,[34,36] whether or not co-expression of engaged genes is dependent on their spatial, but not the linear, proximity remains to be seen. Moreover, it is hypothesized that interacting promoters can influence transcriptional states of each other and that the promoter of one gene can function as an enhancer of other gene.[34] Nevertheless, these proposals are yet not established as fact. Importantly, most of these studies have primarily focussed on intra-chromosomal (referred as '*cis*' in this study) interactions and whether or not distant genes converging from different chromosomes (referred as *trans* interactions) have functional association is yet not clear. Comprehensive statistical analyses of accumulated HiC like datasets can answer several questions pertaining to non-random genome organization. Here, we ask whether we can delineate evolutionary constraints of three-dimensional organization of genome.

Multivariate analyses provide a statistical platform to assess the association of several different functional variables in an unbiased manner. Availability of various genome-wide datasets and high-resolution data of *cis* as well as *trans* chromatin interactions makes budding yeast an ideal candidate for multivariate analysis to identify the potential functional constraints shaping the non-random spatial organization of genome. The article by Duan et al.[22] suggested following key features of three-dimensional organization of budding yeast genome: (i) interactions among the centromeres, (ii) interactions among the sites of early origin and not the late origins, and (iii) interactions among t-RNA genes. A few follow-up studies suggested a link between chromatin interactions and co-expression of involved genes.[37,38] Another report, on the contrary, dismissed the claims of proximity of co-expressed genes in yeast.[39] Moreover, the possibility that *cis* and *trans* chromatin interactions might have been shaped under different evolutionary constraints has not been explored. In this study, using comprehensive statistical analysis, we show that functional and evolutionary constraints of *cis* and *trans* interactions are significantly distinct and are not necessarily associated with co-expression of genes. We show that the *trans* interactions are primarily constrained by coordinated replication through converged early origins, while *cis* interactions are shaped by coordination through late origins and by the minimization of expression noise of engaged genes in an evolutionarily conserved manner.

2. Materials and methods

2.1. Data sources

We obtained the publicly available genome-scale datasets from different sources; details of which are given in Supplementary Table S1.

2.2. Methods

Detailed methodology to process the datasets is given in the Supplementary Material.

2.2.1. Binning of data

The interaction frequency data (frq) were clustered into bins of equal size of 1 unit, and average value of each functional attribute was calculated for each bin. The master tables for the binned and the original data are given in the Supplementary Information.

2.2.2. Correlogram analyses

Correlograms were plotted for the binned data using 'corrgram' R-package (http://cran.r-project.org/web/packages/corrgram/index.html). Pearson's correlation coefficients were calculated using *cor.test()* function in R and *P*-values for multiple comparisons (no. *55*) were corrected using Bonferroni's method. Significant *P*-values after correction are marked with triple asterisk (***) in correlograms.

2.2.3. Partial least squared regression

Partial least squared regression (PLSR) models the relation between input matrix $\mathbf{X}$ ($n \times p$ matrix with n dimensions of p input variables, genomic/functional attributes in this case) and response matrix $\mathbf{Y}$ ($n \times 1$ matrix with n dimensions of 1 response variable, interaction frequency in this case) by decomposing them as following:

$$\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^{\mathrm{T}} + \mathbf{F} \quad (2)$$

Where $\mathbf{T}$ and $\mathbf{U}$ are $n \times r$ matrices with r extracted latent vectors (or scores). $\mathbf{P}$ and $\mathbf{Q}$ are $p \times r$ and $1 \times r$ matrices of X and Y loadings, respectively. $\mathbf{E}$ and $\mathbf{F}$ are $n \times p$ and $n \times 1$ matrices of residuals. In kernel PLS regression, following inner relation between T and U is assumed;

$$\mathbf{U} = \mathbf{TB} + \mathbf{H}$$

where $\mathbf{B}$ is the $r \times r$ diagonal matrix of regression coefficients and $\mathbf{H}$ is matrix of residuals. Accordingly, equation (2) can be rewritten as

$$\mathbf{Y} = \mathbf{TBQ}^{\mathrm{T}} + (\mathbf{HQ}^{\mathrm{T}} + \mathbf{F})$$

which defines the final PLS regression model,

$$\mathbf{Y} = \mathbf{TC}^{\mathrm{T}} + \mathbf{F}^{*}$$

where $\mathbf{C}^{\mathrm{T}} = \mathbf{BQ}^{\mathrm{T}}$ and $\mathbf{F}^{*} = \mathbf{HQ}^{\mathrm{T}} + \mathbf{F}$

'PLS' R-package was used for this (http://cran.r-project.org/web/packages/pls/index.html). Prior to multivariate regression analyses, the columns of X were *Z*-normalized.

2.2.4. Random re-sampling to compile null distributions

To assess the non-random connectivity among replication origins, we used re-sampling approach given by Witten and Noble.[39] Briefly, the positions of the origins in the genome were randomized 10^3 times keeping the chromosomal distribution same as in the original set.

P-value was calculated using following equation:

$$p = \frac{1}{B}\left(1 + \sum_{b=1}^{B} (1)\frac{k_b}{k'_b} > \frac{k}{k'}\right)$$

Where, B = number of re-samplings (no. 1,000); k_b = Observed number of interactions among randomized coordinates; k'_b = Expected number of interactions among the randomized coordinates; k = Observed number of interactions among original coordinates. k' = Expected number of interactions among original coordinates.

We mapped the feature coordinates onto restriction fragments and considered a restriction fragment only once regardless of how many feature coordinates fall inside. This takes care of local clustering of feature coordinates.

3. Results

3.1. Distinct functional attributes of *cis* and *trans* chromatin interactions in yeast

Chromatin in the interphase nucleus is generally present in the form of chromosomal territories. The chromatin loci embedded inside the territory would, therefore, be expected to have greater number of *cis* interactions, while the ones near territorial edge would have abundant *trans* interactions. This is particularly true for the metazoan genomes; the distinction into such intra- and inter-territorial organization of genes is not well studied in yeast however. If yeast genome conforms into distinct territories, certain regions of chromosomes would exhibit greater number of either *cis* or *trans* interactions marking intra-territorial and inter-territorial locations of the loci. Therefore, we first tested whether the number of *cis* interactions of each locus correlated with the *trans* interactions of corresponding locus in the yeast genome. We smoothened the number of *cis* and *trans* interactions using an arbitrary window of 20 kb along the chromosomes and plotted number of *cis* vs. *trans* interactions. We observed that the number of *cis* and *trans* interactions per locus poorly correlated with each other ($\rho = 0.05$) and that there were regions, which were enriched either with *cis* or *trans* interactions (highlighted in ellipses in Fig. 1a). The observed distinction of *cis* and *trans* interactions was also confirmed by plotting the *Z*-normalized number of *cis* and *trans* interactions side-by-side as a function of chromosomal coordinates. As shown in the Fig. 1b, there were domains in the genome having greater number of either *cis* or *trans* chromatin interactions. We further calculated average frequency of all the *cis* and all the *trans* interactions of each genomic locus independently. By plotting average *cis* and *trans* interaction frequencies of corresponding genomic loci against each other, we once again observed regions having higher *cis* interaction frequency and relatively lower *trans* frequency and vice versa supporting that there were regions in the genome having high frequency of either *cis* or *trans* interactions (Supplementary Fig. S1). We further showed an example of chromosome 4 in the rendered 3D model of yeast genome (as provided by Duan et al.) in Fig. 1c, highlighting a region enriched with local clusters of abundant *cis* interactions (dashed box). Rest of the chromosome 4 could be seen as an extended arm extensively intermingling with other chromosomes. Quantitative data for Fig. 1c were shown in the Fig. 1d. The plot clearly showed greater number of *cis* interactions in the boxed area, which corresponds to the same boxed area as in 3D model. Our observations through Fig. 1 highlighted that certain genomic domains can have preferred folding in *cis* while others intermingles extensively with other chromosomes in *trans*, supporting the notion of 'chromosome territories'. This also raised a possibility of *cis* and *trans* interactomes to have evolved

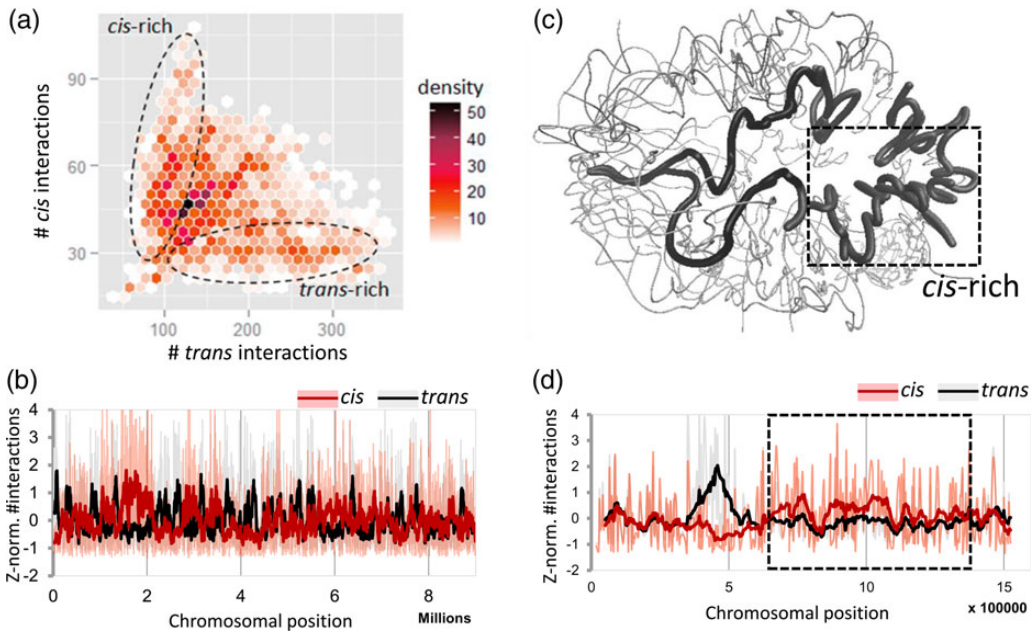

Figure 1. Preferred domains of *cis* and *trans* chromatin interactions. (a) Scatter plot of number of *cis* and *trans* interactions of each locus in yeast genome. X–Y coordinates were binned into 100 bins and represented as shade intensity. (b) Linear view of yeast genome (chr1–16 are concatenated in that order) depicting *Z*-normalized number of *cis* and *trans* chromatin interactions of each locus. Thicker lines represent moving average of 10 values. (c) 3D model of yeast genome, as provided by Duan et al., rendered using VMD software. As an example, a region rich in *cis* interactions on chromosome 4 is highlighted in dashed box. (d) *Z*-normalized number of *cis* and *trans* interactions of each locus on chromosome 4 as a function of chromosomal coordinates. Boxed area highlights the same *cis*-rich region as in (c). Data plotted in a, b, and d are given in the Supplementary Table S2. This figure is available in black and white in print and in colour at *DNA Research* online.

under distinct evolutionary constraints. We, therefore, attempted to delineate the functional attributes that explained best the observed distinction of *cis* and *trans* chromatin interactions by analysing the datasets of *cis* and *trans* chromatin interactions independently.

Multi-dimensional genomic datasets, i.e. data where measurements were done for all the yeast genes across several different time-points or conditions, were used to calculate similarity (mostly Pearson's correlation coefficient; Fig. 2; Table 1; Materials and Methods) between interacting genes. Analysis of correlations among several functional attributes suggested that the frequency of *trans* interactions was best correlated with the similarity in % replicated DNA or co-replication ('rep', Table 1) of interacting loci ($\rho = 0.47$, P-value < 10^{-4}, Fig. 3a). Interestingly, co-expression of genes (i) through cell cycle (cct), (ii) following transcription factor perturbation (tfp), and

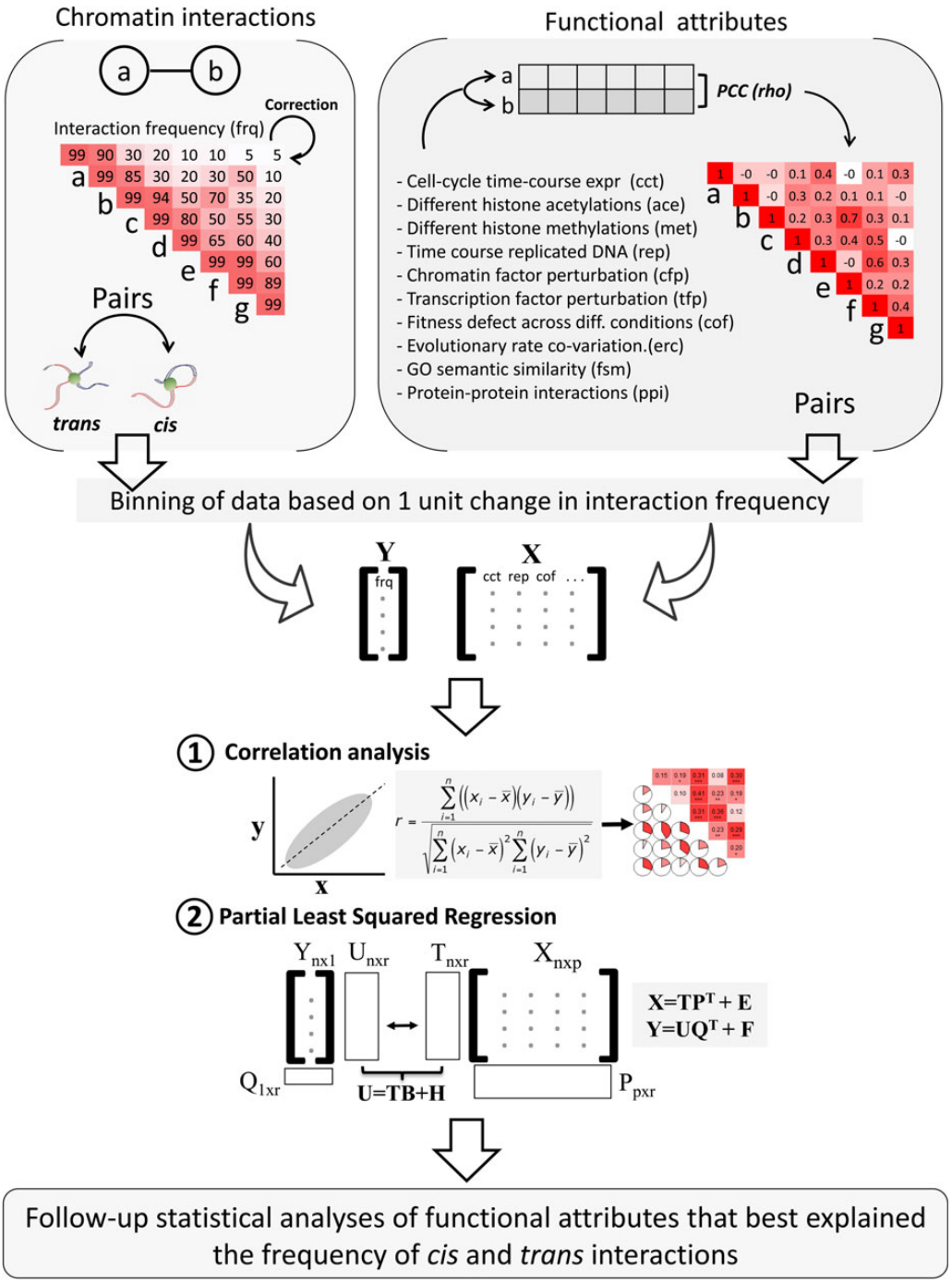

Figure 2. Flow chart of overall analysis. Pairs of interacting genes were obtained from chromatin interaction data and split into *cis* and *trans* pairs. For each interacting pair of genes, epigenetic/functional similarity was calculated, mostly by calculating Pearson's correlation between attribute profiles across conditions or time points. Vectors of *cis* and *trans* interaction frequencies were compiled and binned into equal size of 1 unit change in interaction frequency. The functional similarity scores were accordingly averaged for each bin. The resulting interaction frequency and functional similarity matrices (Y and X) were than subjected to correlation and PLSR analysis. The functional attributes that exhibited best association in the multivariate analyses were subjected to comprehensive follow-up analyses. This figure is available in black and white in print and in colour at *DNA Research* online.

Table 1. Abbreviations used for different kind of functional similarities between genes

Abbreviations	Attribute	Similarity	No. in Supplementary Table S1
frq	Interaction frequency	*cis* and *trans* chromatin interaction frequencies between genes were taken from Duan et al. and normalized using HiCNorm package	1
cct	Cell cycle time course	Pearson correlation between time course expression values of interacting genes	2
cfp	Chromatin factor perturbation	Pearson correlation between expression values of interacting genes across chromatin factor mutant strains	3
tfp	Transcription factor perturbation	Pearson correlation between expression values of interacting genes across transcription factor mutant strains	4
env	Environmental response	Pearson's correlation between expression values of interacting genes across different environmental conditions	5
ace	Acetylation	Pearson's correlation between ChIP enrichment values of interacting gene promoters across different histone acetylations	7
met	Methylation	Pearson's correlation between ChIP enrichment values of interacting gene promoters across different histone methylations	8
rep	Co-replication	Pearson's correlation between % replication DNA of 500 bp genomic bins, as provided by the authors, mapping to interacting pairs of genes	9
cof	Co-fitness	Co-fitness of interacting genes were directly taken from this study	11
ppi	Protein–protein interaction	Socio-Affinity (SA) index, which measures the log-odds of observed number of times two proteins interact relative to the expected value deduced from their frequency in the dataset, were obtained for each protein-protein interaction	14
fsm	Functional similarity	Functional similarity (fsm) was calculated using 'GOSemSim' R-package.	16
erc	Evolutionary rate co-variation	Evolutionary rate co-variation (erc) was taken as it is from the source	17

(iii) following environmental perturbations did not show strong correlation with the *trans* interaction frequency as claimed elsewhere[38] ($\rho = 0.07$, 0.09, and 0.13, respectively; Fig. 3a, Supplementary Fig. S2). On the contrary, *cis* interactions among loci, which were at least 20 kb apart, did not exhibit strong correlation with co-replication ($\rho = 0.14$; P-value $< 10^{-3}$; FDR > 0.01) and were, instead, strongly correlated with the co-fitness of interacting genes as measured through chemical genomic screens ($\rho = 0.25$; P-value $< 10^{-4}$; FDR < 0.01; Fig. 3b). Again, the general cell-cycle-related co-expression of genes showed weak correlation, though slightly higher compared with the *trans* interactions ($\rho = 0.11$ vs. 0.07). To further scrutinize the significance of observed correlations, we rewired the chromatin interactions using the strategy given in the 'Materials and methods' section to obtain null distribution and calculate P-values (Supplementary Fig. S3). Examples of correlograms generated from rewired *cis* and *trans* interactions were shown in Supplementary Fig. S3a and b, and their comparisons with observed values of correlations were drawn in Supplementary Fig. S3c and d. In particular, we confirmed that (i) the correlation of *trans* interaction frequency with the coordinated replication and that of *cis* interaction frequency with the co-fitness of interacting genes were (i) significant compared with rewired control (P-value = 1.0e–09 and 3.3e–09, respectively, FDR < 0.001 for both comparisons, Supplementary Fig. S3a–e); and (ii) significantly different for *cis* and *trans* interactions, i.e. correlation with 'rep' was significantly higher for *trans* (P-value = 3.1e–12) and lower for *cis*, while correlation with co-fitness was significantly higher for *cis* (P-value = 8.5e–08) and lower for *trans* interactions. We also confirmed these observations for the chromatin interactions that were concomitantly captured in the HiC-library generated using *Eco*RI restriction enzyme, highlighting the robustness of the analyses (Supplementary Fig. S4a and b). We further cross-validated the differential association of *cis* and *trans* interaction frequencies with the coordinated replication (rep) and co-fitness (cof) of engaged genes using recently published HiC dataset of exponentially growing budding yeast cells[40] (Fig. 3c).

Correlation analysis does not always imply if a variable independently accounts for the observed response (interaction frequency in this case) or it is due to co-linearity among variables. To address this, we performed PLSR of the variables to identify the ones that explained the observed variance in the frequency of chromatin interactions. The analyses suggested that 32.3 and 36.9% of total variance in *trans* and *cis* interaction data, respectively, were explained by one major component, primarily comprised of co-replication (rep) followed by similar susceptibility to chromatin factor perturbations (cfp) of interacting genes in case of *trans* interactions and co-fitness (cof) followed by functional similarity (fsm) of interacting genes in case of *cis* interactions (Fig. 3d and e). The second best component explained only 5.6 and 3.9% of variance in *trans* and *cis* interaction frequency, respectively. Details of relative contributions of functional attributes to each component were given in the Supplementary Fig. S5. Importantly, second component was also associated with co-replication to a large extent (second best contributor) in case of *trans* interactions and with co-fitness (second best contributor) in case of *cis* interactions (Supplementary Fig. S5). Interestingly, the cell-cycle-related co-expression (cct) of genes, in general, did not show significant contributions to the first two components in PLSR analysis of either *trans* or *cis* interactions. Overall, the PLSR and the correlation analyses concomitantly suggested that the *trans* interactions in yeast were primarily shaped by similarity in replication states of interacting loci, while *cis* interactions were mainly constrained by co-fitness of interacting genes highlighting an overall distinction of constraints shaping *cis* and *trans* chromatin interactomes.

3.2. *trans* and *cis* interactions associate with the coordination of early and late replication, respectively

It has been shown that the early firing, but not the late firing, origins from different chromosomes non-randomly collide with each other in the nuclear space.[22] We first confirmed this observation (P-value =

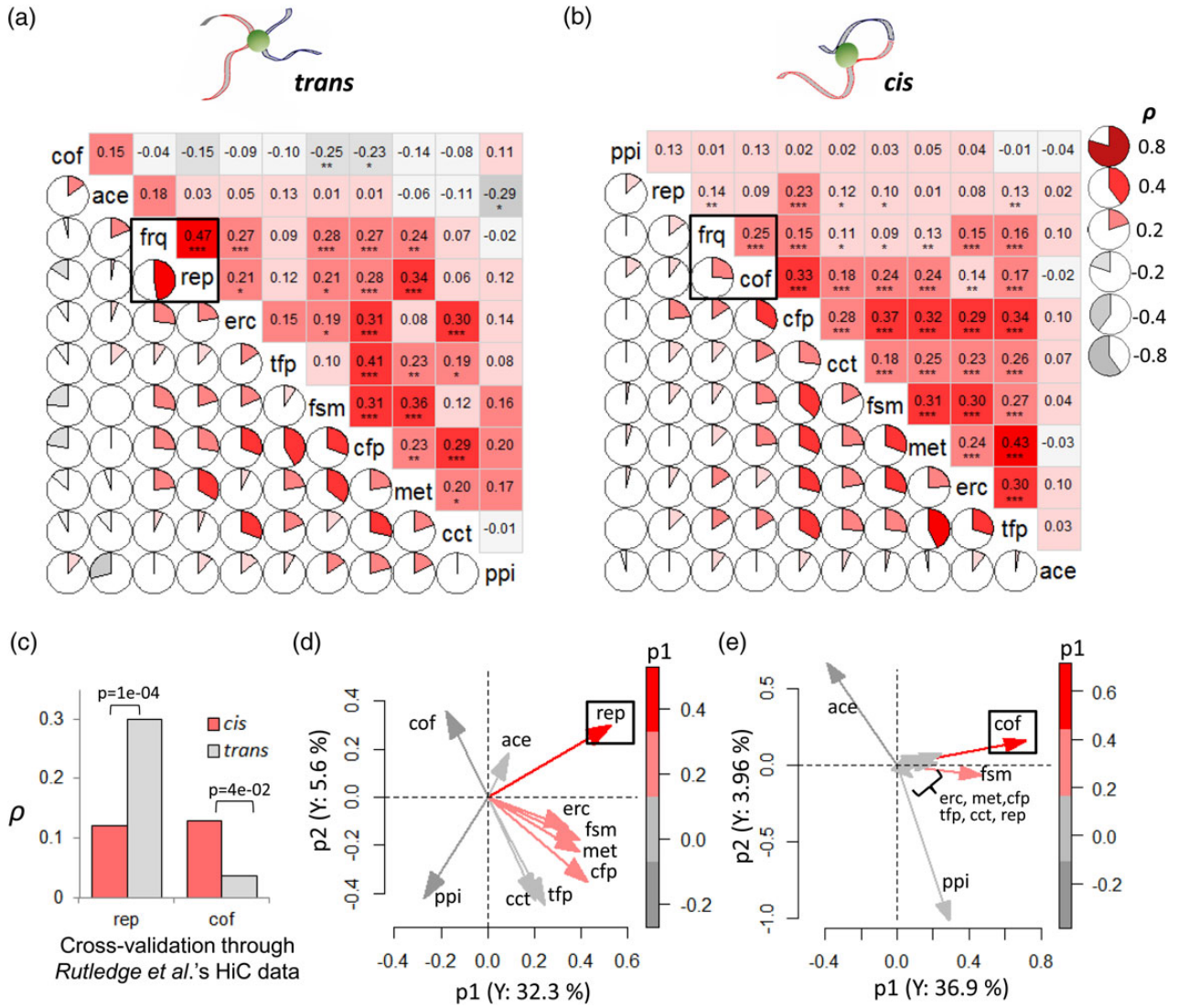

Figure 3. Multivariate analyses of *trans* and *cis* chromatin interactions in yeast genome. (a and b) Correlograms of distinct functional attributes of *trans*- and *cis*-interacting genes. The upper triangle shows the heatmap and the correlation values for each comparison. Lower triangle is a pie chart representation of Pearson's correlation coefficients. Triple asterisks (***) indicate *P*-value < 10^{-4} that corresponds to FDR < 0.01 for multiple comparisons of all 55 correlations between chromatin interaction frequency and other functional variables. Double asterisks indicate *P*-value < 10^{-3}, and single asterisk indicates *P*-value < 10^{-2}. Plots are made using 'corrgram' package on R. Frq, interaction frequency; cof, co-fitness; ppi, protein–protein interaction; cct, cell cycle time course; tfp, transcription factor perturbation; ace, acetylation; met, methylation; erc, evolutionary rate covariation; cfp, chromatin factor perturbation; fsm, functional similarity; rep, replication. Data in correlograms are ordered as per PCA clustering. 'ρ' stands for Pearson's correlation coefficient. Detailed control analyses are given in Supplementary Fig. S3. Details of three-letter abbreviations are given in the Table 1. (c) Cross-validation of differential correlations of *cis* and *trans* interaction frequencies with 'rep' and 'cof' through an independent HiC data of exponentially growing budding yeast cells (Rutledge et al.[40]). 'ρ' stands for Pearson's correlation coefficient. (d and e) Bi-plot of leading two X-loading vectors p1 and p2 (columns of matrix P) of PLSR model for (d) *trans*- and (e) *cis*-interacting gene pairs. The arrows are shaded as per p1 loading. Loading vectors p1 and p2 correspond to leading latent vectors t1 and t2 of matrix T, which explain 32.3 and 5.6% variance of Y ($n \times 1$ matrix of interaction frequency) for *trans* interactions and 36.9 and 3.96% variance of *cis* interaction frequency, respectively. The datasets plotted in a and b are given in the Supplementary Tables S3 and S4. The loadings for all the components are given in the Supplementary Fig. S5. This figure is available in black and white in print and in colour at *DNA Research* online.

3.1e–02, Fig. 4a). Interestingly, we observed that the late, but not the early, firing origins non-randomly interacted in *cis* (*P*-value = 1.9e–02; Fig. 4a), suggesting that the close physical proximity might be a common property of origins of replication and that early origins were *trans* interacting and late origins were *cis* interacting. Early and late origins showed significantly lesser number of *cis* and *trans* interactions with each other compared with null distribution, i.e. early-to-late interactions were under-represented compared with null distribution, strongly suggesting that they do not interact with each other in *cis* or *trans* ($P\text{-value}_{cis} = 0.003$, $P\text{-value}_{trans} = 0.012$; Supplementary Fig. S6). It is noteworthy that the significant *P*-values in the Fig. 4a did not approach extreme partly due to small sample sizes ($n = 78$ and 122 for early and late origins) and stringent null models.

These observations suggested that the correlation between *trans* interaction frequency and correlated replication seen in Fig. 2a might be associated with the non-random *trans* interactions among early origins. On similar lines, relatively weaker correlation ($\rho = 0.14$) of *cis* interaction frequency with the coordinated replication might be

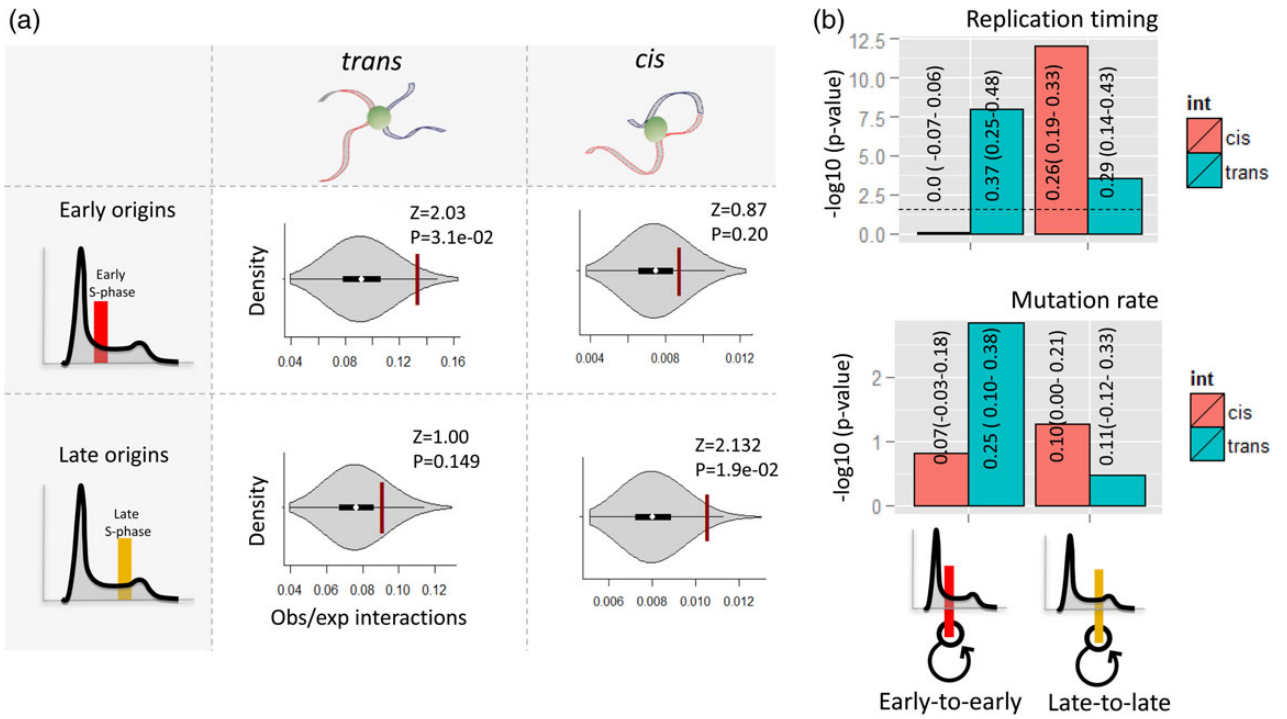

Figure 4. Association of *cis* and *trans* chromatin interactions with the coordinated replication and co-varying mutation rate. (a) Spatial connectivity (observed/expected number of interactions) among CLb5-independent early firing origins and Clb5-dependent late firing origins, with respect to null distribution generated through re-sampling-based method. Cartoon on the left-hand side represents the atypical cell-cycle curve representing number of cells (*y*-axis) as a function of replicated DNA content (*x*-axis). The first peak represents cells in G1 phase with 2N DNA followed by S-phase, and the second peak representing cells in G2 phase with 4N total DNA. The vertical bars in the cartoon represent early and late S-phase respectively. (b) Significance of correlation between interaction frequency and coordinated replication (upper panel), and similarity of mutation rate (lower panel) for *cis* and *trans* interactions. Interactions are split into EE (early-to-early) and LL (late-to-late) categories. Horizontal dotted lines represent *P*-value of 0.05. Over each bar, corresponding Pearson's correlations and 95% confidence values are given. Datasets of b are given in the Supplementary Tables S5 and S6. This figure is available in black and white in print and in colour at *DNA Research* online.

associated with the non-random *cis* interactions among late origins (Fig. 3b). To further explore whether *trans* and *cis* interactions associate with the coordination of early and late replication, respectively, we split the yeast genome into early and late replicating domains as per data from McCune et al.[41] and accordingly compiled early-to-early and late-to-late interaction categories. Early-to-early (EE) interaction category exhibited significant correlation between coordinated replication and the *trans*, but not the *cis*, interaction frequency (ρ_{trans} = 0.37, ρ_{cis} = 0.0; Fig. 4b). On the contrary, in the late-to-late category, *cis* and *trans* interactions showed marginally differing correlations with coordinated replication (ρ_{cis} = 0.26, ρ_{trans} = 0.29; Fig. 4b). However, importantly, the *cis* interaction frequency in the late-to-late category showed greater significance and power of correlation with the coordinated replication compared with the *trans* interaction frequency (P-value$_{cis} = 8.7 \times 10^{-8}$, P-value$_{trans} = 2.7 \times 10^{-4}$; power$_{cis}$ = 0.99 and power$_{trans}$ = 0.83; Fig. 4b). Since the Pearson's correlation coefficient decreases and the power of the observed effect increases exponentially as the sample size increases, we reason that the differing sample sizes of *cis* and *trans* interactions of late replicating regions (n = 711 and 146 binned values, respectively) explain the observed difference in significance of apparently similar correlation values. It is also notable that the *cis* interactions, as a whole, were only weakly associated with the coordinated replication as demonstrated in Fig. 3b, the significance of correlation of late replication with the *cis* interactions was, therefore, non-trivial. This is also reinforced by our earlier observation that late origins interacted in *cis*, but not in *trans*. These results supported that the *trans* and *cis* interactions were organized vis-a-vis to correlated early and late replication states, respectively, possibly marking spatial segregation of early and late replication factories. Furthermore, replication is the major cause of mutations in the genome, and correlation between replication timing and mutation rates has been observed in the past.[42] We, therefore, asked whether spatially proximal loci exhibited similar mutation rates. We obtained synonymous single-nucleotide substitution rates, which were adjusted for neutrality, for this purpose (Supplementary Table S1). Again, we observed stronger correlation between similarity in mutation rates and the *trans* interaction frequency and relatively weaker correlation with the *cis* interaction frequency (ρ_{cis} = 0.14, ρ_{trans} = 0.36; P-value$_{cis}$ = 0.63, P-value$_{trans}$ = 2.2e–05; Supplementary Fig. S7). Early-to-early interactions in *trans* and late-to-late interactions in *cis* exhibited greater significance of correlations with the similarity in mutation rates of engaged genes (P-value$_{EE,cis}$ = 0.15, P-value$_{EE,trans}$ = 1.4e–03, P-value$_{LL,cis}$ = 5.4e–02, P-value$_{LL,trans}$ = 0.46; Fig. 4b), highlighting a possible role of 3D genome organization in mutagenic mechanisms that might have shaped the co-evolution of spatially proximal genes in yeast. We hypothesize that the genes sharing the same sub-nuclear compartment might experience similar microenvironment that causes single-nucleotide mutations during replication or similar deficiency/efficiency of DNA repair factors.

3.3. *trans* and *cis* interactions exhibit differential preference for nucleosome occupancy

Since the process of replication is intricately linked with the chromatin structure, we tested whether the observed distinction between *trans*

and *cis* interactomes was associated with nucleosome occupancy and explains their association with early and late replication, respectively. *Z*-normalized nucleosome occupancy in 3 kb bins across yeast genome was plotted against the number of *cis*, *trans* interactions, and the ratio of the two. The plots clearly showed significantly negative and positive correlations with *cis* and *trans* interactions, respectively ($\rho_{cis} = -0.12$, $\rho_{trans} = 0.13$; $P\text{-value}_{cis} < 2.2e{-}16$, $P\text{-value}_{trans} < 2.2e{-}16$; Fig. 5a). Accordingly, the ratio of number of *cis* and *trans* interactions inversely scaled with the nucleosome occupancy suggesting that *cis* and *trans* interactions, in general, associate with low and high nucleosome occupancy, respectively ($\rho_{cis/trans} = -0.10$; $P\text{-value}_{cis/trans} = 7.2e{-}08$; Fig. 5a). The Pearson's correlation coefficients observed in the Fig. 5a, though statistically significant, were relatively small in magnitude, which can be attributed to large sample size ($n = 4{,}031$). It is also noteworthy that the present study does not aim to claim the scaling of *cis* and *trans* interactions with the nucleosome occupancy, but the distinct level of nucleosome occupancy of regions significantly enriched with *cis* or *trans* interactions. Therefore, we identified the *cis*- and *trans*-rich regions using sliding window (20 kb) approach and applying cut-off for *cis*/*trans* peaks as 50% of maximal window value in the genome (Fig. 5b). Demarcating the regions that were enriched with *cis*- or *trans* interactions clearly highlighted greater nucleosome occupancy in the regions enriched with *trans* interactions and *vice versa* (*P*-value < 2.2e–16; Fig. 5b, boxplot). These observations suggested that the nucleosome-depleted 'open chromatin' regions in the yeast genome were more likely to fold in *cis*, possibly highlighting a mechanism of spatially insulating the highly transcribing open chromatin from the nucleosome-rich repressive or regulated domains. We indeed observed that *cis*-enriched, nucleosome-depleted domains had relatively higher transcriptional activity compared with *trans*-enriched regions ($P = 5.0e{-}03$; Fig. 5b, boxplot). We also confirmed that these correlations were not the artefacts of abundant *trans* interactions among centromeres, which are generally heterochromatized and might show higher nucleosomal occupancy. We removed the regions that were proximal (<50 kb) to centro/telomeres and recalculated the correlations. We consistently found significant negative and positive correlations of nucleosome occupancy with the *cis* and *trans* interaction frequencies, respectively ($\rho_{cis} = -0.12$, $\rho_{cis} = 0.19$; $P\text{-value}_{cis} < 2.2e{-}16$, $P\text{-value}_{trans} < 2.2e{-}16$).

To reconcile the observation in the context of replication, we assessed the nucleosome occupancy around the early origins and compared with that of late origins. To nullify the differential levels of

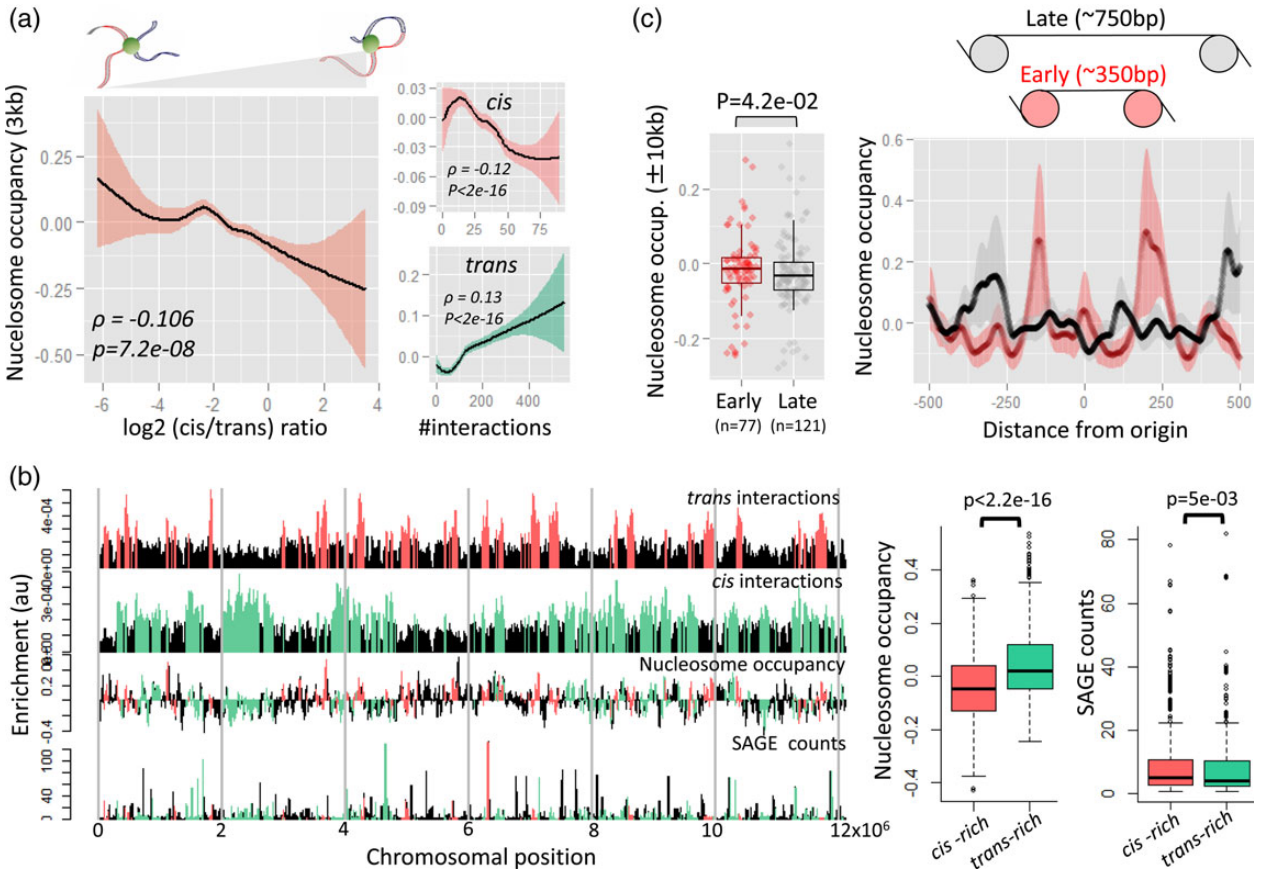

Figure 5. Association of *cis* and *trans* chromatin interactions with the nucleosome occupancy. (a) Left: Regression plot between the ratio of *cis* to *trans* interactions and nucleosomal occupancy. Right: regressions showing relationships between nucleosome occupancy and *cis*/*trans* interaction frequency. (b) Linear view of nucleosome occupancy and gene expression (SAGE counts) aligned along the number of *cis*/*trans* chromatin interactions. *cis*-Enriched and *trans*-enriched regions are highlighted. The peaks were identified as windows having sliding mean (averaged over 10 consecutive windows) greater than the 50% of the maximum window average in the genome. Boxplots on the right-hand side show distributions of nucleosome occupancy and SAGE counts in the *cis*-rich and *trans*-rich regions. (c) Left panel: average nucleosome occupancy (G1 phase) around ±10 kb of early and late origins. Right panel: average nucleosome occupancy profile around early and late origins. Shades represent the standard errors across all the early and late origins respectively. The datasets of b and c are given in Supplementary Tables S7 and S8, respectively. This figure is available in black and white in print and in colour at *DNA Research* online.

replicated DNA at early and late origins during S phase, we considered nucleosome occupancy in G1 phase of cell cycle only. Since we showed that the *trans* interactions were associated with early origins as well as with higher overall nucleosome occupancy, by inference, we expected the early origins to be associated with the higher nucleosome occupancy and vice versa for late origins. Indeed, we confirmed greater overall nucleosome occupancy around early origins compared with late origins (±10 kb, *P*-value = 5.0e-02, Fig. 5c). Interestingly, by analysing the nucleosome occupancy at single nucleosome resolution, we observed that the early origins were located in a narrow (~350 bp) nucleosome-depleted region (NDR) despite having an overall higher nucleosome occupancy (Fig. 5c). Such nucleosomal organization had been shown to facilitate the replication initiation.[43] Late origins, on the contrary, were located in a relatively wider NDR (~750 bp), significance of which is not entirely clear yet (Fig. 5c). However, we speculated that the late origins might need accessibility to additional sequence features for the regulatory factors involved in late replication. Indeed, such sequence features have been proposed earlier by others too.[44,45] Role of *cis* regulatory elements in late firing of origins is further discussed in the Discussion section.

3.4. Association of *cis* interactions with the essential gene clusters

Strong correlation between *cis* interaction frequency and the co-fitness of interacting genes suggested that the genes with similar fitness defects across different environments tend to co-localize more frequently. This would also mean that genes with the same extreme growth defects, like lethality, would tend to interact with greater interaction frequency. We, therefore, tested whether the essential genes were interacting with each other in *cis* with a significantly greater frequency compared with random null model. We extracted the *cis* and *trans* interactions that had essential genes on both restriction fragments. Indeed, the resampling approach uncovered significantly greater frequency of *cis* (*P*-value = 5e–02), but not the *trans* (*P*-value = 0.16), interactions among essential genes (Fig. 6a). Essential genes are generally located in the regions having an overall low nucleosome occupancy and consequently exhibit lower expression noise.[46] We, therefore, propose a hypothesis that the essential genes should also be relatively stable in the three-dimensional nuclear space to keep the expression noise low. Greater mobility might introduce noise in the expression, which can be deleterious in case of essential genes. We presumed that a stable locus would be expected to show: (i) high average value of its interaction frequencies to all its *cis*-interacting partners, that we referred as 'average interaction frequency'; (ii) relatively short-range interactions and lower variance in the genomic distances between interacting loci; and (iii) greater consistency of chromatin interactions across biological replicates in HiC experiments, measured as overlap of interactions of a locus commonly captured in *Eco*RI and *Hin*dIII HiC libraries to all its interactions captured in *Hin*dIII library. Although it is not possible to distinguish biological variations from technical variations in the present scenario, some studies have shown that variation between biological replicates are largely biological than technical and can be attributed to genetic, epigenetic, or

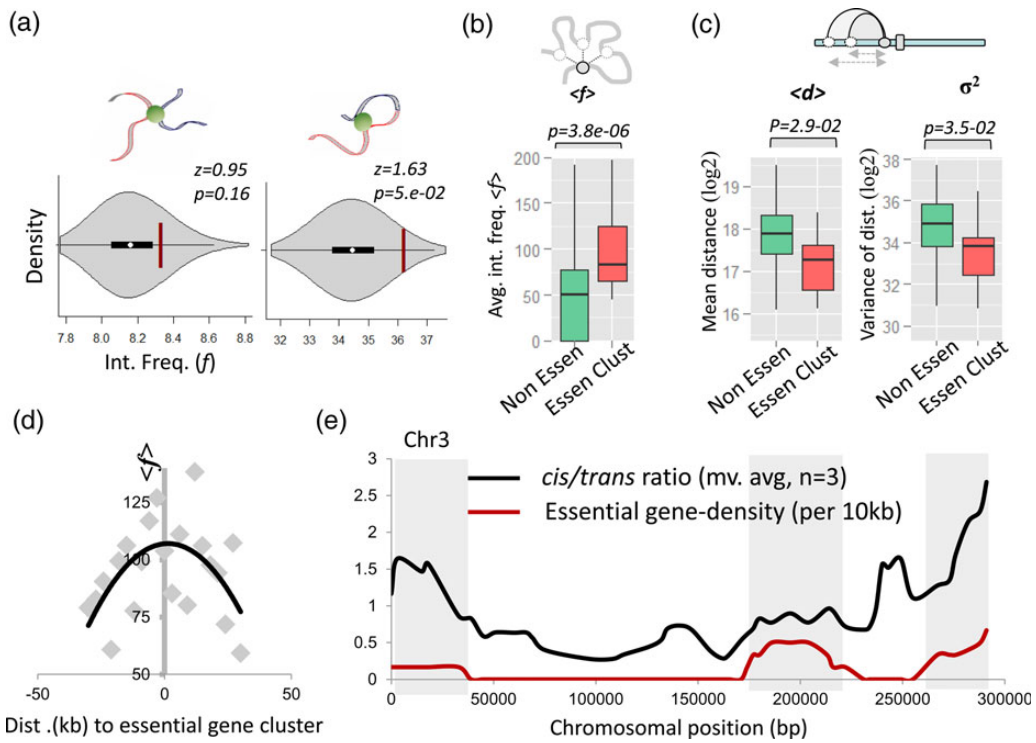

Figure 6. Association of *cis* interactions with the gene essentiality. (a) Average frequency of *trans* and *cis* interactions among essential genes overlaid on the null distribution of average values across 1,000 random samples. Cartoon on the top depicts all *cis* interactions of a locus (grey-coloured node), from which the average interaction frequency of that locus was calculated. (b) Average frequency of *cis* interactions of genomic loci having and not having essential gene clusters. (c) Mean and variance of genomic distances between interaction sites at genomic loci having and not having essential gene clusters. Cartoon on the top represents the genomic distance between *cis*-interacting sites of grey-coloured node. (d) Average *cis* interaction frequency as a function of relative distance from the essential gene cluster (defined as >2 essential genes in 5 kb window). (e) Average *cis* interaction frequency (moving average of three consecutive restriction fragments) and essential gene density along Chromosome 3. Data plotted in b–e are given in the Supplementary Table S9. This figure is available in black and white in print and in colour at *DNA Research* online.

stochastic mosaicism of cells.[47–52] The chromatin loci exhibiting greater consistency of chromatin interactions in biological replicates can, therefore, be considered relatively restrained in the nuclear space compared with the ones showing greater variations. By applying the aforementioned measures, we observed that the loci having essential gene clusters (>2 gene in 5 kb bin) consistently showed: (i) higher average interaction frequency (*P*-value = 3.8e–06; Fig. 6b), (ii) relatively short-range interactions as depicted by mean distance between interacting sites (*P*-value = 2.9e–02; Fig. 6c); (iii) lower variance in genomic distances between interacting sites (*P*-value = 3.5e–02; Fig. 6c); and (iv) greater overlap between biological replicates of HiC data, compared with rest of genome (Supplementary Fig. S8). Further, the average interaction frequency decreases as a function of genomic distance from the essential gene clusters, marking the selective enrichment of high frequency interaction on and around essential gene clusters (Fig. 6d and e, Supplementary Fig. S9). Thus, the higher average interaction frequency, relatively short-range interactions of lower variance, and consistency across HiC replicates might suggest spatially clustered and stable nature of genomic regions having essential gene clusters. We also confirmed the link between gene essentiality and the average interaction frequency analysing average fitness defect, as measure from chemical genomics data, at sites of high average interaction frequency (*P*-value = 9.0e–06; Supplementary Fig. S10).

3.5. *cis* interactions are constrained by minimization of expression noise

To directly test whether or not spatially restrained loci were associated with the lower noise in the expression, we extracted high noise and low noise genes (upper and lower quartile of genome-wide abundance-corrected expression noise data) and assessed their average interaction frequencies, mean and variance of genomic distances between interacting sites, and the overlap ratio of their interactions captured in the HiC replicates. The genes with low expression noise exhibited greater average interaction frequency (*P*-value = 2e–02; Fig. 7a), relatively lower mean and variance of genomic distances between interacting loci ($P\text{-value}_{mean}$ = 9e–03, $P\text{-value}_{variance}$ = 4e–02; Fig. 7b), and greater overlap ratio of interactions in the HiC biological replicates (*P*-value = 3.3e05; Fig. 7), supporting the hypothesis that spatially restrained loci would tend to have lower expression noise. Indeed, the expression noise of essential genes inversely scaled with the average *cis*, but not the *trans* interaction frequency of the loci (Supplementary Fig. S11). Further, it is known that the genes that are toxic when over-expressed also exhibit low expression noise.[46] We, therefore, tested whether these genes are also spatially constrained by high frequency of *cis* interactions. Surprisingly, genes exhibiting over-expression toxicity did not show significantly greater average interaction frequency as in the case of essential gene clusters (Supplementary Fig. S12). We propose that spatially restrained or 'tethered' property is specifically associated with the mechanism that minimizes stochastic loss of expression but not the abrupt gain of expression.

To further scrutinize our observations, we performed following additional analyses. (i) We identified the regions of significantly greater average frequency of *cis* interactions from the interactions commonly present in the two HiC replicates. The distribution plot of average *cis* interaction frequency across yeast genome revealed a distinct population of regions having significantly greater interaction frequency

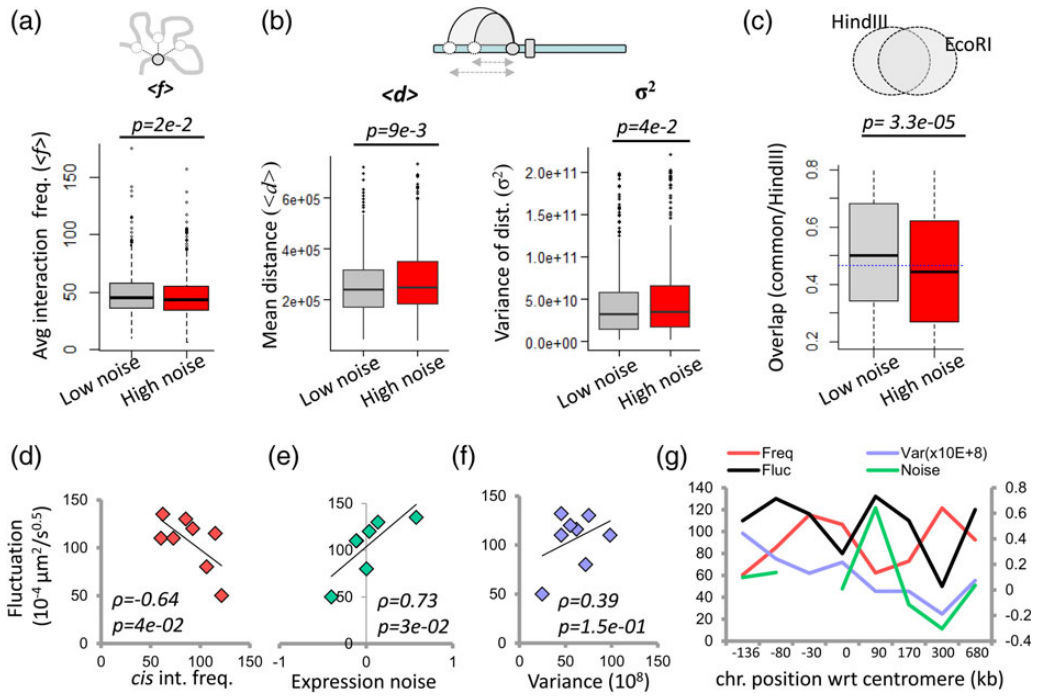

Figure 7. Association of *cis* interactions with expression noise and chromatin fluctuations. (a) Distribution of average *cis* interaction frequencies for low (≤1st quartile) and high noise (≥3rd quartile) genes. Cartoon on the top depicts all *cis* interactions of a locus (grey-coloured node), from which the average interaction frequency of grey-coloured locus was calculated. (b) Distribution of mean and variance of genomic distances between interacting sites for low and high noise genes. Cartoon on the top represents the genomic distance between *cis*-interacting sites of grey-coloured node. (c) Distribution of overlap ratio of chromatin interactions between *Hin*dIII and *Eco*RI HiC replicates for low and high noise genes. Horizontal dotted line represents the genome-wide median value. (d–f) Scatter plot between and experimentally determined chromatin fluctuations and (d) average *cis* interaction frequency, (e) expression noise, and (f) variance of genomic distances between interacting sites on yeast Chromosome 12. (g) Expression noise, average *cis* interaction frequency, variance of genomic distances, and chromatin fluctuations aligned along Chromosome 12. Data plotted in a–c and d–g are given in Supplementary Tables S9 and S10, respectively. This figure is available in black and white in print and in colour at *DNA Research* online.

than the rest (Supplementary Fig. S13a). This is analogous to identifying 'peaks' in ChIP-Seq genomic tracks. The 'peak' nature of certain regions was also confirmed by plotting QQ plot of the average interaction frequency against a random normal distribution (Supplementary Fig. 13b and c). We extracted the genes located in these regions and compared the expression noise with the null distribution. Again, we observed significantly lower (P-value = 1e–03) noise for the genes located in the regions of high average interaction frequency. (ii) Lower diversity of interactions might also represent relatively stable state of a locus. We measured the variation in chromatin environment using coefficient of variation (σ/μ) and observed that the genes having greater variation in their chromatin interactions exhibited higher expression noise (P-value = 3e–02; Supplementary Fig. S14). (iii) More importantly, we compared the experimentally determined chromatin mobility of several loci[53,54] with their average *cis* interaction frequencies, variances of distances between interacting sites, and the expression noise (Fig. 7d–g, Supplementary Fig. S15). The analyses showed striking correlation of chromatin fluctuations with the expression noise ($\rho = 0.78$, Fig. 7d), moderate correlation with the variance of genomic distances between interacting loci ($\rho = 0.39$, Fig. 7e), and strong anticorrelation with the *cis* interaction frequency ($\rho = -0.65$, Fig. 7f), strongly supporting our proposal that the chromatin mobility associates with the expression noise of the underlying genes.

Although we largely restrict all our claims to the budding yeast genome, we now present a few lines of evidence which suggest that minimization of expression noise through physically restrained environment might be a general evolutionary constraint shaping the genome organization in other radically different systems like bacteria and the mammalian cells. The expression noise data had been available for *Escherichia coli* too[55] and very recently genome organization data of *E. coli* have also been generated using Genome Conformation Capture technique.[56,57] It would, therefore, be interesting to test whether the proposed relationship between interaction frequency and expression noise also exists in Bacteria. We performed similar analyses in *E. coli* by taking normalized interaction frequency data for 10 kb bins as provided by Xie et al.,[56] number of essential genes and the maximal expression noise value in each bin. We found strikingly similar association among gene essentiality, expression noise, and average interaction frequency in *E. coli*, as in yeast (Fig. 8a–c). This is an important observation highlighting an evolutionarily conserved association between expression noise and genome organization despite having radically distinct mechanisms of gene expression. Further, genome-wide single-cell gene expression data are now available for several model systems. We obtained one such dataset for mouse embryonic stem cells (mESCs), for which genome-wide chromatin interaction data was also available.[35,58] We calculated the noise in gene expression for mESCs and corrected for the transcript abundance using residuals of lowess fit between transcript abundance and noise. Again, the average interaction frequency of low noise genes was significantly greater than that of high noise genes (P-value = 3.0e–11;

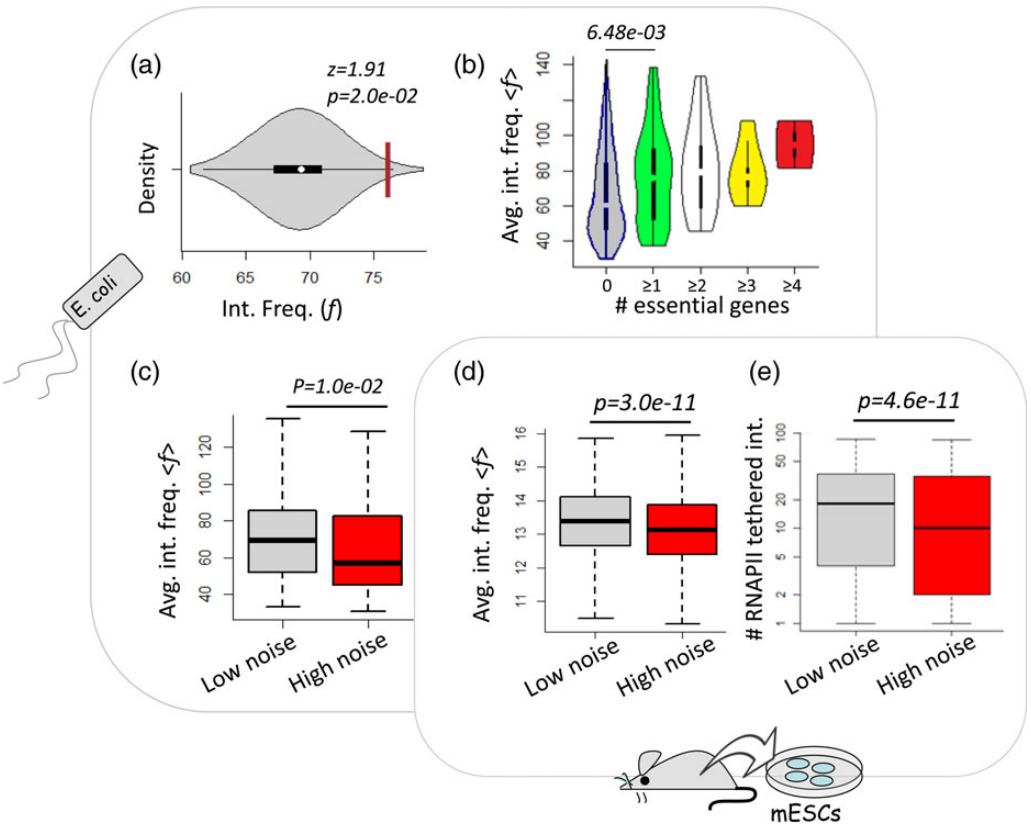

Figure 8. Association of chromatin interactions with the expression noise in other systems (a) Average frequency of interactions among genomic bins having essential genes overlaid on the null distribution of average values across 1,000 random re-samplings in *Escherichia coli*. (b) Violin plots of average interaction frequencies of loci having distinct densities of essential genes per 5 kb in *E. coli* genome. The categories >0, >1, and >2 are inclusive and not mutually exclusive. (c) Distribution of interaction frequencies for low (≤1st quartile) and high noise (≥3rd quartile) genes in *E. coli*. (d) Distribution of interaction frequencies for low (≤1st quartile) and high noise (≥3rd quartile) genes in mESCs. (e) Distribution of number of RNAPII-associated chromatin interactions of low and high noise genes in mESCs. Mapped *E. coli* and mESC datasets are given in Supplementary Table S11 and S12. This figure is available in black and white in print and in colour at *DNA Research* online.

Fig. 8d). Due to lack of whole-genome single-cell gene expression data along with HiC data on other cell types of mammalian genome, our observation on mESCs is presently not generalizable to different cell types in higher eukaryotes.

The above analyses clearly strengthened our proposal that essential genes exhibit lesser variation in its chromatin environment and that such a property contributes in reducing the expression noise of engaged loci.

4. Discussion

It is known that spatial positioning of genes in the nucleus is highly non-random. Genes positioned interior to the nucleus experience transcriptionally permissive environment, while the ones located near the nuclear periphery are generally repressed or lowly expressed.[1] Whether or not spatial co-localization of genes plays an active role in regulating essential genomic functions is presently subjected to intense scrutiny. It has been presumed that genes co-localize to synchronize their transcriptional states.[59,60] Besides sporadic reports of co-expression of individual loci,[19,61,62] genome-wide proposal has been made in case of multi-gene complexes (relatively short-range intra-chromosomal interactions) in mammals[34] and for both intra- and inter-chromosomal interactions in lower eukaryotes like yeast.[37,38] However, these studies are not subjected to comprehensive multivariate analyses and in the absence of other genomic variables, these reports do not present an unbiased view of possible functional constraints shaping the three-dimensional genome organization of eukaryotic genome. Here we analysed several functional attributes of yeast genome to identify the potential constraints of genome-wide chromatin interactions. Our analyses suggested that inter- and intra-chromosomal interactions were under distinct evolutionary constraints. While inter-chromosomal interactions were primarily associated with the coordination of Clb5-independent replication, intra-chromosomal interactions were constrained by the coordination of Clb5-dependent replication and minimization of expression noise of engaged loci. Correlation of interaction frequency with the co-expression of engaged genes was very weak. The correlation coefficient of similar magnitude was also reported earlier by others.[37] However, the authors compared the correlation coefficient with the average correlation of all possible gene pairs in the genome and found the correlation of 0.09 to be significantly higher. We argue that genome-wide average might not serve as an appropriate control, because the sample size of gene pairs will be disproportionally greater for the whole genome compared with the ones that are spatially proximal. Rewired contact networks and re-sampling-based approach can serve as better control. Re-assessment of reports claiming co-localization of co-regulated genes have been questioned elsewhere too.[39] More importantly, in the absence of other functional variables, it is not entirely justified to claim that co-expression if the major constraint of genome organization.

The specific association of *trans* and *cis* interaction frequency with early and the late replication, respectively, supports the notion of spatially segregated early and late replication factories, which had been proposed earlier by several authors.[63–65] Interestingly, a recent report has clearly shown that androgen-induced proximity between TMPRSS2 and ERG genes is due to AR-controlled replication, but not the transcription.[66] Further, the association of preferred domains of *cis* and *trans* interactions with the low and high nucleosome occupancy, respectively, supports the spatial segregation of open and closed chromatin. The observation that the early origins were flanked by well-positioned nucleosome around a narrow trough of nucleosome-depleted region was in line with earlier reports suggesting the role ORC mediated nucleosome positioning in establishment of pre-initiation complex around early origins.[43] Relatively wider NDRs around late origins can be explained through following arguments: (i) Late firing of origins might need additional *cis* regulatory elements in the proximity. It is known that after recycling from early origins, replication factors are present in a very limited amount for late origins. Late origins, therefore, cannot afford the delay caused by the remodelling of nucleosomes to open up the binding sites of replication factors. Constitutively open chromatin and spatial co-localization of late origins therefore might help in efficient and rapid usage of replication factors and ensure the complete replication of the genome. (ii) Late origins are known to be present predominantly in the intergenic regions (IGRs) between convergent genes.[67] The abundant occurrence of transcription termination around convergent IGRs can interfere with the nucleosome stability at late origins. The nucleosome depletion around late origins can thus be an artefact of genomic neighbourhood. (iii) It is known that the radial positioning early origins in the nuclear space is mostly random, while late origins often localize towards nuclear periphery. The accessibility to sequences flaking late origins can help targeting the late origins to nuclear periphery.[68]

The association of chromatin interactions with the mutation rate variation suggests a potential role of spatial genome organization in mutagenic mechanisms. Coordinated replication through spatial convergence might also ascribe similar susceptibility to genetic errors at engaged loci. Such properties had earlier been proposed in the context of cancer genomes.[69,70]

Intra-chromosomal interactions, on the other hand, were also constrained by co-fitness of genes. It has been proposed that the gene pairs with higher co-fitness represent their close functional similarity. It can, therefore, partly be explained by the co-functionality of *cis*-interacting genes as shown in the Fig. 2e. We further attempted to explain the link between fitness defect and genome organization by taking essential genes as example. We hypothesized that the high frequency intra-chromosomal interactions restrict the mobility of engaged loci, while low frequency interactions of a locus might represent a relatively mobile state of chromatin. The presumption that the high interaction frequency and the lower coefficient of variation represent spatially restrained loci in genome is also supported by comparison with the experimentally determined mobility of certain loci[54] (Fig. 8d–g, Supplementary Fig. S12). Restricted mobility of interacting loci might be important to reduce the transcriptional noise of interacting genes. We reconciled this hypothesis by taking the example of essential genes. Essential gene clusters are known to exhibit lower expression noise and consistently remain in transcriptionally permissive open chromatin state. Their non-randomly greater frequency of interactions aligns to our hypothesis. It was noteworthy that we did not consider the chromatin interactions between loci which were <20 kb apart, and therefore, the linear clustering of essential genes would not have impacted this observation. Further, the association of low noise genes with greater mean interaction frequency of loci and *vice versa* across bacteria, yeast, and mouse genomes clearly suggests that spatially stable nature of genes assist in minimizing the stochastic transcriptional errors. Indeed, it has earlier been proposed that the long-range chromatin interactions might occur at the cost of increased expression noise.[71] To which sub-nuclear compartment these loci might tether to? We propose that for active genes, the most likely tethering foci would be transcription factories. Though this would need thorough scrutiny, we explored the RNAPII-associated chromatin interaction data generated through ChIA-PET technique for this purpose. Multi-gene complexes in that data have been proposed to be equivalent to transcription factories.[34] We asked whether abundance-corrected

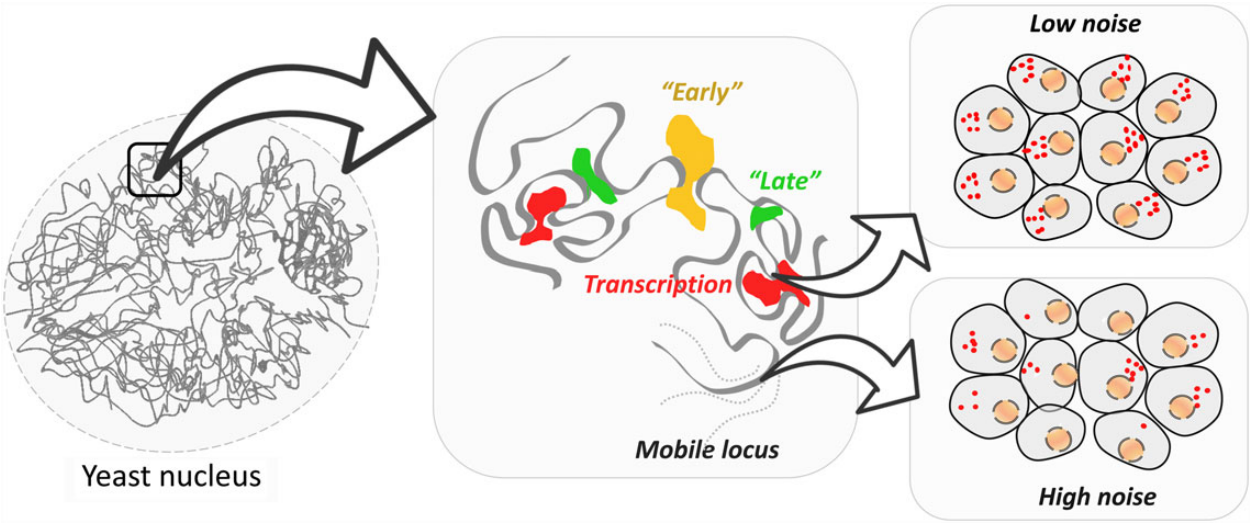

Figure 9. Schematic representation of overall observations. At higher resolution, the yeast genome is organized into preferred domains of inter- and intra-chromosomal interactions, which is associated with the spatial segregation of early and late replication. As inferred from high average interaction frequency and chromatin fluctuations, essential gene loci are hypothesized to be physically restrained possibly by tethering to nuclear sub-compartments or foci. The proposed physically restrained nature of essential genes might be important to mitigate the epigenetic fluctuations or errors that predispose the genes to stochastic variation in expression. This figure is available in black and white in print and in colour at *DNA Research* online.

transcriptional noise associate with these complexes. We observed that the number of RNAPII-tethered promoter-to-promoter interactions, marking multi-gene complexes, was significantly higher for low noise genes compared with high noise genes ($P < 3e{-}11$), suggesting that genes engaged in multi-gene complexes or transcription factories tend to have lower transcriptional noise (Fig. 8e). Other nuclear compartments like nucleoli have been shown experimentally to constrain the chromatin movement, and their disruption leads to increased chromatin mobility.[72]

Altogether, our unbiased approach of analysing distinct functional constraints provides a different perspective of evolution of three-dimensional genome organizations, which appears to be conserved in bacteria, yeast, and mouse. *Escherichia coli* genome is known to have an organization tightly linked with the replication,[57] and here we showed that it is also associated with gene essentiality and expression noise, very similar to what we observed in yeast. Finally, based on our observations, we propose a model for functional constraints shaping the *cis* and *trans* organization of chromatin. This is illustrated in the Fig. 9.

5. Conclusion

The study suggests that there are different set of functional constraints that shape intra- and inter-chromosomal interactomes in eukaryotes. Distinct spatial organization of early and late origins and the underlying coordination of replication strongly support the presence of discrete early and late replication factories. A tethered intra-chromosomal microenvironment might ascribe physical stability to a locus which can be important to reduce the transcriptional noise of engaged loci, particularly, the ones that are functionally indispensable. Therefore, coordinated replication and gene essentiality, not necessarily the co-expression of genes, seem major functional and evolutionary constraints shaping the three-dimensional genome organization of eukaryotes as well as of prokaryotes.

Authors' contribution

A.S. performed most of the analyses. M.B. helped with the data analysis of mouse embryonic stem cells and performed some of the control analyses. K.S.S. conceived and supervised the project. K.S.S. has performed the statistical tests and drawn the figures.

Acknowledgements

Authors acknowledge the financial support from Ministry of Human Resource and Development (MHRD), India.

Conflict of interest statement

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

Funding to pay the Open Access publication charges for this article was provided by Ministry of Human Resource Development (MHRD), India.

References

1. Tanabe, H., Muller, S., Neusser, M., et al. 2002, Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates, *Proc. Natl Acad. Sci. USA*, **99**, 4424–9.
2. Krivega, I. and Dean, A. 2012, Enhancer and promoter interactions-long distance calls, *Curr. Opin. Genet. Dev.*, **22**, 79–85.
3. Petrascheck, M., Escher, D., Mahmoudi, T., Verrijzer, C.P., Schaffner, W. and Barberis, A. 2005, DNA looping induced by a transcriptional enhancer in vivo, *Nucleic Acids Res.*, **33**, 3743–50.
4. Bulger, M. and Groudine, M. 2010, Enhancers: the abundance and function of regulatory sequences beyond promoters, *Dev. Biol.*, **339**, 250–7.
5. Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., et al. 2014, Enhancer loops appear stable during development and are associated with paused polymerase, *Nature*, **512**, 96–100.

6. 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. USA*, **106**, 20222–27.
7. West, A.G. and Fraser, P. 2005, Remote control of gene transcription, *Hum. Mol. Genet.*, **14**, R101–11.
8. Ong, C.T. and Corces, V.G. 2011, Enhancer function: new insights into the regulation of tissue-specific gene expression, *Nat. Rev. Genet.*, **12**, 283–93.
9. Lanctot, C., Cheutin, T., Cremer, M., Cavalli, G. and Cremer, T. 2007, Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions, *Nat. Rev. Genet.*, **8**, 104–15.
10. Fraser, P. and Bickmore, W. 2007, Nuclear organization of the genome and the potential for gene regulation, *Nature*, **447**, 413–7.
11. Osborne, C.S., Chakalova, L., Brown, K.E., et al. 2004, Active genes dynamically colocalize to shared sites of ongoing transcription, *Nat. Genet.*, **36**, 1065–71.
12. Eskiw, C.H. and Fraser, P. 2011, Ultrastructural study of transcription factories in mouse erythroblasts, *J. Cell Sci.*, **124**, 3676–83.
13. Martin, S. and Pombo, A. 2003, Transcription factories: quantitative studies of nanostructures in the mammalian nucleus, *Chromosome Res.*, **11**, 461–70.
14. Cook, P.R. 2010, A model for all genomes: the role of transcription factories, *J. Mol. Biol.*, **395**, 1–10.
15. Sutherland, H. and Bickmore, W.A. 2009, Transcription factories: gene expression in unions? *Nat. Rev. Genet.*, **10**, 457–66.
16. Jackson, D.A., Iborra, F.J., Manders, E.M. and Cook, P.R. 1998, Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei, *Mol. Biol. Cell*, **9**, 1523–36.
17. Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. 2002, Capturing chromosome conformation, *Science*, **295**, 1306–11.
18. Zhao, Z., Tavoosidana, G., Sjolinder, M., et al. 2006, Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions, *Nat. Genet.*, **38**, 1341–7.
19. Fullwood, M.J., Liu, M.H., Pan, Y.F., et al. 2009, An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, **462**, 58–64.
20. Dostie, J., Richmond, T.A., Arnaout, R.A., et al. 2006, Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements, *Genome Res.*, **16**, 1299–309.
21. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., et al. 2009, Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science*, **326**, 289–93.
22. Duan, Z., Andronescu, M., Schutz, K., et al. 2010, A three-dimensional model of the yeast genome, *Nature*, **465**, 363–7.
23. Wang, C., Liu, C., Roqueiro, D., et al. 2015, Genome-wide analysis of local chromatin packing in Arabidopsis thaliana, *Genome Res.*, **25**, 246–56.
24. Sexton, T., Yaffe, E., Kenigsberg, E., et al. 2012, Three-dimensional folding and functional organization principles of the Drosophila genome, *Cell*, **148**, 458–72.
25. Umbarger, M.A., Toro, E., Wright, M.A., et al. 2011, The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation, *Mol. Cell*, **44**, 252–64.
26. Tanizawa, H., Iwasaki, O., Tanaka, A., et al. 2010, Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation, *Nucleic Acids Res.*, **38**, 8164–77.
27. Dixon, J.R., Selvaraj, S., Yue, F., et al. 2012, Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature*, **485**, 376–80.
28. Dixon, J.R., Jung, I., Selvaraj, S., et al. 2015, Chromatin architecture reorganization during stem cell differentiation, *Nature*, **518**, 331–6.
29. Ciabrelli, F. and Cavalli, G. 2015, Chromatin-driven behavior of topologically associating domains, *J. Mol. Biol.*, **427**, 608–25.
30. Pope, B.D., Ryba, T., Dileep, V., et al. 2014, Topologically associating domains are stable units of replication-timing regulation, *Nature*, **515**, 402–5.
31. Zhang, Y., Wong, C.H., Birnbaum, R.Y., et al. 2013, Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations, *Nature*, **504**, 306–10.
32. Jin, F., Li, Y., Dixon, J.R., et al. 2013, A high-resolution map of the three-dimensional chromatin interactome in human cells, *Nature*, **503**, 290–4.
33. Sanyal, A., Lajoie, B.R., Jain, G. and Dekker, J. 2012, The long-range interaction landscape of gene promoters, *Nature*, **489**, 109–13.
34. Li, G., Ruan, X., Auerbach, R.K., et al. 2012, Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, *Cell*, **148**, 84–98.
35. Schoenfelder, S., Furlan-Magaril, M., Mifsud, B., et al. 2015, The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements, *Genome Res.*, **25**, 582–97.
36. Chepelev, I., Wei, G., Wangsa, D., Tang, Q. and Zhao, K. 2012, Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization, *Cell Res.*, **22**, 490–503.
37. Homouz, D. and Kudlicki, A.S. 2013, The 3D organization of the yeast genome correlates with co-expression and reflects functional relations between genes, *PLoS One*, **8**, e54699.
38. Dai, Z. and Dai, X. 2012, Nuclear colocalization of transcription factor target genes strengthens coregulation in yeast, *Nucleic Acids Res.*, **40**, 27–36.
39. Witten, D.M. and Noble, W.S. 2012, On the assessment of statistical significance of three-dimensional colocalization of sets of genomic elements, *Nucleic Acids Res.*, **40**, 3849–55.
40. 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**, 8299–313.
41. McCune, H.J., Danielson, L.S., Alvino, G.M., et al. 2008, The temporal program of chromosome replication: genomewide replication in clb5 {Delta} Saccharomyces cerevisiae, *Genetics*, **180**, 1833–47.
42. Lang, G.I. and Murray, A.W. 2011, Mutation rates across budding yeast chromosome VI are correlated with replication timing, *Genome Biol. Evol.*, **3**, 799–811.
43. Lipford, J.R. and Bell, S.P. 2001, Nucleosomes positioned by ORC facilitate the initiation of DNA replication, *Mol. Cell*, **7**, 21–30.
44. Friedman, K.L., Diller, J.D., Ferguson, B.M., Nyland, S.V., Brewer, B.J. and Fangman, W.L. 1996, Multiple determinants controlling activation of yeast replication origins late in S phase, *Genes Dev.*, **10**, 1595–607.
45. Berbenetz, N.M., Nislow, C. and Brown, G.W. 2010, Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure, *PLoS Genet.*, **6**, e1001092.
46. Batada, N.N. and Hurst, L.D. 2007, Evolution of chromosome organization driven by selection for reduced gene expression noise, *Nat. Genet.*, **39**, 945–9.
47. Blainey, P., Krzywinski, M. and Altman, N. 2014, Points of significance: replication, *Nat. Methods*, **11**, 879–80.
48. De, S. 2011, Somatic mosaicism in healthy human tissues, *Trends Genet.*, **27**, 217–23.
49. Herndon, L.A., Schmeissner, P.J., Dudaronek, J.M., et al. 2002, Stochastic and genetic factors influence tissue-specific decline in ageing C. elegans, *Nature*, **419**, 808–14.
50. Molloy, M.P., Brzezinski, E.E., Hang, J., McDowell, M.T. and VanBogelen, R.A. 2003, Overcoming technical variation and biological variation in quantitative proteomics, *Proteomics*, **3**, 1912–9.
51. Fernandez, L.C., Torres, M. and Real, F.X. 2016, Somatic mosaicism: on the road to cancer, *Nat. Rev. Cancer*, **16**, 43–55.
52. Chalancon, G., Ravarani, C.N., Balaji, S., et al. 2012, Interplay between gene expression noise and regulatory network architecture, *Trends Genet.*, **28**, 221–32.
53. Albert, B., Mathon, J., Shukla, A., et al. 2013, Systematic characterization of the conformation and dynamics of budding yeast chromosome XII, *J. Cell Biol.*, **202**, 201–10.
54. Heun, P., Laroche, T., Shimada, K., Furrer, P. and Gasser, S.M. 2001, Chromosome dynamics in the yeast interphase nucleus, *Science*, **294**, 2181–6.

55. Silander, O.K., Nikolic, N., Zaslaver, A., et al. 2012, A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli, *PLoS Genet.*, **8**, e1002443.
56. Xie, T., Fu, L.Y., Yang, Q.Y., et al. 2015, Spatial features for Escherichia coli genome organization, *BMC Genomics*, **16**, 37.
57. Cagliero, C., Grand, R.S., Jones, M.B., Jin, D.J. and O'Sullivan, J.M. 2013, Genome conformation capture reveals that the Escherichia coli chromosome is organized by replication and transcription, *Nucleic Acids Res.*, **41**, 6058–71.
58. Tang, F., Barbacioru, C., Bao, S., et al. 2010, Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis, *Cell Stem Cell*, **6**, 468–78.
59. Rajapakse, I., Perlman, M.D., Scalzo, D., Kooperberg, C., Groudine, M. and Kosak, S.T. 2009, The emergence of lineage-specific chromosomal topologies from coordinate gene regulation, *Proc. Natl Acad. Sci. USA*, **106**, 6679–84.
60. Szczepinska, T. and Pawlowski, K. 2013, Genomic positions of co-expressed genes: echoes of chromosome organisation in gene expression data, *BMC Res. Notes*, **6**, 229.
61. Schoenfelder, S., Sexton, T., Chakalova, L., et al. 2010, Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells, *Nat. Genet.*, **42**, 53–61.
62. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. and Micheletti, C. 2013, Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19, *PLoS Comput. Biol.*, **9**, e1003019.
63. Dimitrova, D.S. and Gilbert, D.M. 2000, Temporally coordinated assembly and disassembly of replication factories in the absence of DNA synthesis, *Nat. Cell Biol.*, **2**, 686–94.
64. Saner, N., Karschau, J., Natsume, T., et al. 2013, Stochastic association of neighboring replicons creates replication factories in budding yeast, *J. Cell Biol.*, **202**, 1001–12.
65. Gondor, A. and Ohlsson, R. 2009, Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? *Nat. Rev. Genet.*, **10**, 269–76.
66. Coll-Bastus, N., Mao, X., Young, B.D., Sheer, D. and Lu, Y.J. 2015, DNA replication-dependent induction of gene proximity by androgen, *Hum. Mol. Genet.*, **24**, 963–71.
67. Soriano, I., Morafraile, E.C., Vazquez, E., Antequera, F. and Segurado, M. 2014, Different nucleosomal architectures at early and late replicating origins in Saccharomyces cerevisiae, *BMC Genomics*, **15**, 791.
68. Heun, P., Laroche, T., Raghuraman, M.K. and Gasser, S.M. 2001, The positioning and dynamics of origins of replication in the budding yeast nucleus, *J. Cell Biol.*, **152**, 385–400.
69. Liu, L., De, S. and Michor, F. 2013, DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes, *Nat. Commun.*, **4**, 1502.
70. De, S. and Michor, F. 2011, DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes, *Nat. Biotechnol.*, **29**, 1103–8.
71. McCullagh, E., Seshan, A., El-Samad, H. and Madhani, H.D. 2010, Coordinate control of gene expression noise and interchromosomal interactions in a MAP kinase pathway, *Nat. Cell Biol.*, **12**, 954–62.
72. Chubb, J.R., Boyle, S., Perry, P. and Bickmore, W.A. 2002, Chromatin motion is constrained by association with nuclear compartments in human cells, *Curr. Biol.*, **12**, 439–45.