Genome urbanization: Clusters of topologically co-regulated genes delineate functional compartments in the genome of *S. cerevisiae*

Maria Tsochatzidou[1], Maria Malliarou[1], Joaquim Roca[2] and Christoforos Nikolaou[1*]

[1] Computational Genomics Group, Department of Biology, University of Crete, Herakleion 70013 Greece

[2] Molecular Biology Institute of Barcelona (IBMB), Spanish National Research Council (CSIC), Barcelona 08028, Spain

Abstract

The eukaryotic genome evolves under the dual constraint of maintaining co-ordinated gene transcription and performing effective DNA replication and cell division, the coupling of which brings about inevitable DNA topological tension. This is resolved and in some cases even harnessed by the genome through the function of DNA topoisomerases, as has been shown in the concurrent transcriptional activation and suppression of genes upon transient deactivation of topoisomerase II (topoII). The scope of this work is to identify extended genomic domains with similar response to topological stress and to study their structural and functional properties. By analyzing a genome wide run-on experiment upon thermal inactivation of topoII in *S. cerevisiae* we were able to define 116 gene clusters of consistent response (either positive or negative) to topological stress. A comprehensive analysis of these topologically co-regulated gene clusters revealed pronounced preferences regarding their functional, regulatory and structural attributes. Our findings point towards a particular genome compartmentalization, according to which genes that negatively respond to topological stress, are positioned in gene-dense pericentromeric regions, are more conserved and associated to essential functions, while up-regulated gene clusters are preferentially located in the gene-sparse nuclear periphery, associated with secondary functions and under complex regulatory control. This multi-faceted “division of labour” is much resembling a “genome urbanization” process. We propose that genome architecture evolves with a core of essential genes occupying a compact genomic “old town”, whereas more recently acquired, condition-specific genes tend to be located in a more spacious “suburban” genomic periphery.

keywords: genome architecture, DNA topology, Saccharomyces cerevisiae, topoisomerase II, genomic Run-on

**Correspondence to: cnikolaou@biology.uoc.gr*

Significance

In all eukaryotes, the relative positions of genes are constrained by the need for complex transcriptional regulation and effective DNA replication, both of which lead to the accumulation of DNA supercoiling. Here, we perform a concise analysis of the genome architecture of *S. cerevisiae*, by examining the way genes respond to the inactivation of topoisomeraseII. We uncover a fundamental functional compartmentalization, according to which, conserved, essential genes are more prone to topological stress and localize in gene-dense chromatin in the center of the nucleus, contrary to stress-responsive genes, occupying the nuclear periphery, where broader intergenic regions may propel transcription by harnessing topological tension. Our findings suggest a vital role of DNA topological constraints in the evolution of eukaryote genome architecture.

Introduction

The distribution of genes in the genome of eukaryotes is highly non-random. Early genome-wide transcriptome analyses showed the expression of genes to correlate with their linear order along the genome (1). Although it was later shown that this was due to the clustering of constitutive genes (2), such spatial associations have since been used to provide the theoretical framework for links between gene expression and chromatin structure (3) and the inference of protein-protein interaction patterns (4). Non-random gene distribution is also evident in the functional enrichments of gene neighborhoods, with functionally related genes being found in linear proximity more often than expected by chance (5, 6).

The selective pressures underlying the localization of genes are thus of unequal intensity and diverse nature and a number of seemingly irrelevant characteristics may shape the overall genome architecture in evolution (7). Among those DNA supercoiling plays a prominent role. The structure of the eukaryotic nucleus is affected by a number of processes such as DNA replication, RNA transcription and the constant ebb and flow of gene activation and repression. These processes are imposing topological constraints in the form of supercoiling, both types of which (positive and negative) may be found in localized areas of the human genome (8). It was recently shown that such structurally-defined areas may form part of extended “supercoiling domains”, where chromatin conformation correlates with the density of topoisomerases I and II (9). The connection between topological attributes and gene expression appears to be so strong, that in *Drosophila melanogaster*, regions of negative supercoiling, created through the inhibition of topoisomerase I, have increased nucleosome turnover and recruitment of RNA-PolII molecules positively correlating with transcription levels (10). Accumulated positive supercoiling, on the other hand, precludes the formation of transcription initiation complexes (11, 12), a fact indicative of the association between topological constraints and gene expression.

In the budding yeast (*Saccharomyces cerevisiae)*, the organization of genes in linear space has also been attributed to common regulatory mechanisms (13). Yeast's distinguishing feature is the overall gene density, with genes covering ~70% of the total genome (14). Despite its reduced size of only 12Mbp, the transcription dynamics of the yeast genome is highly complex, with genes being expressed in tandem and in operon-like transcripts, with varying sizes of gene upstream and downstream regions (15). Transcription directionality in such a highly streamlined genome also plays a crucial role in the regulatory process, with a number of bidirectional promoters (16) exerting control over coupled gene pairs. The interplay between DNA structure and gene regulation is manifest in a number of cases where gene expression is modulated through three-dimensional loops formed at gene boundaries (17). Thus, even in a small eukaryotic genome, there is a strong association between gene organization, DNA structure and gene expression. We have recently demonstrated the regulatory role of topoisomerase II (topoII) in this context, through a Genomic transcription Run-On (GRO) experiment (18),

that showed different sets of genes responding in opposite ways to the accumulation of topological stress during transcriptional elongation.

In this work, we sought to investigate how the response to the accumulation of topological stress may extend beyond single gene promoters to affect broader genomic regions. Based on a previously published genome-wide dataset (18) we first defined clusters of genes that are differentially affected by topoII deactivation and then went on to assess the extent of functional and structural preferences within them. We were able to detect intricate associations between DNA topology and the distribution of genes in linear order and to show how the two may be linked to other organizational characteristics such as gene spacing and transcriptional directionality. Our results are suggestive of a subtle dynamics of evolution of genome architecture, which we describe as “Genome Urbanization” and according to which the relative position of genes in the nucleus reflects a broader functional, structural and regulatory compartmentalization.

Results

Non-random Clustering of topologically Co-Regulated Genes

We first sought to define domains with concordant response to DNA topological stress. Starting form our initial dataset of differential GRO values for 5414 yeast protein-coding genes (Supplementary File 1), we constructed gene clusters on the basis of genes with similar response to topological tension being found in adjacent positions more often than expect by chance, in a way similar to the one described in (7)(Figure 1A, see Methods).

In total there were 116 clusters with more than 7 genes and 180 clusters containing 6 or more genes. In order to assess the significance of the observed clustering, we implemented a bootstrapping approach upon a randomization process that consisted of 1000 permutations of our initial dataset (see Methods). Figure 1C shows the mean distribution of gene numbers in clusters for these 1000 permutations along their standard deviation. We found that the observed number of clusters with 6 or more genes had a bootstrap p-value of 0.043, while for clusters with >=7 genes this was 0.0008. Of these significantly long (>=7 genes) clusters, 50 comprised exclusively up-regulated genes and 66 exclusively down-regulated ones (median number of genes=8 for both types, Supplementary File 2). Based on the way they were defined, we chose to refer to them as “Topologically co-regulated gene clusters” (TCGC) and went on to characterize them in terms of various properties.

Positional Preferences of Topologically Co-regulated Gene Clusters

The distribution of TCGC (Figure 1D) suggests a non-random localization along the

genome. Up-regulated (red) gene clusters tend to be found towards the outer boundaries of linear chromosomes, while down-regulated ones (blue) show a tendency for their center, often in close proximity to the centromeres. In some cases, clusters appear to assemble in super-clusters as in the case of the right arm of chromosome 12 or the left arms of chromosomes 6 and 7. A straight-forward analysis of TCGC distance from the centromeres showed statistically significant opposing preferences for the up- and down-regulated gene clusters to be located away from and close to centromeres respectively ($p<=0.05$, Supplementary Figure 2).

TCGC formation and centromere distances reflect only one-dimensional tendencies, whereas the eukaryotic nucleus is organized in three-dimensions, where chromosomes interact in space forming inter-chromosomal domains (19). In order to gain insight into possible higher-level positional preferences, we performed an enrichment analysis of clusters occurring in specific three-dimensional domains of the yeast genome as described in (20) and analyzed in chromosomal networks in (21). We found down-regulated TCGC to be preferentially located in the center of the nucleus, described in the model of (21) as an extensive "community" of pericentromeric interchromosomal interactions. Up-regulated ones, on the other hand, were mostly found enriched in the periphery, which is constituted by the subtelomeric regions and the right arm of chromosome 12 (Supplementary Figure 3).

Opposing Functional and Regulatory Preferences in different types of TCGC

TopoII is essential for yeast cells and its prolonged deactivation is bound to cause a general shutdown of cellular activity. The fact, however, that a significant proportion of yeast genes respond to its transient deactivation with increased transcription levels indicates the existence of a positive effect for a subset of cellular functions. We performed a functional enrichment analysis for genes belonging in the up- and down-regulated clusters separately. A functional enrichment analysis at the level of Gene Ontology (Figure 2A) shows extensive differences between the two types of TCGC, a fact indicative of their nuclear compartmentalization being echoed in their functional roles. Three main clusters are apparent: a) Functions enriched in both types of clusters include secondary metabolism and DNA maintenance. b) GO terms that are enriched in up-regulated clusters and depleted in down-regulated ones, represent functions related to cellular transport, the metabolism of co-factors and general stress response. c) Down-regulated-specific GO terms contain basic cellular functions associated with RNA transcription and processing, translation and the nuclear environment. A general pattern suggests that the localization of clusters among chromosomes is also reflected in their functions with up-regulated gene clusters being mostly enriched in peripheral functions, unrelated to the core nuclear processes, the opposite being the main characteristic of genes within down-regulated clusters.

In order to investigate differences in the regulatory potential of up- and down-regulated gene clusters, we obtained the full set of genomic coordinates for conserved

Transcription Factor Bind Sites (TFBS) of 102 different transcriptional regulators as originally compiled in (22) and went on to assess TFBS enrichment in the two types of TCGC. Enrichments were calculated as logarithms of observed over expected ratios, taking into consideration the number of TFBS for each transcription factor and the length of the gene clusters analyzed, while statistical significance was assessed through 1000 random permutations of the TFBS coordinates (Supplementary Methods). Figure 2B highlights the transcription factors whose binding sites are found more or less frequently than expected by chance for both types of TCGC. Down-regulated gene clusters tend to be mostly depleted of TFBS, partly explained by the fact that they are enriched in constitutively expressed genes and thus subject to less complex regulation. From a previous analysis at the level of genes on the same dataset we know down-regulated genes to be enriched in essential functions, with constant expression levels and mostly depleted of TATA-boxes (18). Up-regulated TCGC, on the other hand, show the exact opposite pattern, with the great majority of the TFBS being enriched, a fact indicative of more complex regulation, with significant enrichments for factors related to chromatin structure, DNA surveillance and amino acid transport (see Supplementary Information).

The general preferences that arise paint a greater picture regarding the organization of genes in topologically co-regulated clusters, one that has up-regulated and down-regulated genes not only occupying distinct areas of the nucleus but also being assigned with different tasks. This positional-functional compartmentalization is also reflected on a number of structural attributes of these clusters, discussed in the following.

Gene Spacing and Directionality of Transcription in TCGC

During transcription, DNA torsional stress accumulates with different sign ahead of and behind the gene's transcription start site. This makes the size of both the gene and the preceding intergenic spacer, as well as the relative direction of transcription in relation to adjacent genes highly relevant for the dissipation of topological tension. The effect of topoII deactivation has been shown to be generally independent from the size of the majority of yeast genes (23), but it is strongly inhibitory in the case of long transcripts (24). The situation is very different when one looks, instead, into the surrounding intergenic space. When we ranked the complete set of yeast genes according to their GRO values and plotted them against the mean size of both intergenic spacers separating each gene from the previous and the next as seen on the linear chromosome we found a a clear positive correlation (p-value<=10^{-12}) between the log-size of the intergenic regions and the GRO value, which is highly indicative of transcription-induced topological stress being more readily dissipated in genes with long upstream (and downstream) regions (Supplementary Figure 2).

The association between DNA topology and structural genomic features is expected to be more pronounced in the series of adjacent genes with similar GRO values. In order to study the effect of intergenic space in co-regulated gene clusters, we employed a more

relaxed criterion in the definition. We thus obtained all possible arrays of 7 contiguous genes, ranked them according to their mean GRO value and kept the top and bottom 200 non-overlapping such arrays as up-regulated and down-regulated clusters. These contained the complete set of our TCGC but also a number of additional gene clusters that showed consistent behaviour in their response to topological stress, although not entirely positive or negative in terms of GRO value. We then expanded these clusters on either side in order to comprise 11 genes each (see Methods for details) and compared the average intergenic space along them as shown in Figure 3A. Up-regulated clusters showed intergenic regions of significantly increased size compared to the genomic average (which is about 660bp), an increase that, moreover, appeared to be inflated towards the central genes in the cluster. Genes in down-regulated clusters were, on the other hand, flanked by much shorter intergenic regions and did so consistently, with little fluctuation. Besides their reduced potential for resolving topological stress, shorter intergenic regions provide shorter available genomic space for transcription factors, which may account for the marked under-representation of TFBS in down-regulated gene clusters (Supplementary Figure 5B).

Figure 3A is strongly indicative of the impact of genomic architecture on the maintenance of topological equilibrium in the nucleus. Genes flanked by shorter intergenic spacers will be more prone to the accumulation of supercoiling on either side of the transcription bubble and are therefore expected to be more sensitive to the lack of topoII, while genes that allow for the dissipation of topological strain into longer, untranscribed, nearby regions are predictably more resilient. This dependence, already evident at the level of individual genes, is further accentuated in gene clusters, which points to the existence of synergistic effects between nearby genes in the resolution of topological stress.

Such synergistic effects may be accentuated by the directionality of transcription of consecutive genes. Gene clusters with more “streamlined” directionality patterns are expected to be able to accommodate DNA supercoiling in a more effective manner, using alternating positive and negative supercoiling to “propel” transcription. In order to test this hypothesis, we searched our gene cluster dataset for specific patterns of gene directionality. We split clusters in three categories depending on whether the central gene in the cluster a) formed part of a series of co-directional transcriptional units or b) was belonging to a pair of divergently or c) convergently transcribed genes. We then compared the GRO values of the central gene in each category. The results, shown in Figure 3B, are indicative of a mild, yet significant association between gene directionality patterns and response to topoII deactivation. Genes lying midway in clusters of co-directional transcription have in general higher GRO values, while genes belonging to convergent pairs have difficulty in dealing with topological tension. Divergently transcribed genes lie somewhere in the middle in terms of sensitivity as reflected in their average GRO values. By calculating a simple index of gene directionality changes within a cluster, we found co-directionality to be a general, quantifiable characteristic of TCGC. The proportion of changes in the transcription direction of genes in a cluster is higher in down-regulated gene clusters, contrary to the mean size of co-directional gene runs, which is higher in up-regulated ones

(Supplementary Figure 6). We thus see that different structural attributes in terms of relative gene distances and directionality may provide a mechanistic framework for their coordinated response to topoII deactivation.

Different Conservation Constraints in TCGC

In order to investigate how the properties described above may be constrained through evolution, we performed an analysis of conservation at two levels. First, we analyzed the mean sequence conservation per cluster as aggregate phastCons scores (25), obtained from a genome-wide alignment of six *Saccharomyces* species (26). Average sequence conservation (excluding intergenic space) was negatively correlated with the mean GRO value for the 116 TCGC ($p<=0.01$), confirming that down-regulated clusters are significantly more constrained in terms of sequence conservation (Supplementary Figure 5A). Increased conservation for down-regulated gene clusters doesn't come as a surprise given their functional preferences described in previous sections. Genes in up-regulated clusters on the other hand appear to be under more moderate sequence constraint, a fact which could be indicative of their less essential role, or their more recent acquisition through gene duplication (27).

We next turned to more complex conservational features that also take into account synteny relationships, reflected upon the position and transcriptional direction of genes in related species. We made use of data from the Yeast Gene Order Browser (YGOB; http://wolfe.gen.tcd.ie/ygob) (28) that contains a detailed catalog of orthologous genes between a number of yeast species. We collected all orthologous gene pairs between *S. cerevisiae* and two of its closest species in the *sensu stricto* complex, *S. paradoxus* and *S. mikatae*. We analyzed them separately for up- and down-regulated clusters by calculating a simple measure of "directional conservation", as described in Methods. Given that syntenic regions are by definition under sequence constraint we were not surprised to see that genes in down-regulated clusters were characterized by both high sequence and directional conservation as may be seen in Figure 3C. What was rather interesting was the corresponding position of genes in up-regulated clusters in the same two-dimensional constraint space. While we already knew that sequence constraints were more relaxed in these regions, we found a significant proportion of genes with high values of directional conservation, suggesting that up-regulated gene clusters tend to maintain the directionality patterns even under milder sequence constraints. It thus seems, that keeping a co-directional gene layout confers a relative advantage to genomic regions that are otherwise less conserved in terms of sequence.

Discussion

The existence of clusters of topologically co-regulated genes (TCGC) implies that

eukaryotic genes may be synergistically orchestrated in gene neighborhoods with particular characteristics. By persistently analyzing the defined topologically co-regulated gene clusters at various levels, we were able to outline a general overarching pattern, according to which the yeast genome may be broadly divided in two compartments that have, in time, assumed radically different architectures and operational roles. These two distinct "territories" were effectively traced by the way genes responded to topological strain. On one hand, we found clusters of down-regulated genes to be preferentially located towards the centromeric part of chromosomes, occupying the center of the nucleus. These consist of highly conserved genes associated with essential functions that are predominantly located in close proximity to each other and with highly variable patterns of gene directionality. In striking contrast, clusters of up-regulated genes are prevalent in the nuclear periphery, enriched towards the telomeres and under more relaxed sequence constraints as they are associated with secondary functions. Contrary to their down-regulated counterparts, these clusters preferentially contain co-directionally transcribed genes that are also separated by long interegenic spacers. Their predominantly stress-responsive functional roles are likely the reason for their up-regulation under topological stress, while their structural organization may enable them to harness DNA supercoiling to achieve increased transcription levels.

Such compelling disparity at all studied levels points towards a general pattern of genome architecture. This very much resembles an urbanization process, that has over evolution demarcated an "old-town" at the centromeric part of the nucleus, formed by tightly crammed ancient genes and a "suburban genome" at the chromosomal outskirts, where newly acquired genes occupy greater spaces with an ordered directionality that resembles tract housing (Figure 4). This "Genome Urbanization" is echoed in various genomic features that we have discussed in the context of TCGC. When looking at the sequence conservation of genes as a function of their distance from the centromere we find a weak negative correlation, with the 5% most distant genes being significantly less conserved than the 5% most proximal (n=638, t.test p-value=0.005). Similar discrepancy is observed when looking at the intergenic space length (n=508, t.test p-value<10^{-6}). It thus appears that the division of the genome in domains with specific "architectural" characteristics may well extend beyond DNA topology. Our findings indicate that the Genome Urbanization scheme is likely a general feature, that allows the nucleus to dissipate DNA topological stress more effectively, but whose functions are likely to extend to gene functionality (29), regulation programs (16, 30) and genome evolution (31).

A particularly important element to consider is that of transcriptional plasticity. The over-representation of stress responsive genes in up-regulated clusters points towards an organization of the genome, in which genes that need to readily modulate their expression levels according to environmental conditions are positioned in specific areas of the genome. Recent works have provided interesting links between plasticity and genomic features that resemble the ones we find to be hallmarks of the "suburban genome", namely non-essentiality, complex regulation and gene duplication (32). The size of the intergenic space between genes has also been shown to widely shape

expression variability (33).

The concept of "Genome Urbanization" may extend to more complex eukaryotes, albeit not in a straight-forward manner. The size, gene density and evolutionary dynamics of the unicellular *S. cerevisiae* make the delineation of domains more clear-cut, while the complexity of gene-sparse genomes from multicellular organisms with the requirements for spatio-temporal expression patterns is bound to be reflected upon a more entangled genome architecture (34). The advent of new experimental approaches for the study of genome conformation in three dimensions provides a solid framework for testable hypotheses that will deepen our understanding of the evolution of genome organization.

Methods

GRO data

Data were obtained from a genome-wide Genomic Run-On (GRO) experiment conducted in triplicates on a yeast strain lacking topoisomerase I and carrying a thermosensitive topoisomerase II (JCW28 - top1Δ, top2ts). GRO was conducted as described in (35) and data were analyzed as previously described in (18).

Gene Clustering

Arrays of adjacent genes with the same sign of GRO values were joined in clusters that encompassed the genomic segment from the farthest upstream to the farthest downstream gene. Clusters of >=7 genes were selected on the basis of a bootstrapping analysis as suggested in (7). This was performed by conducting 10000 random permutations of gene order while keeping the same GRO values. We used functions from the BedTools Suite (36) to control for unaltered gene sizes and chromosomal distributions. Gene number distributions of the derived clusters were calculated alongside the mean values and standard deviation of number of clusters for the 10000 random gene sets. We then compared the observed values with the expected under randomness asking that the observed value be at least greater than the mean of the 10000 permutations by two standard deviations. Clusters with >=7 genes occurred in less than 0.1% of the simulations (bootstrap value p=0.0008) and were divided into up-regulated and down-regulated, depending on the mean GRO value of all genes in each cluster (Supplementary File 2).

Gene and intergenic space size and direction of transcription

Each chromosome was scanned in overlapping 11-gene windows and for each step we recorded: the full list of 11 GRO values, mean GRO value of the central 7 genes and gene

lengths and mean intergenic space lengths for all genes. The top/bottom 200 non-overlapping clusters in terms of mean GRO value were analyzed at the level of gene and intergenic spacer lengths (Figure 3A). We used the same list to obtain patterns of gene directionality as arrays of seven genes (Figure 3B). GRO values of the central gene were analyzed for three characteristic patterns corresponding to a) co-directional genes (central 5 genes transcribed in the same direction) b) the central gene being a member of a divergent or c) a convergent gene pair.

Gene Directionality Conservation Index

We obtained orthologous gene coordinates for *S. paradoxus* and *S. mikatae* from the Yeast Gene Order Browser (YGOB) (28). For each genomic region of *S. cerevisiae* we calculated the ratio of genes retaining their position and direction of transcription in the other two species. A value of 1/N, N being the number of genes in the region, was added to the score if both the gene's position and direction was maintained in the other two species. This led to measure of directionality conservation on a scale of 0 (no retention of direction) to 1 (absolute retention of direction). The contour map of Figure 3C was formed by splitting the two-dimensional space in a 10x10 grid and assigning each bin with the proportion of clusters falling in the corresponding sequence/direction conservation value range (bins of 0.1 for each). The final value assigned to each of the 10x10 bin was the log2(ratio) of up/down-regulated cluster frequency. Values >0 corresponded to an enrichment of up- and values < 0 to an enrichment of down-regulated clusters.

Data Availability

Raw data are deposited as GEO database (http://www.ncbi.nlm.nih.gov/geo) with accession number GSE16673, while the relative GRO values for a set of 5414 genes were mapped on chromosomal coordinates for the SGD/saCcer1 version (Oct. 2003) (Supplementary File 1). In-house scripts for the analysis of data are available upon request.

References

1. Caron H, et al. (2001) The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* 291(5507):1289–92.
2. Lercher MJ, Urrutia AO, Hurst LD (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet* 31(2):180–3.
3. Batada N, Hurst L (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39(8):945–949.
4. Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: A fingerprint of proteins that physically interact. *Trends Biochem Sci* 23(9):324–328.
5. Lee JM, Sonnhammer ELL (2003) Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res* 13(5):875–882.
6. Tiirikka T, Siermala M, Vihinen M (2014) Clustering of gene ontology terms in genomes. *Gene* 550(2):155–164.
7. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5(4):299–310.
8. Ljungman M, Hanawalt PC (1992) Localized torsional tension in the DNA of human cells. *Proc Natl Acad Sci U S A* 89(13):6055–6059.
9. Naughton C, et al. (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* 20(3):387–95.
10. Teves SS, Henikoff S (2014) Transcription-generated torsional stress destabilizes nucleosomes. *Nat Struct Mol Biol* 21(1):88–94.
11. Roca J (2011) Transcriptional inhibition by DNA torsional stress. *Transcription* 2(2):82–85.
12. Joshi RS, Piña B, Roca J (2010) Positional dependence of transcriptional inhibition by DNA torsional stress in yeast chromosomes. *EMBO J* 29(4):740–8.
13. Kruglyak S, Tang H (2000) Regulation of adjacent yeast genes. *Trends Genet* 16(3):109–111.
14. Goffeau A, et al. (1996) Life with 6000 genes. *Science (80-)* 274(5287):546–567.
15. David L, et al. (2006) A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci U S A* 103(14):5320–5.
16. Xu Z, et al. (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457(7232):1033–7.
17. Tan-Wong SM, et al. (2012) Gene loops enhance transcriptional directionality. *Science* 338(6107):671–5.

18. Nikolaou C, et al. (2013) Topoisomerase II regulates yeast genes with singular chromatin architectures. *Nucleic Acids Res* 41(20):9243–56.

19. Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. *Mol Cell* 49(5):773–82.

20. Duan Z, et al. (2010) A three-dimensional model of the yeast genome. *Nature* 465(7296):363–7.

21. Hoang S a, Bekiranov S (2013) The network architecture of the Saccharomyces cerevisiae genome. *PLoS One* 8(12):e81972.

22. Harbison C, Gordon D, Lee T, Rinaldi N (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*:99–104.

23. Pedersen JM, et al. (2012) DNA Topoisomerases Maintain Promoters in a State Competent for Transcriptional Activation in Saccharomyces cerevisiae. *PLoS Genet* 8(12). doi:10.1371/journal.pgen.1003128.

24. Joshi RS, Piña B, Roca J (2012) Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic Acids Res*:gks626–.

25. Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034–50.

26. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423(6937):241–54.

27. Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B (2009) Evolution of Gene Order in the Genomes of Two Related Yeast Species Evolution of Gene Order in the Genomes of Two Related Yeast Species. 2009–2019.

28. Byrne KP, Wolfe KH (2005) The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15(10):1456–61.

29. Gehlen LR, et al. (2012) Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus* 3(October 2015):370–383.

30. Zhu J, et al. (2008) Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 40(7):854–61.

31. Sugino RP, Innan H (2012) Natural selection on gene order in the genome reorganization process after whole-genome duplication of yeast. *Mol Biol Evol* 29(1):71–9.

32. Lehner B (2010) Conflict between noise and plasticity in yeast. *PLoS Genet* 6(11):e1001185.

33. Bajić D, et al. (2012) Balancing noise and plasticity in eukaryotic gene expression.

BMC Genomics 13(1):343.

34. Bagadia M, Singh A, Sandhu KS (2016) Three dimensional organization of genome might have guided the dynamics of gene order evolution in eukaryotes. *Genome Biol Evol* 8(3):evw050.

35. García-Martínez J, Aranda A, Pérez-Ortín JE (2004) Genomic Run-On Evaluates Transcription Rates for All Yeast Genes and Identifies Gene Regulatory Mechanisms. *Mol Cell* 15(2):303–313.

36. Quinlan A, Hall I (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841.

37. Chouvardas P, Kollias G, Nikolaou C (2016) Inferring active regulatory networks from gene expression data using a combination of prior knowledge and enrichment analysis. *BMC Bioinformatics* 17 Suppl 5:181.

38. Taddei A, Gasser SM (2012) Structure and function in the budding yeast nucleus. *Genetics* 192(1):107–29.

Figure Legends

Figure 1

A. Schematic representation of cluster calls. Contiguous genes with similar (positive or negative) GRO values were joined in gene clusters, which were defined as the genomic region spanning the chromosomal space from the fartherst upstream to the farthest downstream gene boundary.

B. Location of genes and clusters with GRO values in part of chromosome 4.

C. Distribution of number of genes in clusters. Real clusters (green) show a skewed distribution towards larger sizes as compared to the mean of 10000 random permutation of GRO values. Differences are significant for gene numbers >=6.

D. Distribution of 116 topologically co-regulated gene clusters (TCGC) in the yeast genome.

Figure 2

A. GO term enrichment heatmap of TCGC of both types. Enrichments were calculated based on a modified Gene Set Enrichment Analysis (37). Only GO terms with an adjusted p-value<=0.05 (at 5% FDR) for at least one of the two TCGC types are reported.

B. Volcano plot showing enrichments of transcription factor binding sites for 102 different transcriptional regulators compiled by (22). Enrichments are shown as log2 based observed/expected ratios. Values >0 indicate enrichment and values <0 indicate depletion (see Methods). P-values correspond to 1000 bootstraps for each transcriptional regulator.

Figure 3

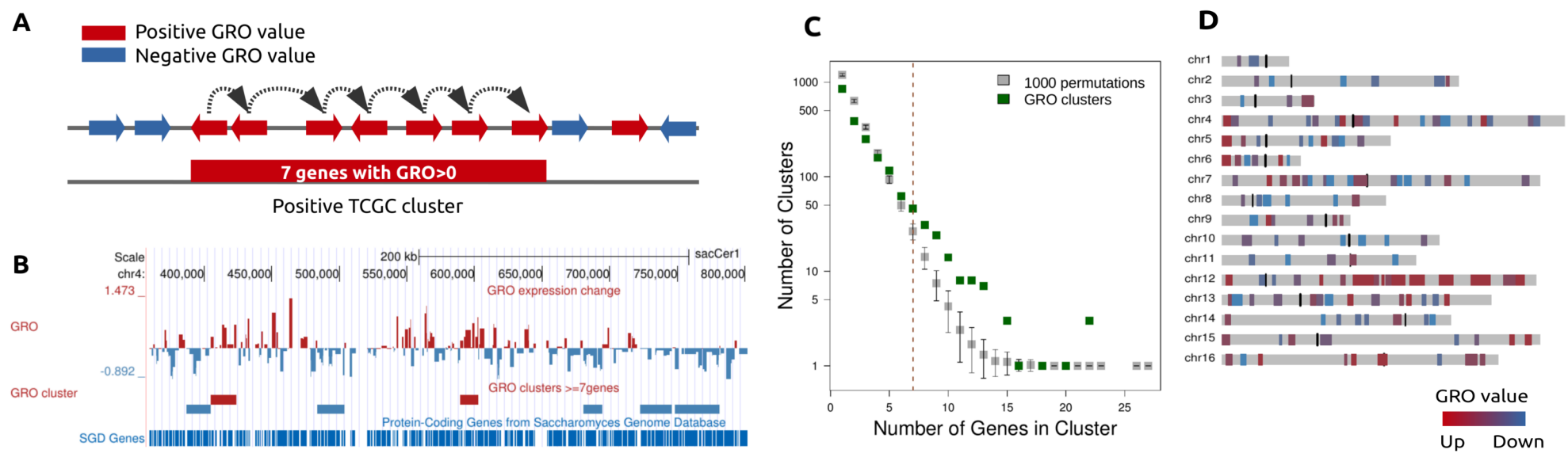
A. Top, mean intergenic region length for clusters of 11 consecutive genes. Each line corresponds to the mean values calculated for the top/bottom 200 clusters based on the central 7 GRO values (see Methods for details). Bottom, same analysis for gene size. Shaded bands correspond to 95% confidence intervals.

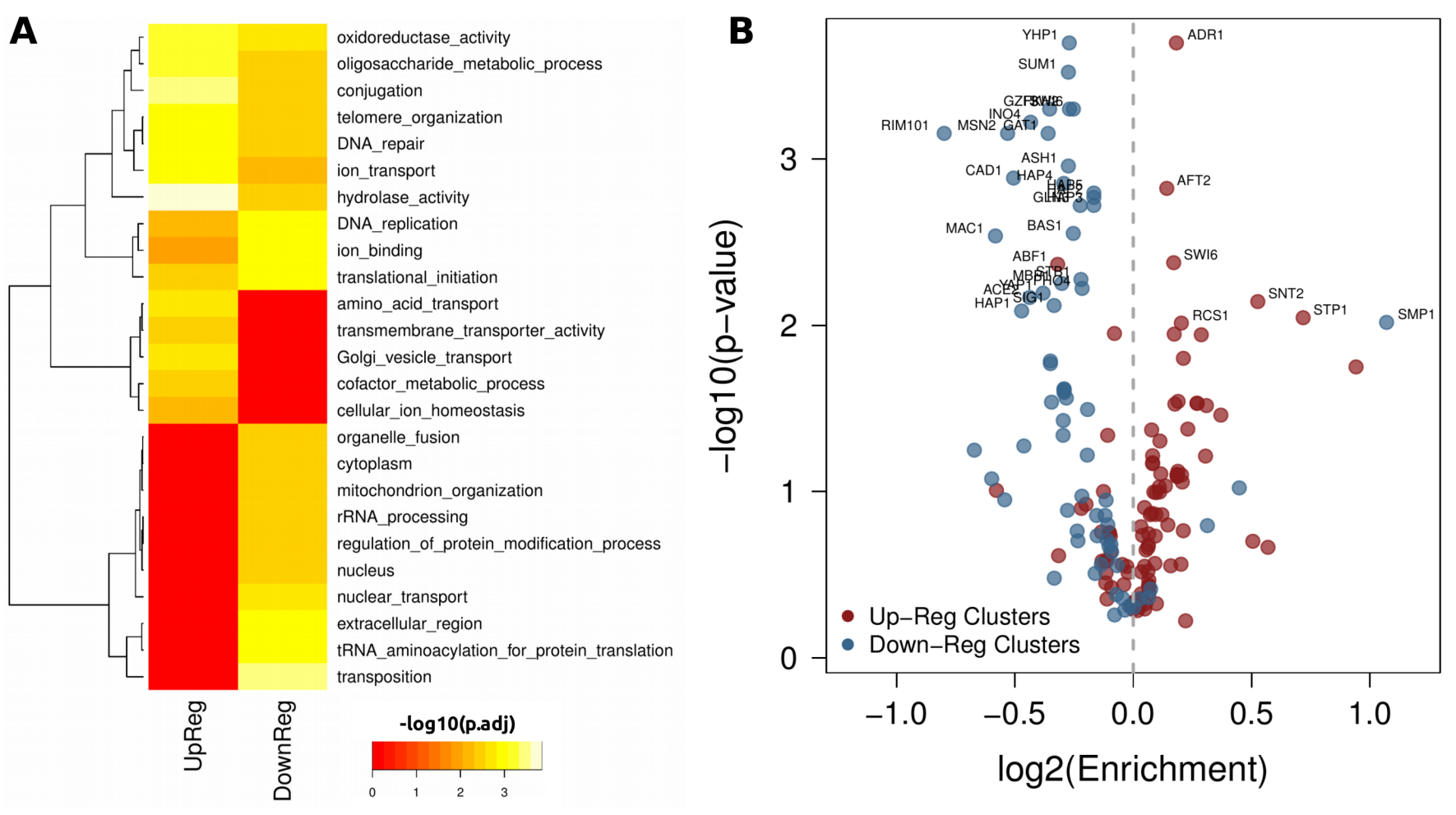
B. Distribution of GRO values of genes lying in the center of 5-cluster genes with different directionality patterns defined on the basis of transcriptional direction (N co-directional=36, N divergent=29, N convergent=25). P-values calculated on the basis of a Mann-Whitney U test.

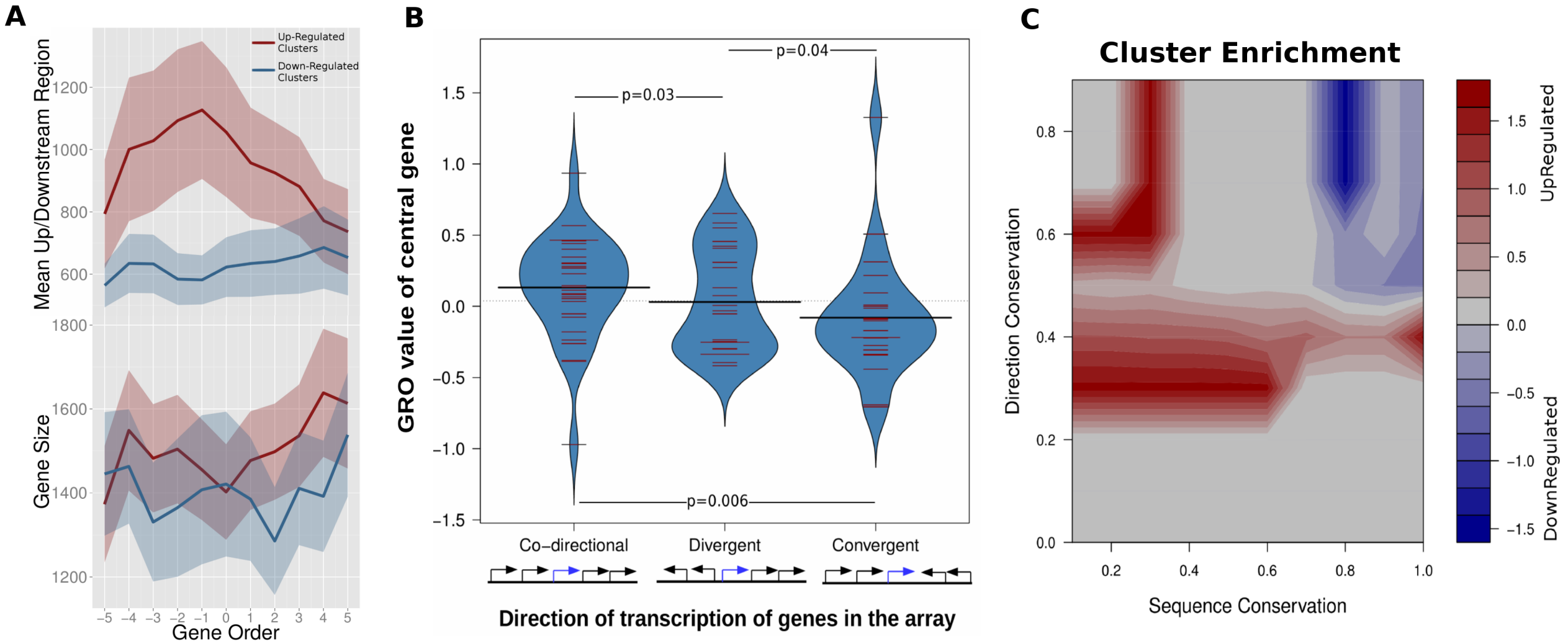
C. Contour heatmap of enrichment of different types of TCGC in areas defined by mean sequence conservation (as above, x-axis) and a transcriptional direction index (y-axis) defined as the proportion of genes retaining relative gene position and directionality in two closely related species. Enrichments were calculated as log2(ratios) of proportion of up-regulated/down-regulated clusters having values in a 10x10 value grid (see Methods).

Figure 4

Genome Urbanization in *S. cerevisiae*. A schematic of the yeast interphase nucleus is shown based on the Rabl configuration (38). Pericentromeric regions correspond to what we call the “Old city center” with enrichment in gene clusters down-regulated under topoII deactivation. The genome in these areas may be compared to the crammed houses of a medieval town separated by narrow, intertwined alleys. Genes in the “old town” are more conserved, associated with essential functions and located within tighter genomic spaces with fewer transcription factor binding sites and entangled directionality. Genomic regions at the nuclear periphery are resembling a “suburban landscape” where more recently acquired (and less conserved) genes are spaced in co-directional operon-like arrays, separated by longer intergenic sequences, reminiscent of the tract housing of modern city suburbia.

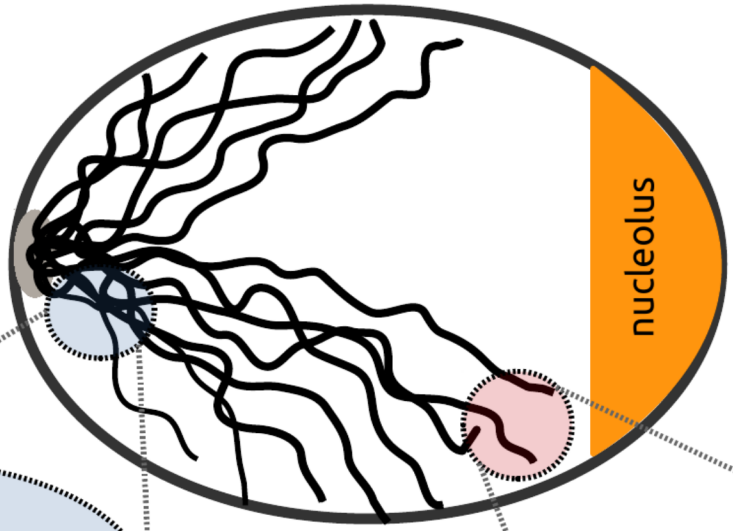

A
Positive GRO value
Negative GRO value
7 genes with GRO>0
Positive TCGC cluster
B
Scale
chr4:
400,000
450,000
500,000
200 kb
550,000
600,000
650,000
700,000
750,000
800,000
sacCer1
1.473
GRO
-0.892
GRO expression change
GRO cluster
GRO clusters >=7genes
Protein-Coding Genes from Saccharomyces Genome Database
SGD Genes
C
1000 permutations
GRO clusters
Number of Clusters
1000
500
100
50
10
5
1
0
5
10
15
20
25
Number of Genes in Cluster
D
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
GRO value
Up
Down


A
oxidoreductase_activity
oligosaccharide_metabolic_process
conjugation
telomere_organization
DNA_repair
ion_transport
hydrolase_activity
DNA_replication
ion_binding
translational_initiation
amino_acid_transport
transmembrane_transporter_activity
Golgi_vesicle_transport
cofactor_metabolic_process
cellular_ion_homeostasis
organelle_fusion
cytoplasm
mitochondrion_organization
rRNA_processing
regulation_of_protein_modification_process
nucleus
nuclear_transport
extracellular_region
tRNA_aminoacylation_for_protein_translation
transposition
UpReg
DownReg
-log10(p.adj)
0
1
2
3
B
YHP1
ADR1
SUM1
RIM101
MSN2
GAT1
INO4
ASH1
CAD1
HAP4
AFT2
MAC1
BAS1
ABF1
SWI6
SNT2
STP1
SMP1
RCS1
HAP1
SIG1
−log10(p−value)
log2(Enrichment)
Up−Reg Clusters
Down−Reg Clusters


A
Up-Regulated Clusters
Down-Regulated Clusters
Mean Up/Downstream Region
1200
1000
800
600
1800
Gene Size
1600
1400
1200
−5 −4 −3 −2 −1 0 1 2 3 4 5
Gene Order
B
p=0.04
p=0.03
p=0.006
GRO value of central gene
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Co−directional
Divergent
Convergent
Direction of transcription of genes in the array
C
Cluster Enrichment
Direction Conservation
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0
Sequence Conservation
UpRegulated
DownRegulated
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

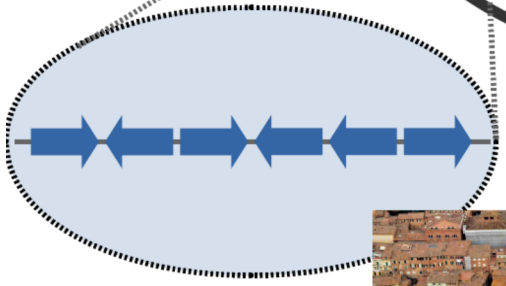
The "Old city center"

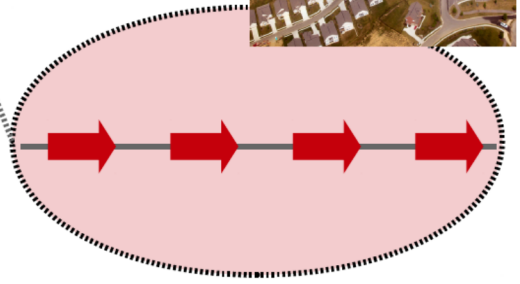
- Ancient, conserved genes
- Constitutive functions
- Simple regulation
- Small intergenic space
- Complex directionality



The "Suburban Genome"

- Newly acquired, duplicated, less conserved genes
- Stress-responsive, secondary functions
- Complex regulation
- Long intergenic space
- "Streamlined" transcription





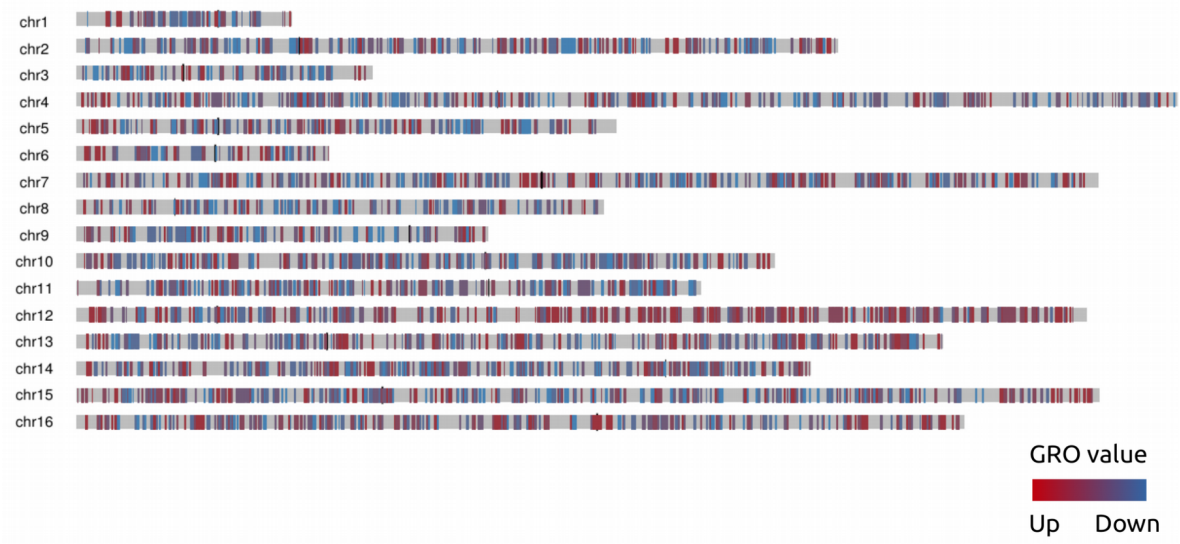
Genome urbanization: Clusters of topologically co-regulated genes delineate functional compartments in the genome of *S. cerevisiae*

Maria Tsochatzidou, Maria Malliarou, Joaquim Roca and Christoforos Nikolaou

Supplementary Information

GRO analysis of 5414 protein coding genes

The relative position of each of the 5414 protein coding genes measured in our GRO experiment was depicted on the corresponding chromosome using a color code for the the GRO value (red: positive, blue: negative). The tendency of genes with similar GRO response to form clusters is obvious for some clear cases in the right arm of chr12 and at the telomeres of most chromosomes.

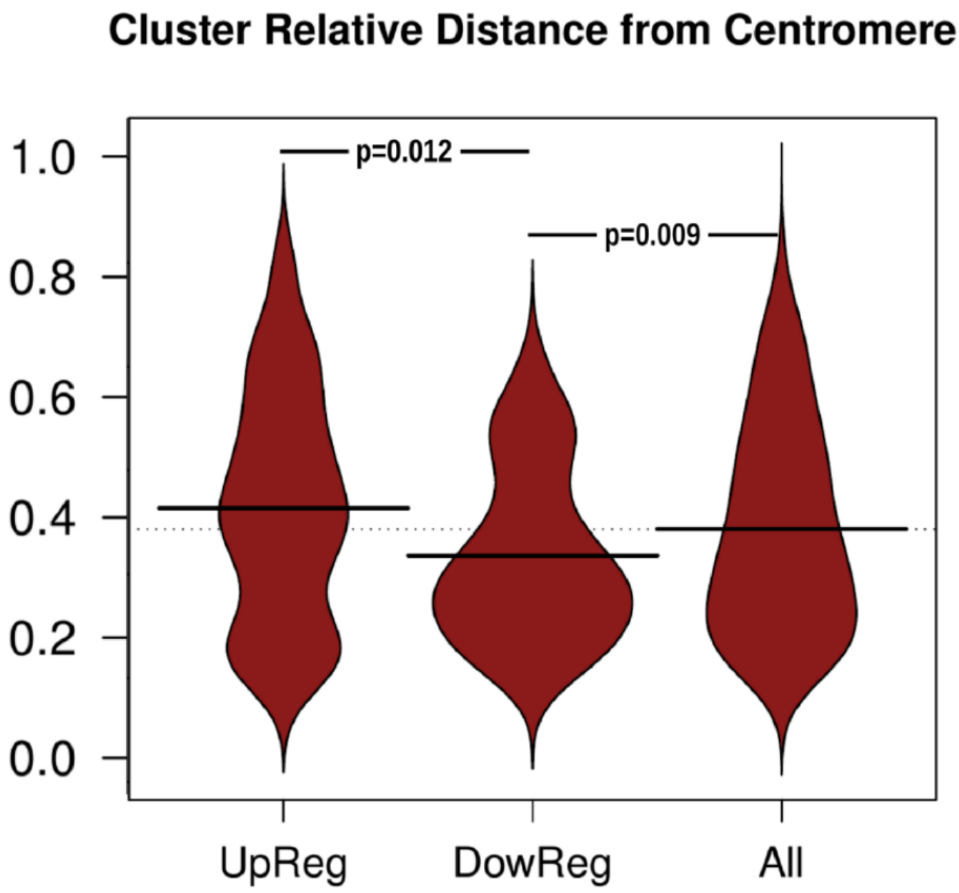


Supplementary Figure 1

Genomic distribution of 5414 genes in the yeast genome according to their GRO values.

Positional Preferences of TCGC

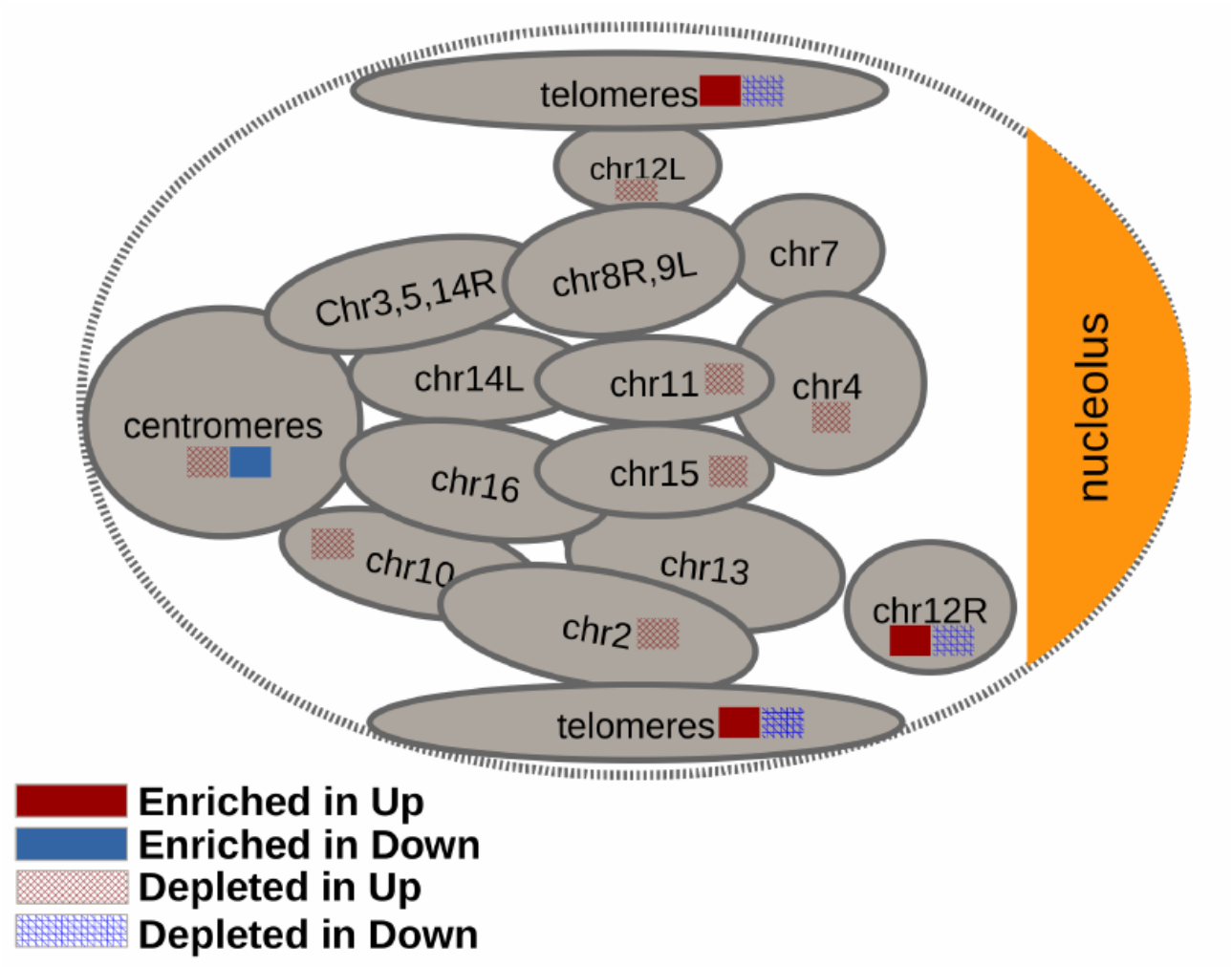
Preferential localization of TCGC was analyzed at linear and three-dimensional levels. Linear preferences were assessed as the relative distance of each cluster from the centromere of the corresponding chromosome (see Supplementary Methods). We found down-regulated TCGC to be preferentially positioned in the pericentromeric regions, with significantly smaller distances from the centromere compared to up-regulated ones which showed an opposite tendency.



Supplementary Figure 2. Topologically co-regulated gene clusters show different positional preferences, with down-regulated clusters being significantly closer to the corresponding chromosome centromeres. Cluster-to-centromere distances were scaled to the corresponding chromosomal arm length as described in Methods. P-values calculated on the basis of a Mann-Whitney U test.

In order to gain insight into possible higher-level positional preferences we performed an enrichment analysis of clusters occurring in specific three-dimensional domains of the yeast genome as described in (1) and analyzed in chromosomal networks by (2). The results recapitulate the already mentioned opposite tendencies for pericentromeric localization. We found down-regulated TCGC to be preferentially located in the center of the nucleus, described in the model of (2) as an extensive “community” of pericentromeric interchromosomal interactions. Up-regulated ones, on the other hand

were mostly found enriched in the periphery as may be seen in the schematic representation of the nucleus in Supplementary Figure 3 (adopted from the authors' model). Another particular nuclear niche for up-Regulated clusters is the right arm of chromosome 12 corresponding to the region downstream of the nucleolus-associated rDNA cluster locus, which is known to act as a barrier for chromosomal interactions (1). This region of chromosome 12 is predicted to exhibit limited connectivity to other regions in terms of chromosomal interactions and is expected to form a rather autonomous domain in the yeast nucleus. It is therefore reasonable to assume that the rDNA locus functions as a chromatin "barrier" that allows areas of the same chromosomes to adopt radically different behaviour. Interestingly, other such barriers to the spread of repressive chromatin have also been associated with the telomeres of budding yeast (3, 4), which fits well with the extensive over-representation of up-regulated clusters towards telomeric regions.



Supplementary Figure 3. Schematic showing the relative enrichments in yeast genome network communities obtained from (2) based on 4C data from (1). Enrichments were calculated as observed/expected ratios of overlaps between TCGC and network community coordinates (see Supplementary Methods). Only statistically significant enrichments or depletions ($p<=0.05$) are reported. Figure is adopted from the network reconstruction by (2).

Distinct regulatory modes in up- and down-regulated TCGC

Analysis of TFBS enrichment showed significant differences between the two types of gene clusters as may be seen in Supplementary Table 1. Down-regulated clusters were only enriched in one factor, SMP1, which is involved in osmotic stress regulation, a process that is associated with DNA topological stress. Most of the significant enrichments of down-regulated genes are negative (under-representations), the most notable of which include the STP1 degron, the chromatin remodeller UME6 and the multidrug resistance-oxidative stress response STB5.

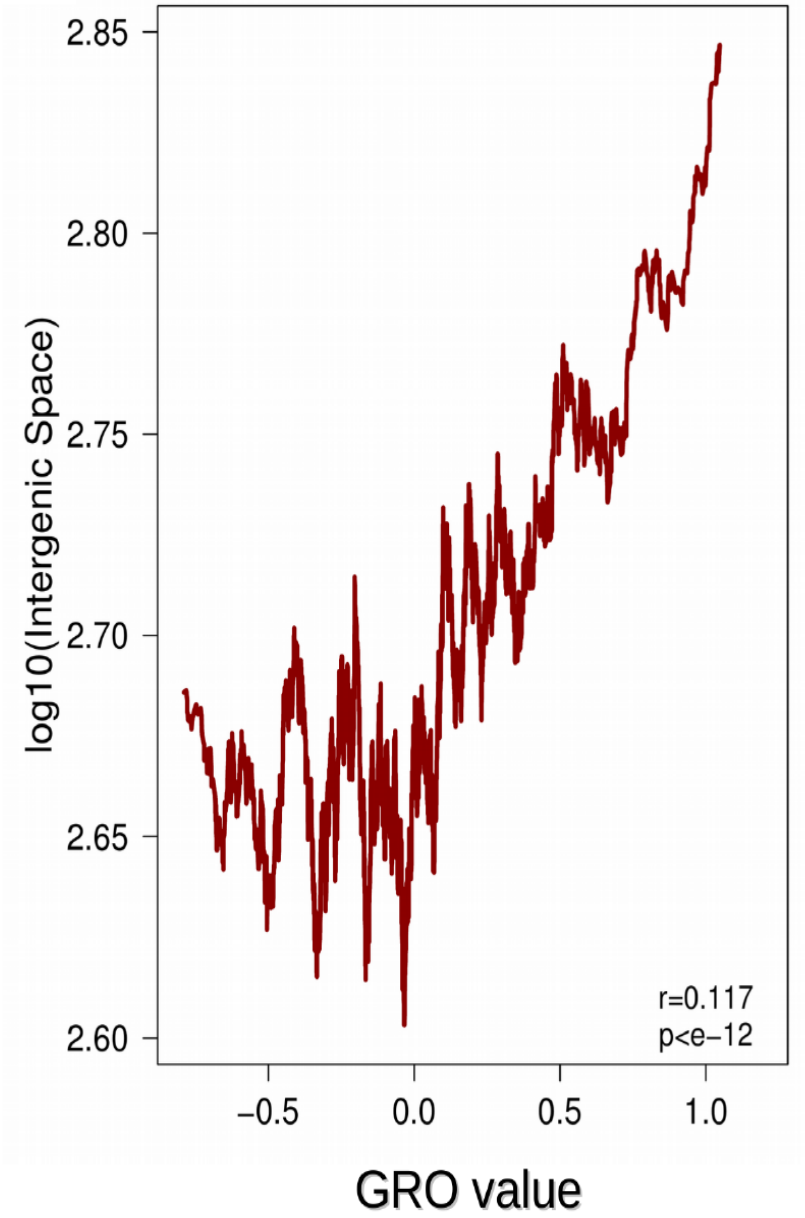
In Up-regulated clusters, significant enrichment was found for factors such as AFT1(RCS1), AFT2 and SWI6, that are involved in chromatin surveillance and STP1 (depleted in down-regulated TCGC) and SNT2, which is a key regulator of the amine and amino-acid transport. The homeostasis of polyamines in particular, has been known to be central for cell growth and to regulate chromatin structure by directly affecting DNA topology and nucleosome stability (5). Up-regulated gene clusters are depleted in binding sites for ABF1, a regulator of DNA replication strongly associated with autonomously replicating sequences (ARS) (6). This could be explained given the fact that accumulation of DNA torsional stress is bound to stall if not shut down the process of DNA replication as a homeostatic mechanism.

TFBS	Enrichment (Up)	p-value(UP)	Enrichment (Down)	p-value (Down)
ABF1	0.802	0.004	1.002	0.498
ACE2	1.020	0.442	0.739	0.007
ADR1	1.135	0.000	0.746	0.000
AFT2	1.103	0.002	0.790	0.000
ARR1	1.137	0.079	0.817	0.025
ASH1	1.079	0.094	0.827	0.001
BAS1	1.084	0.078	0.839	0.003
CAD1	1.065	0.271	0.704	0.001
CIN5	1.007	0.459	0.815	0.000
DAL80	1.004	0.459	0.814	0.000
DAL82	0.928	0.046	0.812	0.000
FKH1	1.028	0.184	0.878	0.000
FKH2	1.046	0.209	0.830	0.001
GAT1	0.933	0.177	0.779	0.001
GCN4	1.024	0.304	0.807	0.000
GCR1	0.922	0.262	0.814	0.046
GLN3	0.968	0.275	0.856	0.002
GZF3	0.935	0.186	0.783	0.001
HAP1	1.047	0.374	0.721	0.008
HAP2	1.058	0.068	0.891	0.002
HAP3	1.058	0.067	0.891	0.002
HAP4	1.068	0.137	0.816	0.001
HAP5	1.058	0.061	0.891	0.002
HSF1	1.046	0.180	0.733	0.000
INO2	1.062	0.101	0.792	0.000
INO4	1.088	0.138	0.740	0.001
MAC1	1.239	0.030	0.668	0.003
MBP1	1.099	0.093	0.811	0.006
MET31	1.205	0.029	0.785	0.017
MET32	1.205	0.030	0.785	0.016
MET4	1.920	0.018	0.947	0.553
MOT3	0.982	0.283	0.866	0.000
MSN2	1.047	0.342	0.692	0.001
MSN4	1.140	0.029	0.701	0.000
PDR3	1.484	0.217	0.000	0.012
PHD1	1.038	0.225	0.785	0.000
PHO4	1.034	0.282	0.861	0.006
RCS1	1.151	0.010	0.722	0.000
REB1	1.005	0.475	0.829	0.000
RGT1	1.081	0.050	0.774	0.000
RIM101	1.236	0.061	0.575	0.001
RLM1	1.292	0.035	0.825	0.130
RLR1	0.917	0.100	0.785	0.000
RPN4	1.046	0.366	0.787	0.029
SIG1	1.023	0.412	0.793	0.008
SKN7	1.158	0.016	0.921	0.139
SKO1	1.140	0.076	0.577	0.000
SMP1	1.419	0.199	2.100	0.010
SNT2	1.440	0.007	0.726	0.053
SOK2	1.034	0.125	0.800	0.000
SPT23	1.023	0.163	0.858	0.000
STB1	1.127	0.011	0.858	0.005
STB5	1.152	0.080	0.538	0.000
STE12	0.985	0.307	0.874	0.000
STP1	1.645	0.009	0.374	0.002
SUM1	1.044	0.218	0.827	0.000
SUT1	1.220	0.011	0.898	0.139
SWI4	1.130	0.030	0.875	0.032
SWI5	1.056	0.135	0.778	0.000
SWI6	1.126	0.004	0.839	0.001
TEC1	1.055	0.043	0.780	0.000
THI2	0.911	0.175	0.822	0.027
TYE7	0.926	0.260	0.815	0.037
UME6	1.047	0.371	0.525	0.000
YAP1	1.173	0.042	0.768	0.006
YAP3	1.137	0.081	0.817	0.025
YAP5	1.137	0.081	0.817	0.024
YAP6	1.137	0.079	0.817	0.024
YAP7	1.051	0.138	0.805	0.000

Supplementary Table 1: TFBS enrichments for 70 transcription factors that were either enriched (Red, Enr>=1.1, p<=0.05) or depleted (Blue Enr<=0.9, p<=0.05) in one of the two TCGC types.

Correlation of GRO value with intergenic space

A clear positive correlation was found when analyzing each gene's GRO value against the mean intergenic space that surrounds it. Intergenic space for each gene was calculated as the arithmetic mean of the gene upstream and downstream regions for each transcriptional unit, starting at the transcription start (TSS) and ending at the transcription termination site (TTS) (see Supplmentary Methods). The results suggest that long intergenic spaces may allow for more effective dissipation of topological stress and may partly explain the existence of up-regulated genes upon topoII inactivation.



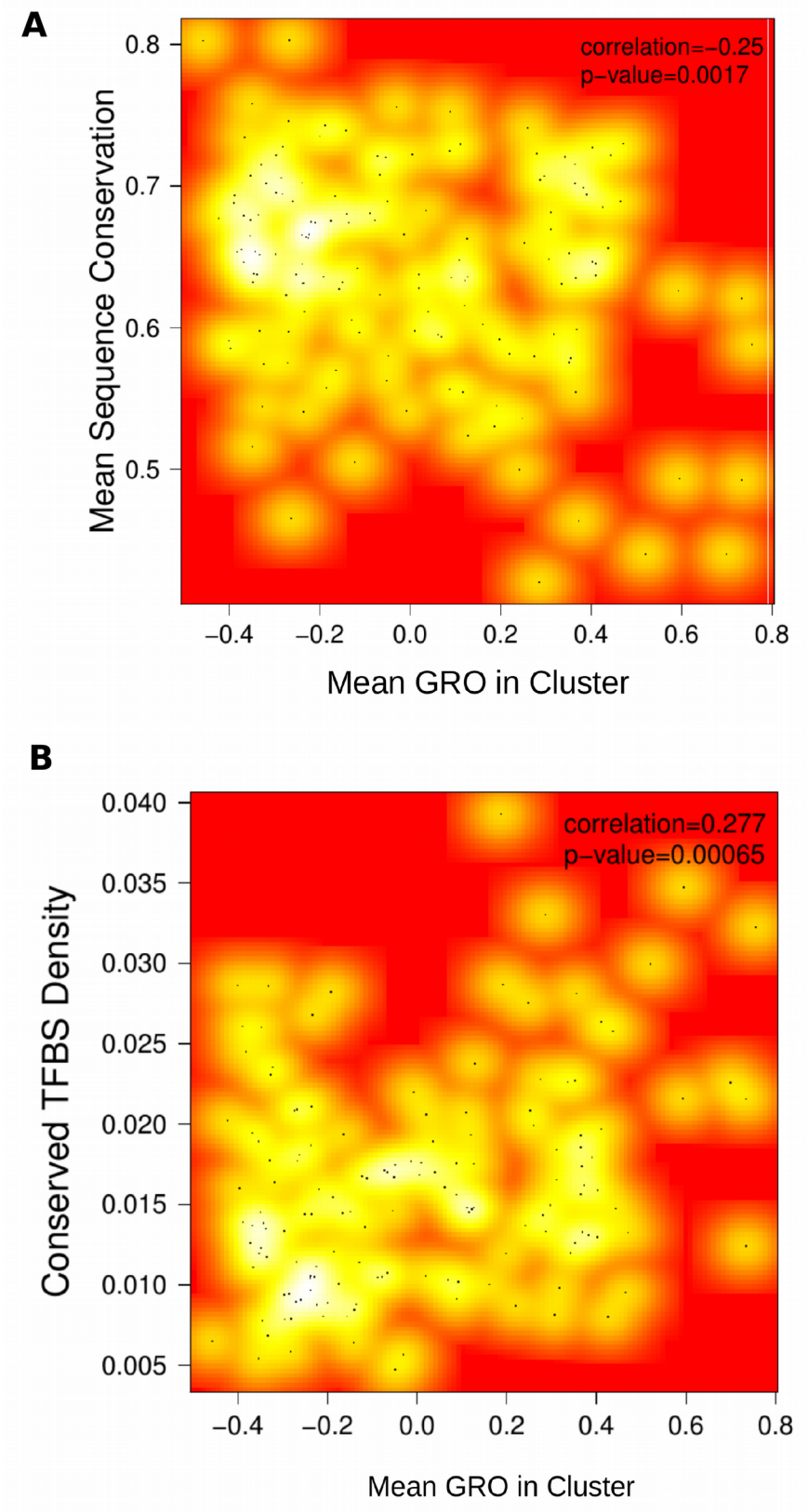
Supplementary Figure 4

Ranked scatterplot of GRO value and logged mean intergenic region length for the complete set of 5414 genes. Intergenic space was calculated as the arithmetic mean of upstream and downstream intergenic regions for each gene (see Methods). Correlation measured as Spearman's rho.

Sequence conservation and TFBS density in TCGC

In order to assess the sequence conservation of TCGC we calculated the mean phastCons score of each cluster (see Supplementary Methods). We then examined how these values correlated with the mean GRO value of each TCGC and found a significant ($p<=0.01$) negative correlation. This tendency was also suggested by examining individual genes in a previous study by our group (7), however, it is when looking at broader chromosomal domains, such as the ones delineated by TCGC that it becomes more pronounced. Increased conservation for down-regulated gene clusters doesn't come as a surprise given their functional preferences described in previous sections. Genes is up-regulated clusters, on the other hand, appear to be under more moderate sequence constraint, a fact which could be indicative of their less essential role, or their translocation as duplicated genes from a different chromosome (8).

Transcription factor binding site (TFBS) density was calculated as the overlap percentage of TFBS (compiled in (9)) against the total length of each cluster. A strong positive correlation ($p<=0.001$) was found when compared with the mean GRO value of each TCGC, a fact that may initially appear counter-intuitive given the low conservation of up-regulated (high GRO) clusters. It may, however, be explained on the basis of two previous observations. First, that up-regulated clusters are predominantly non-essential, stress-responsive genes with more complex regulation patterns and second, because of their clear tendency for longer intergenic spacers, which can accommodate a greater number of transcription factor binding sites.



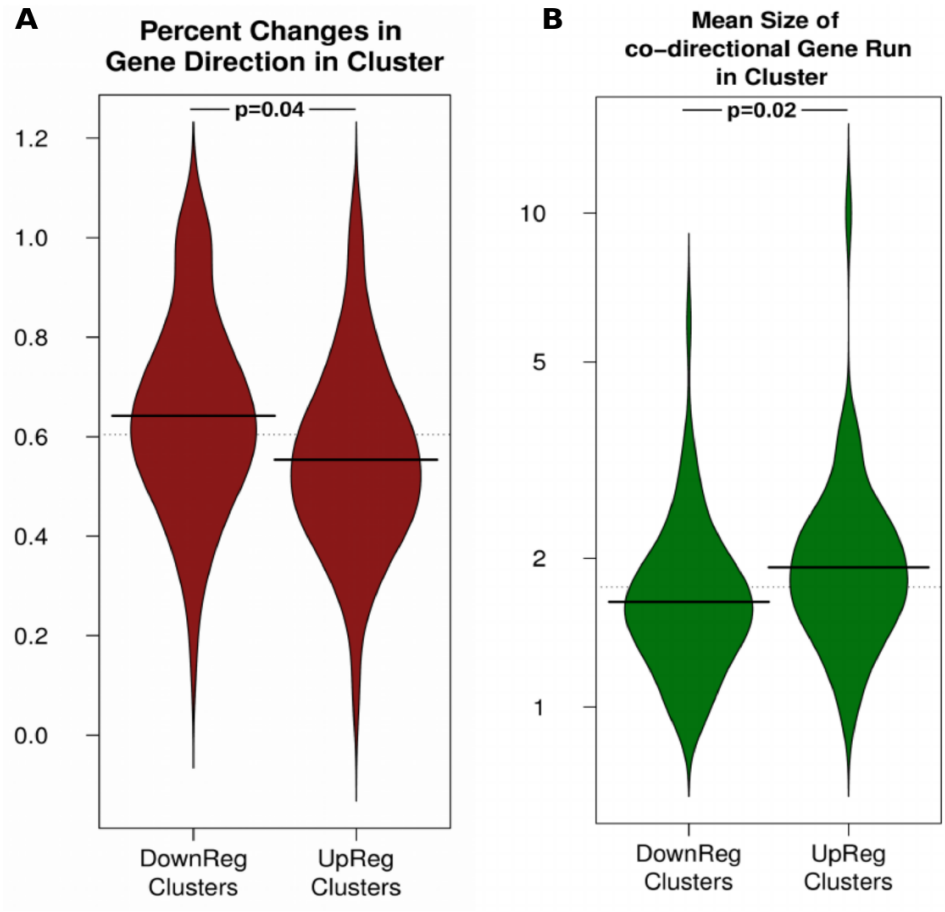
Supplementary Figure 5

A. Heatmap scatterplot of TCGC GRO values against mean sequence conservation, calculated as mean aggregate phastCons score for all genes in each cluster (N=116). Correlation measured as Pearsons r.

B. Transcription factor binding site (TFBS) density against GRO value for 116 TCGC. Correlation measured as Pearsons r. TFBS compiled by (9)

Quantification of gene directionality

Gene directionality was quantified with two different measures as described in Supplementary Methods. Both measures showed significant differences for the two types of TCGC, supporting the tendency for up-regulated genes to contain more co-directional genes (longer runs of co-directionality and smaller number of changes in gene transcription direction).



Supplementary Figure 6

Comparison of gene co-directionality measures for up- and down-regulated TCGC.

A. Percentage of changes in gene transcription direction as number of times transcription direction changes within a cluster divided by the number of genes in the cluster +2 (flanking genes). P-value calculated on the basis of Student's t-test.

B. Size of co-directional gene runs within TCGC. Gene runs are taken as the number of consecutive genes with the same direction. P-value calculated on the basis of Students t-test.

An example of an up-regulated TCGC

Together our findings suggest a characteristic profile for up-regulated gene clusters with subtelomeric localization, containing non essential, co-directional genes with low sequence conservation, separated by long intergenic spacers. This profile matches some well-described yeast gene clusters such as the DAL cluster, which enables *S. cerevisiae* to effectively metabolize allantoin as nitrogen source (10). Although it is not part of our strict 50 up-regulated TCGC, the DAL cluster shares many of their characteristics, being located in the subtelomeric region of the right arm of chromosome 9 and containing 6 consecutive genes with positive GRO values. The cluster appears to have been formed through recent rearrangements but has maintained the relative gene directionality within the *sensu stricto* yeast complex. Gene order within the DAL cluster has been shown to be crucial not only to the cell's fitness under nitrogen starvation but also to the coordinated expression of the genes themselves (11). More generally, it was recently shown that genome rearrangements in yeast although not lethally disruptive may affect gene transcription levels in a pervasive manner, extending beyond the areas of rearrangement (12). We believe that such phenomena may well be reflections of the intricate interplay between gene transcription and changes in DNA topology that we bring forward with our analyses.

Supplementary Methods

Formation of gene clusters

Contiguous genes with positive or negative values were joined and clusters were defined as the uninterrupted regions spanning the genomic space from the first to the last segment in an all-positive (up-regulated) or all-negative (down-regulated) gene series (see Figure 1A). The number of genes included in clusters ranged from 1 (single-gene clusters) to 31, with a mean value of 2.7.

Positional enrichments of gene clusters

Genomic coordinates for yeast centromeres were obtained from SGD (http://www.yeastgenome.org/locus/centromere). Cluster-centromere distances were calculated as the sequence length between the most proximal cluster boundary to the central point of the centromeric coordinates. Distances were then scaled with the size of the chromosomal arm extending from the central point of the centromere to the chromosome's boundary, so as to be represented in a range of 0 (i.e. overlapping the centromere) to 1 (i.e. lying at the edge of the corresponding chromosomal arm).

We used the clustering of yeast coordinates in network communities described in (2) using original 4C data from (1). We then calculated the enrichment of our TCGC, separately for up- and down-regulated ones in the 13 distinct level-1 communities (Supplementary Table 7 from (2)). Enrichment was calculated on the basis of an observed over expected ratio of overlaps between the two sets of genomic coordinates (TCGC and each corresponding community) and was statistically assessed on the basis of 1000 random permutations of cluster coordinates as described in the main text (see Methods). Overlaps with a bootstrap p-value less or equal to 0.01 were deemed significant. The complete table of enrichments for all 13 communities is available as Supplementary File 3.

Functional Enrichment

We employed a modified gene set enrichment functional analysis at TCGC level to analyze concerted over-representations of Gene Ontology terms (www.geneontology.org). Enrichment was calculated based on a hypergeometric test for each gene cluster and controlled for multiple comparisons at a 5% FDR (13). GO terms with significant enrichment (adjusted p-value <=0.05) in at least one of the two types of TCGC were recorded.

Transcription factor binding site Enrichment

Conserved Transcription Factor Binding Sites (TFBS) were obtained from the UCSC Genome Browser's Transcriptional Regulatory Code track. These corresponded to a compendium of 102 transcriptional regulators based on a combination of experimental results, cross-species conservation data for four species of yeast and motifs from the

literature compiled by (9). Enrichment in TF binding was calculated as in the case of chromosomal communities described above. Enrichments were assessed as ratios of observed over expected overlaps and p-values were obtained as bootstrap values from 1000 random permutations of cluster coordinates.

Intergenic space size

We used genomic coordinates downloaded from UCSC (SGD/saCcer1). Intergenic distances were calculated as the full length of regions spanning the genomic space between two consecutive genes, using transcription initiation and termination as boundaries, regardless of gene transcription direction. We assigned to each gene a mean intergenic space length to be the arithmetic mean of the lengths of gene upstream and downstream intergenic regions. For genes at chromosomal boundaries, one of the two intergenic regions were set to be equal to the distance from the gene boundary to the corresponding chromosomal start/end.

Sequence conservation and TFBS density

Sequence conservation was calculated as aggregate phastCons scores (14) obtained from UCSC and based on a multiple alignment of 7 *Saccharomyces* species. Mean conservation was taken as the mean phastCons score for a given region. For each cluster we removed intergenic space and calculated the mean aggregate phastCons score for all genes in the cluster.

TFBS density was calculated as the percentage of the length of each TCGC overlapping with conserved TFBS as compiled in (9).

Supplementary Files

Supplementary File 1

Relative GRO values of 5414 protein coding genes for yeast JCW28 strain mapped on chromosomal coordinates of sacCer1 (Oct. 2003, SGD/sacCer1).

Supplementary File 2

Genomic coordinates of 116 topologically co-regulated gene clusters (TCGC) with the number of contained gene and mean GRO value.

Supplementary File 3

Enrichments of TCGC with the 13 genomic network communities from (2).

Supplementary References

1. Duan Z, et al. (2010) A three-dimensional model of the yeast genome. *Nature* 465(7296):363–7.

2. Hoang S a, Bekiranov S (2013) The network architecture of the Saccharomyces cerevisiae genome. *PLoS One* 8(12):e81972.

3. Pryde FE, et al. (1999) Limitations of silencing at native yeast telomeres. *EMBO J* 18(9):2538–2550.

4. West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* 16(3):271–88.

5. Matthews HR (1993) Polyamines, Chromatin Structure and Transcription. *Bioessays* 15(8):561–566.

6. Buchman AR, Kimmerly WJ, Rine J, Kornberg RD (1988) Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. *Mol Cell Biol* 8(1):210–25.

7. Nikolaou C, et al. (2013) Topoisomerase II regulates yeast genes with singular chromatin architectures. *Nucleic Acids Res* 41(20):9243–56.

8. Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B (2009) Evolution of Gene Order in the Genomes of Two Related Yeast Species Evolution of Gene Order in the Genomes of Two Related Yeast Species. 2009–2019.

9. Harbison C, Gordon D, Lee T, Rinaldi N (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*:99–104.

10. Wong S, Wolfe KH (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet* 37(7):777–782.

11. Naseeb S, Delneri D (2012) Impact of chromosomal inversions on the yeast DAL cluster. *PLoS One* 7(8):e42022.

12. Naseeb S, et al. (2016) Widespread Impact of Chromosomal Inversions on Gene Expression Uncovers Robustness via Phenotypic Buffering. *Mol Biol Evol* 33(7):1679–96.

13. Chouvardas P, Kollias G, Nikolaou C (2016) Inferring active regulatory networks from gene expression data using a combination of prior knowledge and enrichment analysis. *BMC Bioinformatics* 17 Suppl 5:181.

14. Siepel A, et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8):1034–50.