bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genome urbanization: Clusters of topologically co-regulated genes delineate spatial, structural and 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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Background: 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 tension in DNA topology. 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 and distinct areas of the genome are expected to be differentially affected by DNA topological constraints. The scope of this work is to identify positional and structural preferences in the distribution of genes, relative to their response to topo-II deactivation.

Results: By analyzing a genome wide run-on experiment upon thermal inactivation of topo II in *S. cerevisiae* we were able to define 117 gene clusters of concerted 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.

Conclusions: This multi-faceted "division of labour" is much resembling a "genome urbanization" process with a core of essential genes occupying a compact genomic "old town", whereas more recently acquired, condition-specific genes are located in a more spacious "suburban" genomic periphery.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

The distribution of genes in the genome of eukaryotes is highly non-random. Along with the first genome-wide analyses, researchers were able to realize that genes are organized in linear space in an ordered way, that moreover, correlates with their expression levels. The existence of extended regions of increased gene expression (RIDGE) in the human genome was reported with the first genome-wide transcriptome map (Caron et al., 2001). Although it was later shown that this effect was due to the clustering of constitutive but not tissue-specific genes (Lercher, Urrutia, & Hurst, 2002), spatial associations of genes have since been used to provide the theoretical framework for links between gene expression and chromatin structure (Batada & Hurst, 2007) and the inference of protein-protein interaction patterns (Dandekar, Snel, Huynen, & Bork, 1998). 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 (Lee & Sonnhammer, 2003; Tiirikka, Siermala, & Vihinen, 2014).

The selective pressures underlying the localization of genes are thus of unequal intensity and diverse nature and a number of other, seemingly irrelevant characteristics may shape the overall genome architecture in evolution (Hurst, Pál, & Lercher, 2004). The implementation of chromosome conformation capture analyses (Dekker, Rippe, Dekker, & Kleckner, 2002; Lieberman-Aiden et al., 2009) has further demonstrated an underlying order existing in the three-dimensional space, where the impact of structural constraints is much more explicit (Bickmore & van Steensel, 2013; Fraser & Bickmore, 2007; Gilbert & Fraser, 2015). 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 (Ljungman & Hanawalt, 1992). More recently, it was 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 (Naughton et al., 2013). 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, were shown to have increased nucleosome turnover and recruitment of RNA-PolII molecules that resulted in elevated transcription levels (S. S. Teves & Henikoff, 2014). The association of supercoiling and its control by topoisomerases with gene regulation is evident in the way that accumulated positive supercoiling precludes the formation of transcription initiation complexes (Joshi, Piña, & Roca, 2010; Roca, 2011).

In the budding yeast (*Saccharomyces cerevisiae)*, the organization of genes in linear space has also been attributed to common regulatory mechanisms (Kruglyak & Tang, 2000). Yeast's distinguishing feature is the overall gene density, with genes covering ~70% of the total genome (Goffeau et al., 1996). Despite its reduced size of only 12Mbp, the transcription dynamics of the yeast genome is highly complex, with genes being

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

expressed in tandem and in operon-like transcripts, with varying sizes of gene upstream and downstream regions (David et al., 2006). Transcription directionality in such a highly streamlined genome also plays a crucial role in the regulatory process, with a number of bidirectional promoters (Xu et al., 2009) 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 (Tan-Wong et al., 2012). Thus, even in a small eukaryotic genome, there is a strong association between DNA structure and gene expression. We have recently demonstrated the regulatory role of topoisomerase II (topo II) in this context, through a Genomic transcription Run-On (GRO) experiment (Nikolaou et al., 2013), that showed different sets of genes responding in completely different ways to the accumulation of topological stress during transcriptional elongation.

In this work, we sought to investigate how this relationship between topo-II deactivation and gene expression may extend beyond single gene promoters to affect broader genomic regions and how these regions may be spanning gene clusters with particular properties. Based on a previously published genome-wide dataset (Nikolaou et al., 2013) we employed a series of novel analyses in order, first to define clusters of genes that are differentially affected by topo-II deactivation and then 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

Genes responsive to topological stress are clustered non-randomly

Having previously assessed genome-wide changes in transcription levels upon topo II thermal inactivation (Nikolaou et al., 2013), we first aimed to define clusters of genes affected in a similar way by topo-II deactivation. Our initial dataset comprised differential GRO values for 5414 yeast protein-coding genes (Supplementary File 1 and Supplementary Figure 1). We defined 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 (Hurst et al., 2004). 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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1A). The number of genes included in clusters ranged from 1 (single-gene clusters) to 31, with a mean value of 2.7. The distribution of the number of genes in clusters was significantly skewed to higher values as may be seen in Figure 1C.

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 bars. 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.

topologically co-regulated gene clusters are enriched in distinct regions of the yeast genome

In Figure 1D the relative positions and mean GRO values of 116 gene clusters with >=7 genes are shown on a linear representation of the yeast genome. The cluster distribution suggests a non-random placement along the different chromosomes. 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 (see Methods) showed statistically significant opposing preferences for the up- and down-regulated gene clusters to be located away from and close to centromeres respectively (Figure 2A).

The tendencies described above were calculated on the basis of linear chromosomes acting as independent units, which is of course far from the reality of the eukaryotic nucleus, where chromosomes interact in three-dimensional space forming inter-chromosomal domains (Gibcus & Dekker, 2013). 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 (Duan et al., 2010) and analyzed in chromosomal networks by (Hoang & Bekiranov, 2013). 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 (Hoang & Bekiranov, 2013) 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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

schematic representation of the nucleus in Figure 2B (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 (Duan et al., 2010). 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 (Pryde et al., 1999; West, Gaszner, & Felsenfeld, 2002), which fits well with the extensive over-representation of up-regulated clusters towards telomeric regions.

Radically different functions and regulatory modes in different types of TCGC

Topoisomerase II 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 responds to its transient deactivation with increased transcription levels indicates the existence of a positive effect for a subset of cellular functions. To this end, we performed a functional enrichment analysis for genes belonging in the up- and down-regulated clusters separately. Figure 3A shows a summary of the GO categories found to be enriched among the genes contained in at least one of the two types of TCGC. Three main clusters are apparent. One constitutes of functions enriched in both types of clusters. These are diverse functions including secondary metabolism and DNA maintenance. A second, smaller cluster is formed by GO terms that are enriched in up-regulated clusters and which are also depleted in down-regulated ones. These encompass functions related to cellular transport, the metabolism of co-factors and stress-response. A third, down-regulated gene specific cluster contains basic cellular functions associated with RNA transcription and processing, translation and the nuclear environment. Overall, the clustering of functional enrichments in Figure 3A shows extensive differences between the two types of TCGC, a fact indicative of their nuclear compartmentalization being echoed in their functional roles. A general pattern suggests that the localization of clusters among chromosomes is also weakly 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.

We next turned 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 (Harbison, Gordon, Lee, & Rinaldi, 2004) and went on to assess TFBS enrichment in the two types of TCGC. Enrichments were calculated on the basis of

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

an observed over expected ratio, 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 (see Methods for details). The results are summarised in Figure 3B, where transcription factors whose binding sites are found more or less frequently than expected by chance are highlighted for both types of TCGC. Down-regulated gene clusters tend to be mostly depleted of TFBS. This was expected since they were shown above to be enriched in constitutively expressed genes, which are 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 essential, with constant expression levels and mostly depleted of TATA-boxes (Nikolaou et al., 2013). TFBS depletion in down-regulated genes may also be explained on the basis of structural properties of the clusters that we discuss in the next section.

Up-regulated TCGC, on the other hand, show the exact opposite pattern, with most of the TFBS being enriched. Significant enrichment was found for factors such as AFT1(RCS1), AFT2 and SWI6, that are involved in chromatin surveillance and STP1 and SNT2, which are key regulators 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 (Matthews, 1993). Up-regulated gene clusters are depleted in binding sites for ABF1, a regulator of DNA replication strongly associated with autonomously replicating sequences (ARS) (Buchman, Kimmerly, Rine, & Kornberg, 1988). 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.

Both analyses described herein at functional and regulatory levels need to be interpreted with caution. The genes analyzed are not necessarily the ones most affected by topo-II deactivation but those with similar response to it, that also tend to cluster in close vicinity. In this sense, it is reasonable to expect broader enrichments that do not directly reflect the topological impact of topo II inactivation. On the other hand, the general preferences that arise, namely the enrichment of down-regulated clusters in constitutive, essential functions and their depletion in transcription factor binding sites 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.

Genes in topologically co-regulated clusters have markedly different structural attributes in terms of gene spacing and transcriptional directionality

We next shifted our focus on the properties of the TCGC in terms of gene length and directionality. During transcription, DNA torsional stress accumulates with different sign

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 topo II deactivation has been shown to be generally independent from the size of the majority of yeast genes (Pedersen et al., 2012), but it is strongly inhibitory in the case of long transcripts (Joshi, Piña, & Roca, 2012). The situation is very different when one looks, instead, into the surrounding intergenic space. Here we show that there is very significant dependence of the GRO response on the size of the intergenic spacers both upstream and downstream of the gene. In Figure 4A 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 (see Methods). There is 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.

This association between DNA topology and structural genomic features is more pronounced in the clusters of genes with similar GRO values, like the TCGC described above. In order to study the effect of intergenic space in gene clusters of homogeneous size, we employed a more relaxed criterion in their 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 4B. 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 they 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, combined with an increased nucleosomal density (Nikolaou et al., 2013) can account for the marked under-representation of TFBS in down-regulated gene clusters (see Supplementary Figure 2).

Figures 4A and 4B are strongly indicative of the impact of genomic architecture on the maintenance of topological equilibrium in the nucleus. Genes with short intergenic space 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 topo II, while genes that allow for the dissipation of topological strain to long, untranscribed, nearby regions are predictably more resilient. This dependence, already evident at the level of individual genes, is further accentuated in gene clusters (Figure 4B), which points to the existence of synergistic effects between genes in the resolution of topological stress.

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gene transcription directionality may also have a role in the way topological stress is resolved under active transcription. Tandemly transcribed genes are not uncommon in eukaryotes and in many cases they have been associated with common functional roles, reminiscent of prokaryotic operons. Gene clusters with all the characteristics of operons are abundant in *C. elegans* (Blumenthal et al., 2002; Nimmo & Woollard, 2002), while uninterrupted transcripts exist in yeast (David et al., 2006). On the other hand, gene pairs with divergent transcriptional direction are usually controlled by common regulatory sequences, organized in bi-directional promoters (Takai & Jones, 2004; Xu et al., 2009). We hypothesized that the directionality of transcription of consecutive genes in yeast may correlate with the coordinated response to topological strain by broader gene communities. In this sense, gene clusters with more "streamlined" directionality patterns are expected to be able to accommodate supercoiling in a more effective manner and be thus less sensitive to topo II inactivation. 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 tandemly transcribed genes 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 cluster category. The results, shown in Figure 4C are indicative of a mild, yet significant association between gene directionality patterns and response to topo-II 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 (see Supplementary Figure 3).

Having already described contrasting preferences in terms of the gene clusters' positional distribution and functional roles of contained genes, we see that different structural attributes in terms of relative gene distances and directionality may provide a mechanistic framework for their coordinated response to topo-II deactivation. We next sought to investigate how such structural features may be constrained through evolution.

Different conservation constraints in topologically co-regulated gene clusters

In order to gain insight on the constraints existing in TCGC, we performed an analysis of conservation at two levels. First, we analyzed the mean sequence conservation per cluster as aggregate phastCons scores (Siepel et al., 2005), obtained from a genome-wide alignment of six *Saccharomyces* species (Kellis, Patterson, Endrizzi, Birren, & Lander, 2003). Figure 5A shows the average sequence conservation, as mean phastCons score, to

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be negatively correlated with the mean GRO value for the 116 TCGC, confirming that down-regulated clusters are significantly more constrained in terms of sequence conservation. This tendency was also suggested by examining individual genes in a previous study by our group (Nikolaou et al., 2013), however, it is when looking at broader chromosomal domains, such as the ones delineated by TCGC that it becomes apparent. 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 (Fischer, Neuvéglise, Durrens, Gaillardin, & Dujon, 2009).

We next turned to more complex conservational features that also take into account synteny relationships as 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) (Byrne & Wolfe, 2005) that contains a detailed catalog of orthologous genes between a number of yeast species before and after the whole-genome duplication event (Kellis, Birren, & Lander, 2004). We collected all orthologous pairs between *S. cerevisiae* and two of its closest species in the *sensu stricto* complex, *S. paradoxus* and *S. mikatae* (Rokas, Williams, King, & Carroll, 2003). We analyzed them separately for up- and down-regulated clusters by calculating a simple measure of "directional conservation", as described in Methods. In brief, we considered every gene, whose relative position and direction was retained, to be positively contributing to this measure, while rearranged or opposite direction genes were assigned negative values. We then described the two types of TCGC in terms of both raw sequence and transcriptional direction constraints. Given that syntenic regions are by definition under sequence constraint we were not surprised to see that down-regulated clusters were characterized by both high sequence and high synteny as may be seen in Figure 5B, where a contour heatmap of enrichment shows down-regulated clusters to be preferentially located in the top right corner, indicative of high values for both measures. What was rather interesting was the corresponding position of up-regulated clusters in the same two-dimensional constraint space. While we already knew from our conservation analysis (Figure 5A) that sequence constraints were more relaxed in these regions, we found a significant proportion of genes retaining their syntenic and directionality relationships even if at lower sequence conservation values. Figure 5B suggests 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.

Figure 5B indicates a different type of constraint existing at the level of gene organization that may be underlying the way these genes respond to topo-II deactivation. Accumulation of supercoiling is known to stall or even entirely stop transcription, however under certain circumstances its dissipation may facilitate or even promote transcriptional elongation (S. Teves & Henikoff, 2014; S. S. Teves & Henikoff, 2014). This dissipative potential may be crucial for genes belonging to up-regulated

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

clusters, as we have seen those to be generally linked to the way the cell responds to stress, including a broad number of secondary metabolic pathways. 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 (Wong & Wolfe, 2005). Albeit, the DAL cluster is not included in our strictly defined topologically co-regulated gene clusters, it shares many of the characteristics of up-regulated clusters, being located in the subtelomeric region of the right arm of chromosome 9 and containing 6 consecutive genes with positive GRO values. As discussed in (Wong & Wolfe, 2005) 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 (Naseeb & Delneri, 2012). 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 (Naseeb et al., 2016). We believe that this 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.

Discussion

Genome "Urbanization". A model for genome organization in functionally and structurally autonomous gene "neighborhoods"

Carefully designed experiments can always serve for the testing of novel hypotheses. In this work we have revisited a published dataset (Nikolaou et al., 2013) under a different perspective, aiming at investigating a link between DNA topology and the overall genome architecture in a simple eukaryote. Having already noted, that a number of yeast genes seem to respond to topological stress in radically different ways, we set out to examine whether this behaviour could be accounted for simply by local or more extended functional and structural attributes. Our primary finding was that the response to DNA topological constraints may be synergistically orchestrated in gene neighborhoods with particular characteristics. The existence of clusters of genes with differential transcriptional response to the inactivation of topo II, was thus the starting point for a detailed study of the particularities of non-random gene organization through the lens of DNA topology.

By persistently analyzing the defined topologically co-regulated gene clusters at various levels, we were thus 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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

"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 in regions with high nucleosomal density. These consisted of highly conserved genes associated with essential functions that were predominantly located in close proximity to each other, with small intergenic areas in between them and with highly variable patterns of gene directionality. In striking contrast, clusters of up-regulated genes were prevalent in the nuclear periphery, enriched towards the telomeres and under more relaxed sequence constraints as they were associated with secondary functions. Contrary to their down-regulated counterparts, these clusters contained co-directionally transcribed genes that were also separated by long interegenic spacers.

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 6). This "Genome Urbanization" is echoed in various genomic features that we have discussed in the context of topologically co-regulated gene clusters. 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 extend beyond DNA topology that has been the basis of our study. 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 exceed topological aspects and extend to gene functionality (Gehlen et al., 2012), regulation programs (Xu et al., 2009; Zhu et al., 2008) and genome evolution (Sugino & Innan, 2012).

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 and bear specific characteristics. 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 (Lehner, 2010). The size of the intergenic space between genes has also been shown to widely shape expression variability (Bajić et al., 2012).

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 (Bagadia, Singh, & Sandhu, 2016). The advent of new experimental

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

We used data 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 (García-Martínez, Aranda, & Pérez-Ortín, 2004) and data were analyzed as previously described in (Nikolaou et al., 2013).

Gene Clustering

Linear clusters of genes were called based on their GRO value. 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 (Figure 1A). Clusters of >=7 genes were selected on the basis of a bootstrapping analysis as suggested in (Hurst et al., 2004). This was performed by conducting 10000 random permutations of gene order while keeping the same GRO values. We used functions from the BedTools Suite (Quinlan & Hall, 2010) to control for unaltered gene sizes and chromosomal distributions. Gene number distributions of the derived clusters were then 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 cases (bootstrap value p=0.0008).

Clusters with >=7 genes were divided into up-regulated and down-regulated, depending on the mean GRO value of all genes in each cluster (Supplementary File 2).

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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Hoang & Bekiranov, 2013) using original 4C data from (Duan et al., 2010). 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 (Hoang & Bekiranov, 2013)). 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 above. 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. GO terms with significant enrichment 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 (Harbison et al., 2004). 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.

Gene Size and Direction of transcription

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.

Consecutive gene clusters used in the analysis of intergenic space size and gene directionality (see Results) were created in the following way: Each chromosome was scanned in overlapping 11-gene windows and for each step we recorded: the full list of 11 GRO values, gene lengths and mean intergenic space lengths (see above). We used this list to obtain patterns of gene directionality as arrays of seven genes (7 being the minimum size of the selected clusters). GRO values of the central gene were analyzed

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for three characteristic patterns corresponding to a) co-directional gene (all 7 genes transcribed in the same direction) b) middle divergent gene pair (the central gene being member of a divergent gene pair) c) the central gene being a member of a convergent gene pair.

Sequence conservation

Sequence conservation was calculated as aggregate phastCons scores (Siepel et al., 2005) 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.

Retention of orthologous gene directionality

We obtained orthologous gene coordinates for *S. paradoxus* and *S. mikatae*, being the closest species to *S. cerevisiae* from the Yeast Gene Order Browser (YGOB) (Byrne & Wolfe, 2005). 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 direction conservation on a scale of 0 (no retention of direction) to 1 (absolute retention of direction). The contour map of Figure 5B 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.

Author Contributions

MT and MM conducted bioinformatic analyses for the positional and functional enrichments, JR planned and conducted the GRO experiment and co-wrote the paper, CN conceived of the study, performed analyses and drafted the paper. All authors have read and approved the final manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 clusters as compared to the mean of 1000 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. 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.

B. Schematic showing the relative enrichments in yeast genome network communities obtained from (Hoang & Bekiranov, 2013) based on 4C data from (Duan et al., 2010). Enrichments were calculated as observed/expected ratios of overlaps between TCGC and network community coordinates (see Methods). Only statistically significant enrichments or depletions (p<=0.05) are reported. Figure is adopted from the network reconstruction by (Hoang & Bekiranov, 2013).

Figure 3

A. GO term enrichment heatmap of TCGC of both types. Enrichments were calculated on a modified Gene Set Enrichment Analysis (Chouvardas, Kollias, & Nikolaou, 2016). Only GO terms with an adjusted p-value (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 (Harbison et al., 2004). Enrichments are shown as log-2 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 4

A. 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

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

upstream and downstream intergenic regions for each gene (see Methods). Correlation measured as Spearman's rho.

B. Top, mean intergenic region length (same as in A) 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.

C. Distribution of GRO values of genes lying in the center of 7-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.

Figure 5

A. Heatmap scatterplot of TCGC GRO values against mean sequence conservation, calculated as aggregate phastCons score for the entire cluster (N=116). Correlation measured as Spearman rho.

B. 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 (see Methods for details). Enrichments were calculated as log2(ratios) of proportion of up-regulated/down-regulated clusters having values in a 10x10 value grid (see Methods for details).

Figure 6

Genome Urbanization in *S. cerevisiae*. A schematic of the yeast interphase nucleus is shown based on the Rabl configuration (Taddei & Gasser, 2012). Pericentromeric regions correspond to what we call the “Old city center” with enrichment in gene clusters down-regulated under topo-II 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.

Supplementary Figure 1

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

Supplementary Figure 2

Transcription factor binding site (TFBS) density against GRO value for 116 TCGC. Correlation measured as Spearman Rank rho. TFBS compiled by (Harbison et al., 2004)

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 3

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 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.

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (Hoang & Bekiranov, 2013).

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bibliography

Bagadia, M., Singh, A., & Sandhu, K. S. (2016). Three dimensional organization of genome might have guided the dynamics of gene order evolution in eukaryotes. *Genome Biology and Evolution*, *8*(3), evw050. http://doi.org/10.1093/gbe/evw050

Bajić, D., Poyatos, J. F., Enard, W., Khaitovich, P., Klose, J., Zöllner, S., … Giaever, G. (2012). Balancing noise and plasticity in eukaryotic gene expression. *BMC Genomics*, *13*(1), 343. http://doi.org/10.1186/1471-2164-13-343

Batada, N., & Hurst, L. (2007). Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nature Genetics*, *39*(8), 945–949. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17660811?itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum&ordinalpos=3

Bickmore, W. A., & van Steensel, B. (2013). Genome Architecture: Domain Organization of Interphase Chromosomes. *Cell*, *152*(6), 1270–1284. http://doi.org/10.1016/j.cell.2013.02.001

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., … Kim, S. K. (2002). A global analysis of Caenorhabditis elegans operons. *Nature*, *417*(6891), 851–854. http://doi.org/10.1038/nature00831

Buchman, A. R., Kimmerly, W. J., Rine, J., & Kornberg, R. D. (1988). Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, *8*(1), 210–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3275867

Byrne, K. P., & Wolfe, K. H. (2005). The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Research*, *15*(10), 1456–61. http://doi.org/10.1101/gr.3672305

Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., … Versteeg, R. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science (New York, N.Y.)*, *291*(5507), 1289–92. http://doi.org/10.1126/science.1056794

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. http://doi.org/10.1186/s12859-016-1040-7

Dandekar, T., Snel, B., Huynen, M., & Bork, P. (1998). Conservation of gene order: A fingerprint of proteins that physically interact. *Trends in Biochemical Sciences*. http://doi.org/10.1016/S0968-0004(98)01274-2

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., … Steinmetz, L.

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

M. (2006). A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(14), 5320–5. http://doi.org/10.1073/pnas.0601091103

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–11. http://doi.org/10.1126/science.1067799

Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., … Noble, W. S. (2010). A three-dimensional model of the yeast genome. *Nature*, *465*(7296), 363–7. http://doi.org/10.1038/nature08973

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. http://doi.org/10.1101/gr.212701

Fraser, P., & Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature*, *447*(7143), 413–7. http://doi.org/10.1038/nature05916

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

Gehlen, L. R., Gruenert, G., Jones, M. B., Rodley, C. D., Langowski, J., O'Sullivan, J. M., & O'Sullivan, J. M. (2012). Chromosome positioning and the clustering of functionally related loci in yeast is driven by chromosomal interactions. *Nucleus*, *3*(October 2015), 370–383. http://doi.org/10.4161/nucl.20971

Gibcus, J. H., & Dekker, J. (2013). The hierarchy of the 3D genome. *Molecular Cell*, *49*(5), 773–82. http://doi.org/10.1016/j.molcel.2013.02.011

Gilbert, D. M., & Fraser, P. (2015). Three Dimensional Organization of the Nucleus: adding DNA sequences to the big picture. *Genome Biology*, *16*(1), 181. http://doi.org/10.1186/s13059-015-0751-9

Goffeau, A., Barrell, G., Bussey, H., Davis, R. W. W., Dujon, B., Feldmann, H., … Oliver, S. G. G. (1996). Life with 6000 genes. *Science*, *274*(5287), 546–567. Retrieved from http://www.scopus.com/inward/record.url?eid=2-s2.0-10244239321&partnerID=40&md5=7cdabf14f159e536d40657780ecbf960

Harbison, C., Gordon, D., Lee, T., & Rinaldi, N. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature*, 99–104. http://doi.org/10.1038/nature02886.Published

Hoang, S. a, & Bekiranov, S. (2013). The network architecture of the Saccharomyces cerevisiae genome. *PloS One*, *8*(12), e81972. http://doi.org/10.1371/journal.pone.0081972

Hurst, L. D., Pál, C., & Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

order. *Nature Reviews. Genetics*, *5*(4), 299–310. http://doi.org/10.1038/nrg1319

Joshi, R. S., Piña, B., & Roca, J. (2010). Positional dependence of transcriptional inhibition by DNA torsional stress in yeast chromosomes. *The EMBO Journal*, *29*(4), 740–8. http://doi.org/10.1038/emboj.2009.391

Joshi, R. S., Piña, B., & Roca, J. (2012). Topoisomerase II is required for the production of long Pol II gene transcripts in yeast. *Nucleic Acids Research*, gks626–. http://doi.org/10.1093/nar/gks626

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*(6983), 617–624. http://doi.org/10.1038/nature02424

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., & Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, *423*(6937), 241–54. http://doi.org/10.1038/nature01644

Kruglyak, S., & Tang, H. (2000). Regulation of adjacent yeast genes. *Trends in Genetics*, *16*(3), 109–111. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Regulation+of+adjacent+yeast+genes#0

Lee, J. M., & Sonnhammer, E. L. L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. *Genome Research*, *13*(5), 875–882. http://doi.org/10.1101/gr.737703

Lehner, B. (2010). Conflict between noise and plasticity in yeast. *PLoS Genetics*, *6*(11), e1001185. http://doi.org/10.1371/journal.pgen.1001185

Lercher, M. J., Urrutia, A. O., & Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genetics*, *31*(2), 180–3. http://doi.org/10.1038/ng887

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, *326*(5950), 289–93. http://doi.org/10.1126/science.1181369

Ljungman, M., & Hanawalt, P. C. (1992). Localized torsional tension in the DNA of human cells. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(13), 6055–6059.

Matthews, H. R. (1993). Polyamines, Chromatin Structure and Transcription. *Bioessays*, *15*(8), 561–566. http://doi.org/10.1002/bies.950150811

Naseeb, S., Carter, Z., Minnis, D., Donaldson, I., Zeef, L., & Delneri, D. (2016). Widespread Impact of Chromosomal Inversions on Gene Expression Uncovers Robustness via Phenotypic Buffering. *Molecular Biology and Evolution*, *33*(7), 1679–96. http://doi.org/10.1093/molbev/msw045

Naseeb, S., & Delneri, D. (2012). Impact of chromosomal inversions on the yeast DAL

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cluster. *PloS One*, *7*(8), e42022. http://doi.org/10.1371/journal.pone.0042022

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J. G., Mati, I. K., Eijk, P. P., … Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nature Structural & Molecular Biology*, *20*(3), 387–95. http://doi.org/10.1038/nsmb.2509

Nikolaou, C., Bermúdez, I., Manichanh, C., García-Martinez, J., Guigó, R., Pérez-Ortín, J. E., & Roca, J. (2013). Topoisomerase II regulates yeast genes with singular chromatin architectures. *Nucleic Acids Research*, *41*(20), 9243–56. http://doi.org/10.1093/nar/gkt707

Nimmo, R., & Woollard, A. (2002). Widespread organisation of C. elegans genes into operons: Fact or function? *BioEssays*, *24*(11), 983–987. http://doi.org/10.1002/bies.10181

Pedersen, J. M., Fredsoe, J., Roedgaard, M., Andreasen, L., Mundbjerg, K., Kruh??ffer, M., … Andersen, A. H. (2012). DNA Topoisomerases Maintain Promoters in a State Competent for Transcriptional Activation in Saccharomyces cerevisiae. *PLoS Genetics*, *8*(12). http://doi.org/10.1371/journal.pgen.1003128

Pryde, F. E., Louis, E. J., Aparicio, O., Gottschling, D., Aparicio, O., Billington, B., … Voytas, D. (1999). Limitations of silencing at native yeast telomeres. *The EMBO Journal*, *18*(9), 2538–2550. http://doi.org/10.1093/emboj/18.9.2538

Quinlan, A., & Hall, I. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841. Retrieved from http://bioinformatics.oxfordjournals.org/cgi/content/short/26/6/841?rss=1

Roca, J. (2011). Transcriptional inhibition by DNA torsional stress. *Transcription*, *2*(2), 82–85. http://doi.org/10.4161/trns.2.2.14807

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, *425*(6960), 798–804. http://doi.org/10.1038/nature02053

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., … Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, *15*(8), 1034–50. http://doi.org/10.1101/gr.3715005

Sugino, R. P., & Innan, H. (2012). Natural selection on gene order in the genome reorganization process after whole-genome duplication of yeast. *Molecular Biology and Evolution*, *29*(1), 71–9. http://doi.org/10.1093/molbev/msr118

Taddei, A., & Gasser, S. M. (2012). Structure and function in the budding yeast nucleus. *Genetics*, *192*(1), 107–29. http://doi.org/10.1534/genetics.112.140608

Takai, D., & Jones, P. A. (2004). Origins of bidirectional promoters: computational

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

analyses of intergenic distance in the human genome. *Mol Biol Evol*, *21*(3), 463–7. http://doi.org/10.1093/molbev/msh040

Tan-Wong, S. M., Zaugg, J. B., Camblong, J., Xu, Z., Zhang, D. W., Mischo, H. E., … Proudfoot, N. J. (2012). Gene loops enhance transcriptional directionality. *Science (New York, N.Y.)*, *338*(6107), 671–5. http://doi.org/10.1126/science.1224350

Teves, S., & Henikoff, S. (2014). DNA torsion as a feedback mediator of transcription and chromatin dynamics. *Nucleus (Austin, Tex.)*, *5*(3), 1–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24819949

Teves, S. S., & Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes. *Nature Structural & Molecular Biology*, *21*(1), 88–94. http://doi.org/10.1038/nsmb.2723

Tiirikka, T., Siermala, M., & Vihinen, M. (2014). Clustering of gene ontology terms in genomes. *Gene*, *550*(2), 155–164. http://doi.org/10.1016/j.gene.2014.06.060

West, A. G., Gaszner, M., & Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. *Genes & Development*, *16*(3), 271–88. http://doi.org/10.1101/gad.954702

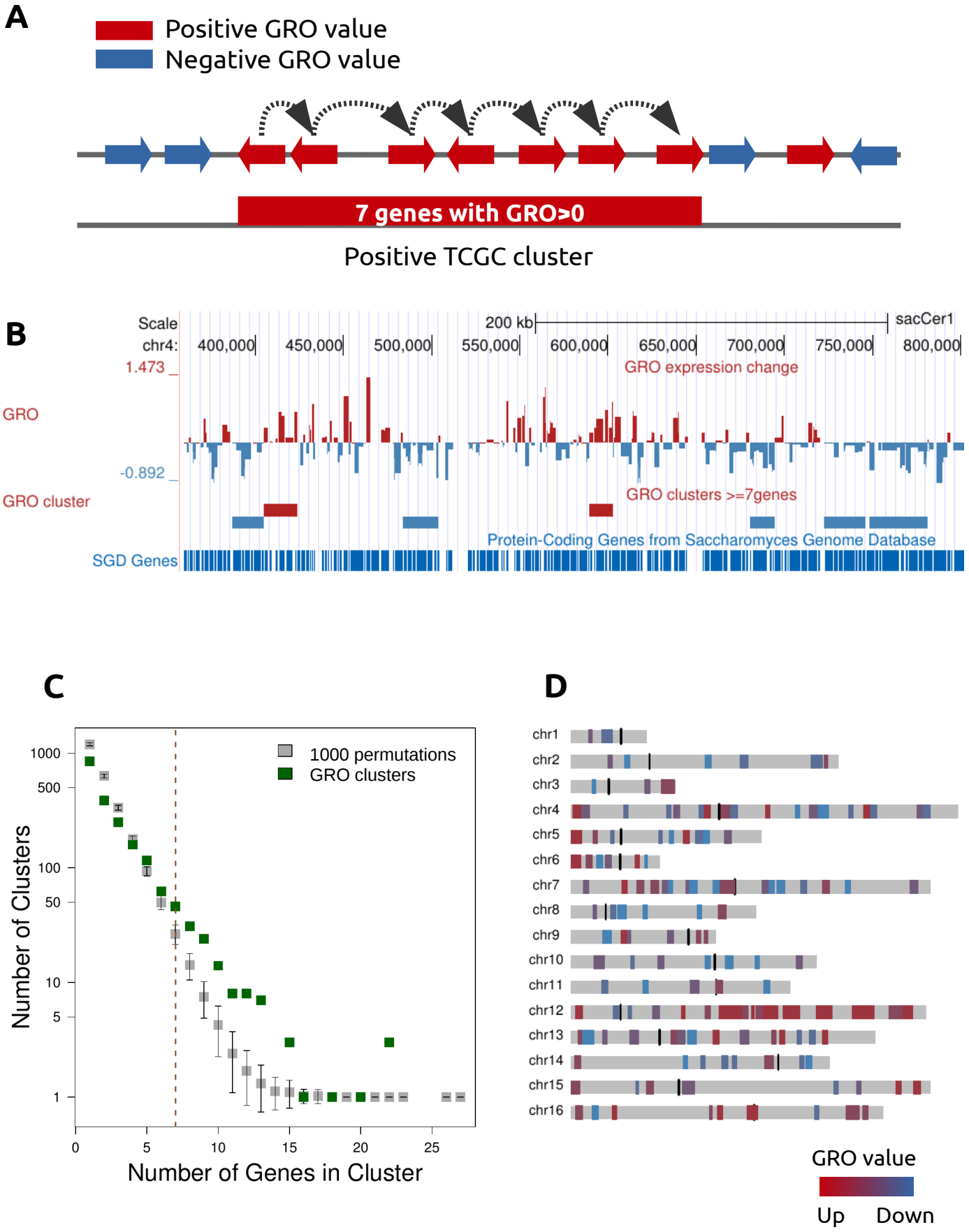
Wong, S., & Wolfe, K. H. (2005). Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nature Genetics*, *37*(7), 777–782. http://doi.org/10.1038/ng1584

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., … Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, *457*(7232), 1033–7. http://doi.org/10.1038/nature07728

Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., … Schadt, E. E. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, *40*(7), 854–61. http://doi.org/10.1038/ng.167

bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

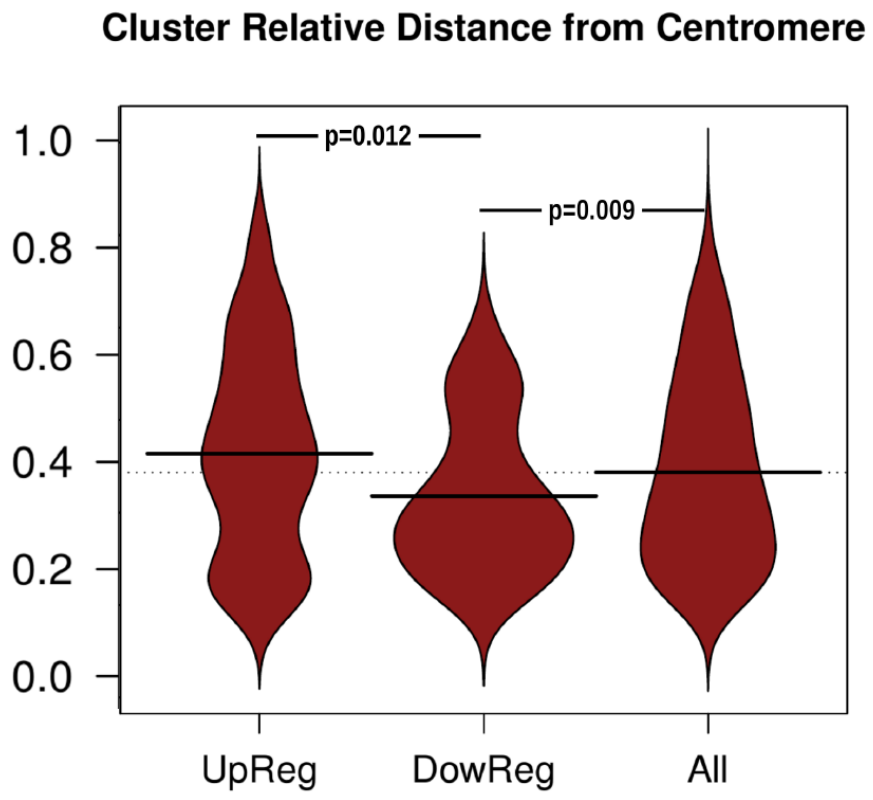
Figure 1



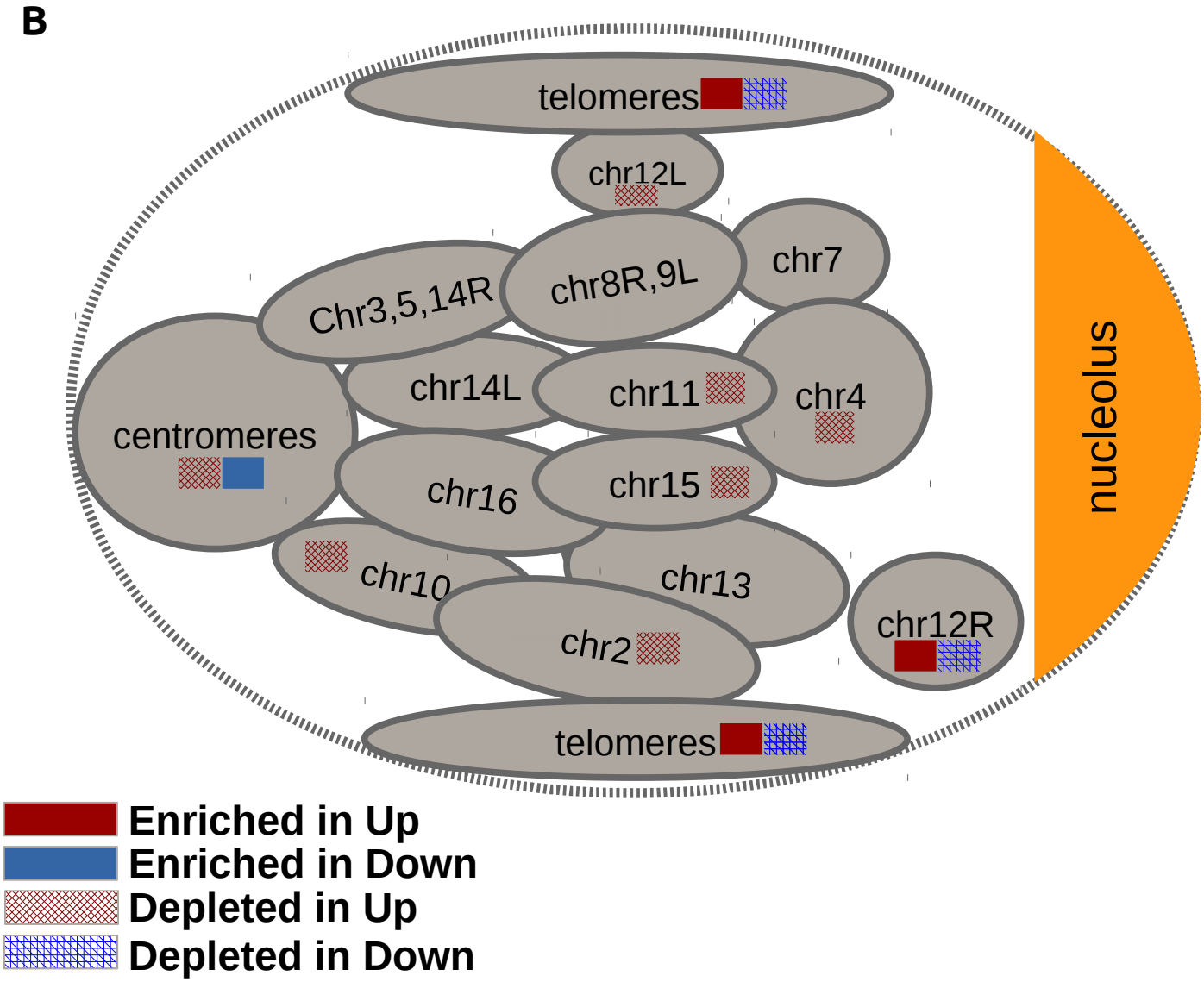
bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2

A



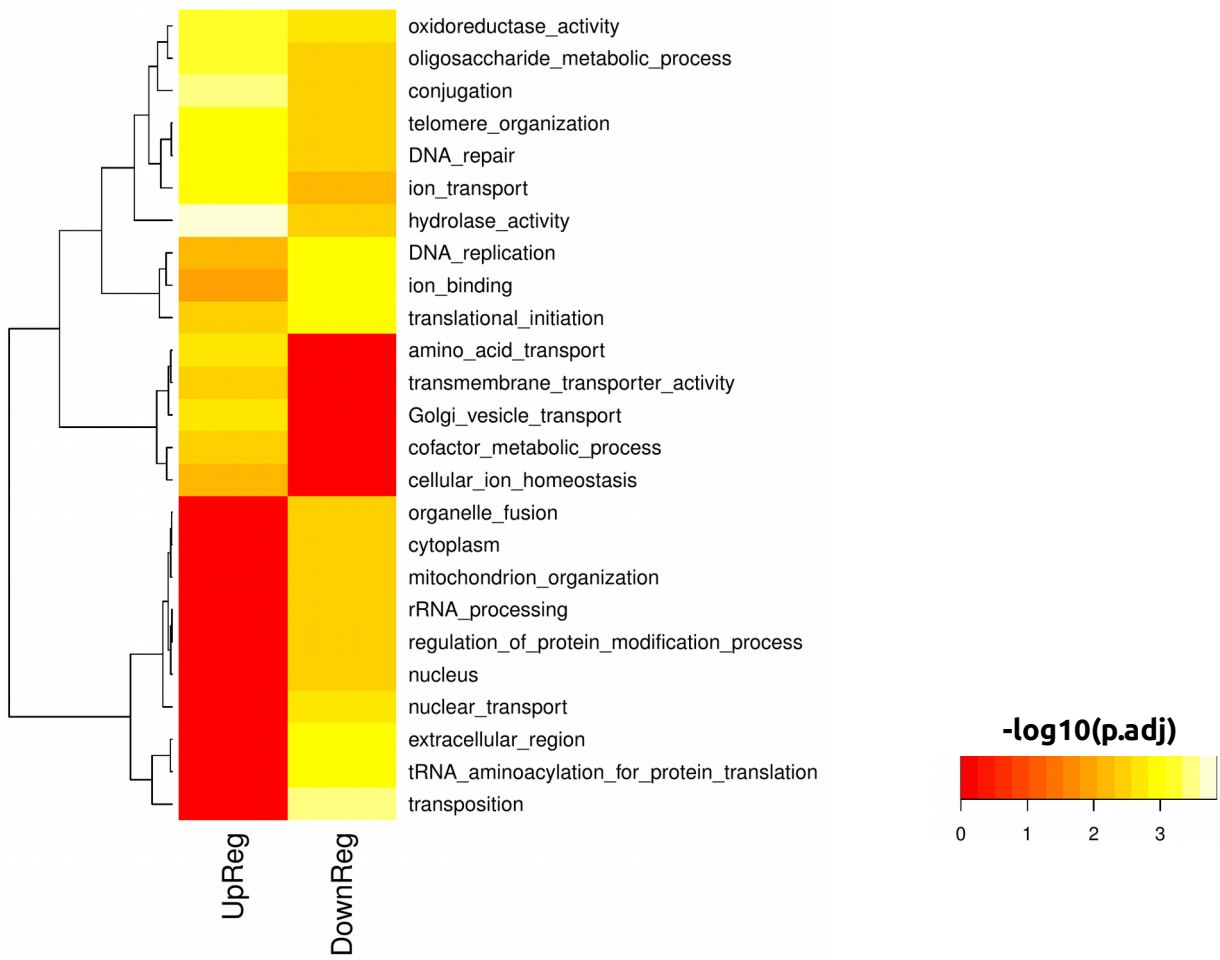
B



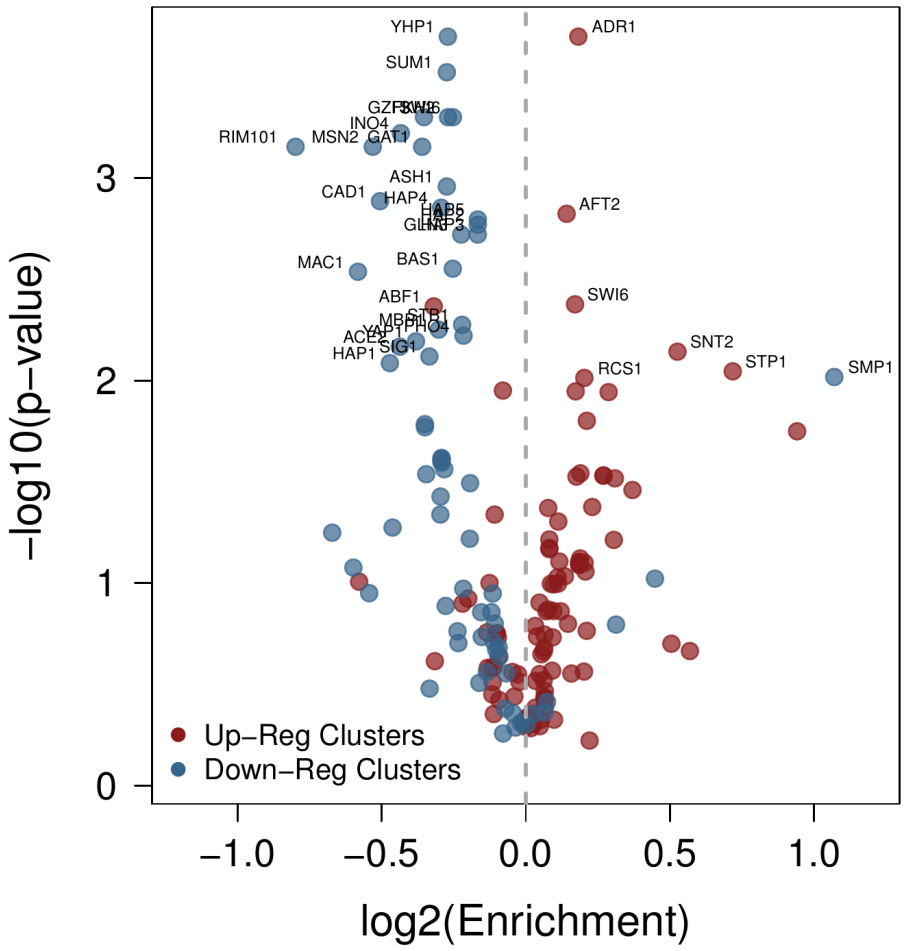
bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3

A

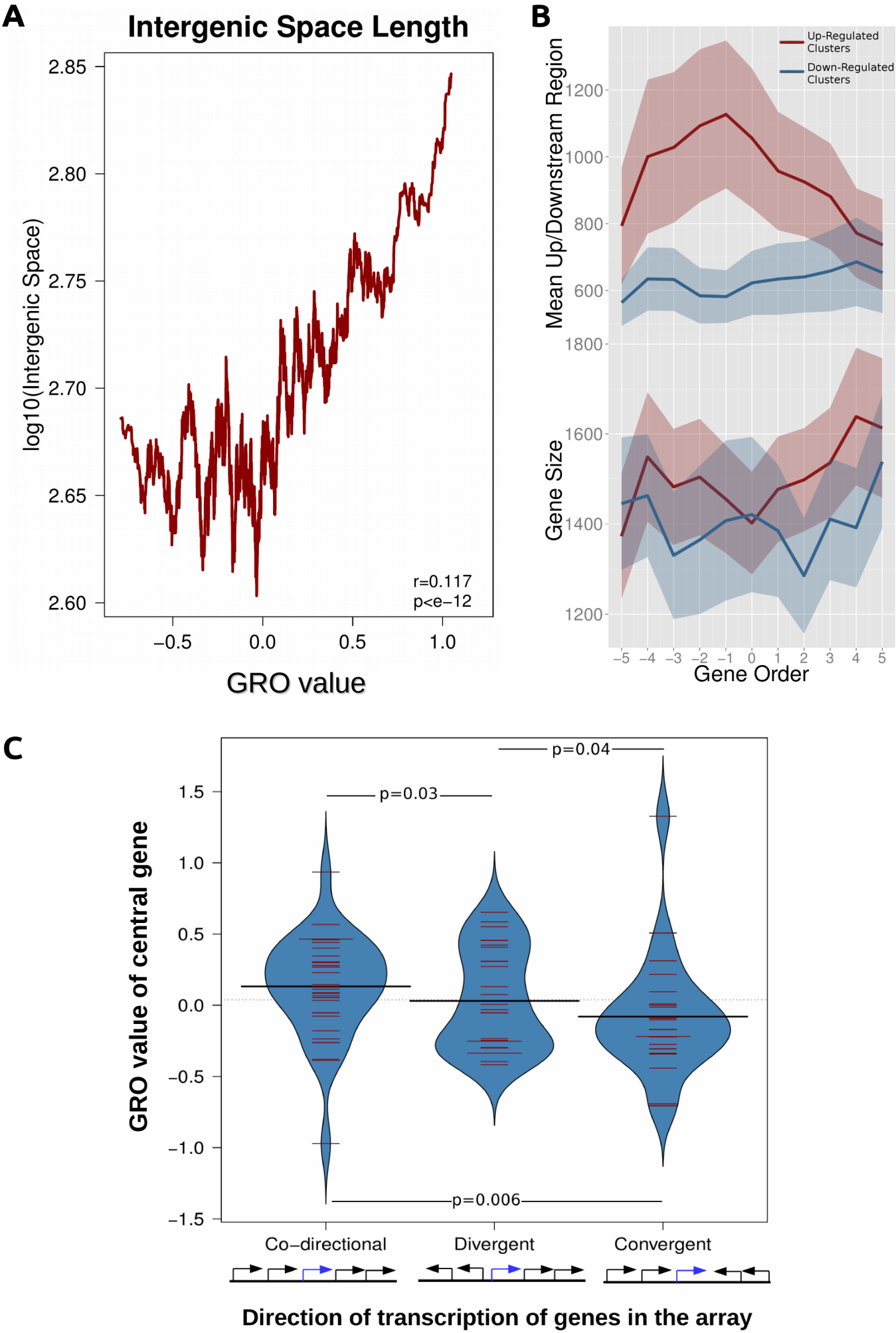


B



bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

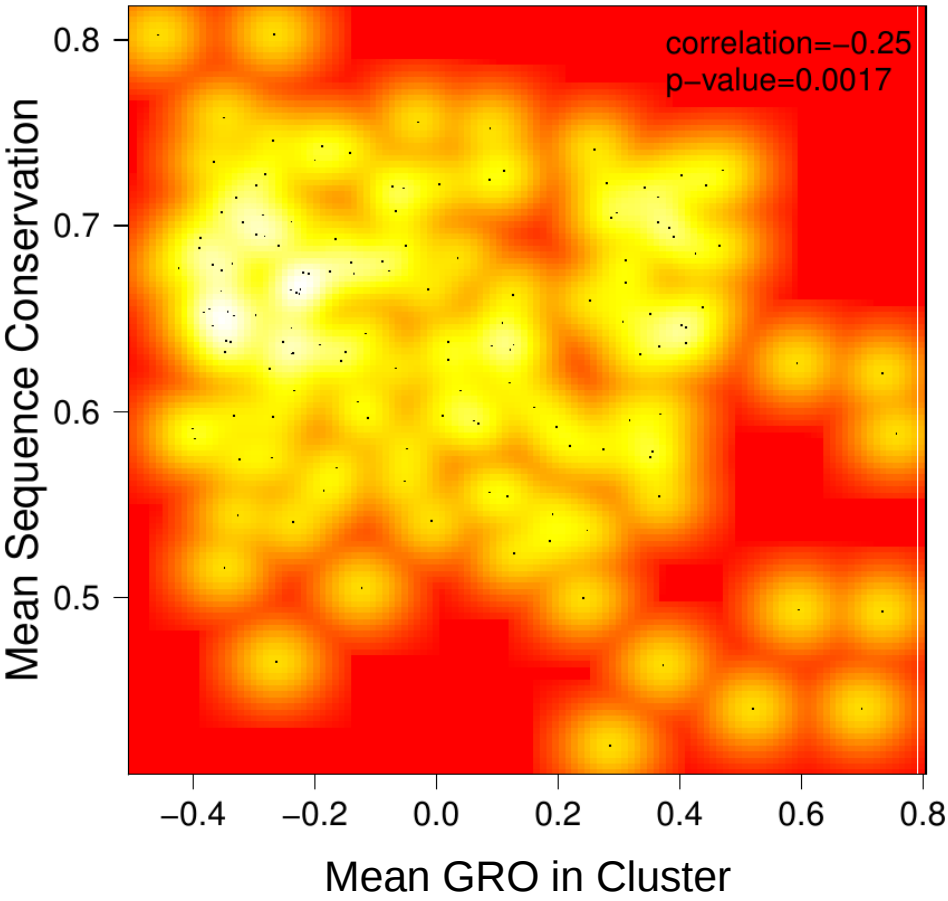
Figure 4



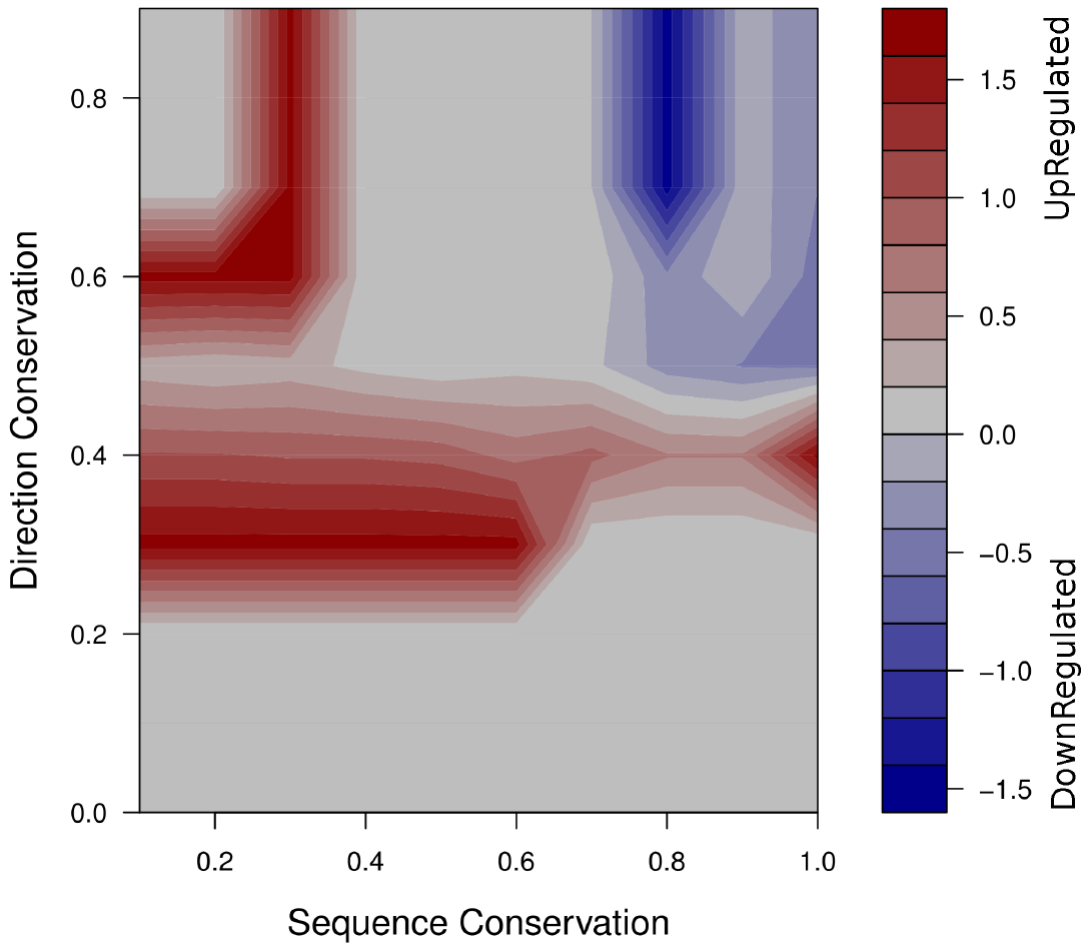
bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5

A

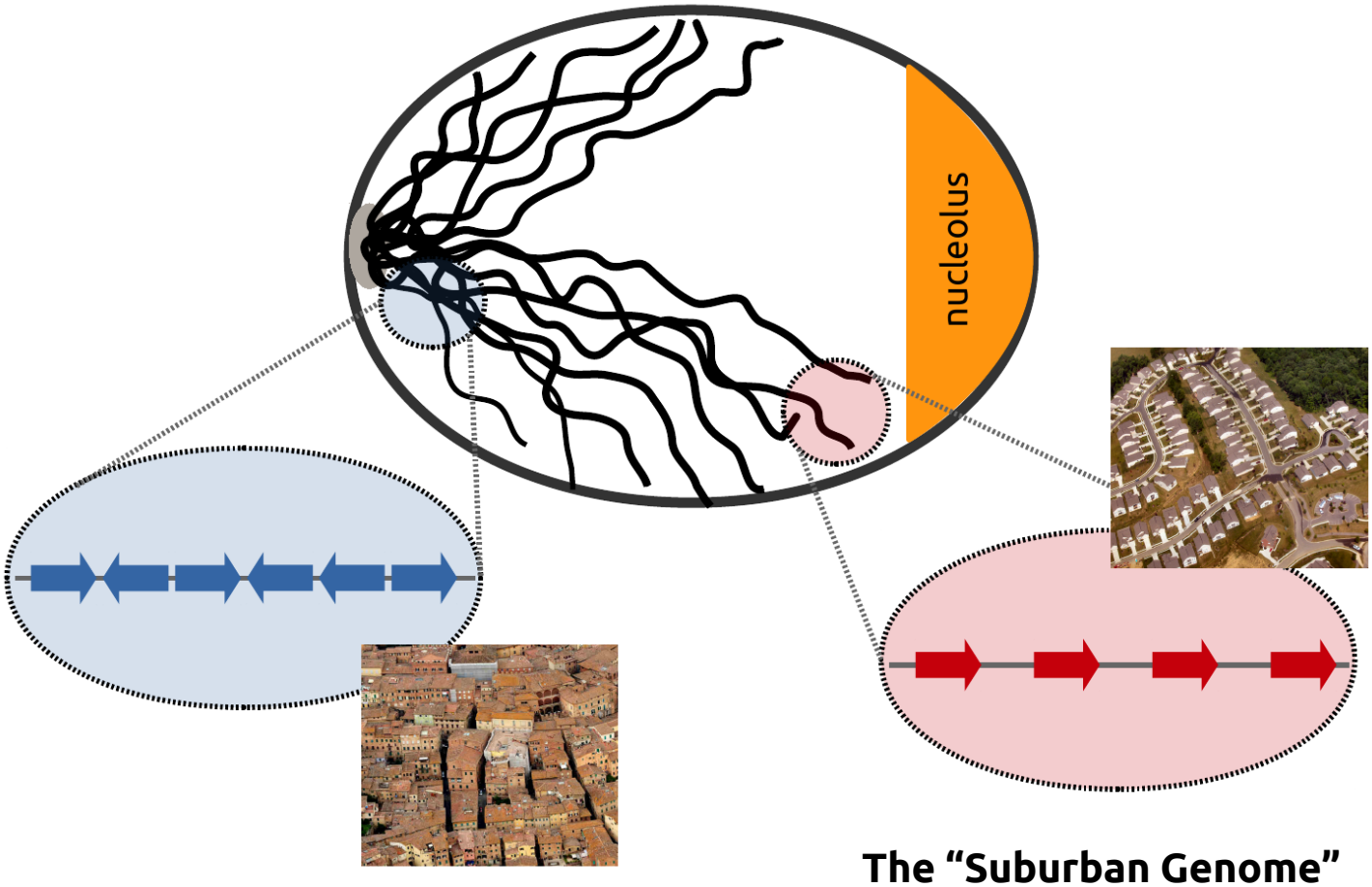


B



bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6



The "Old city center"

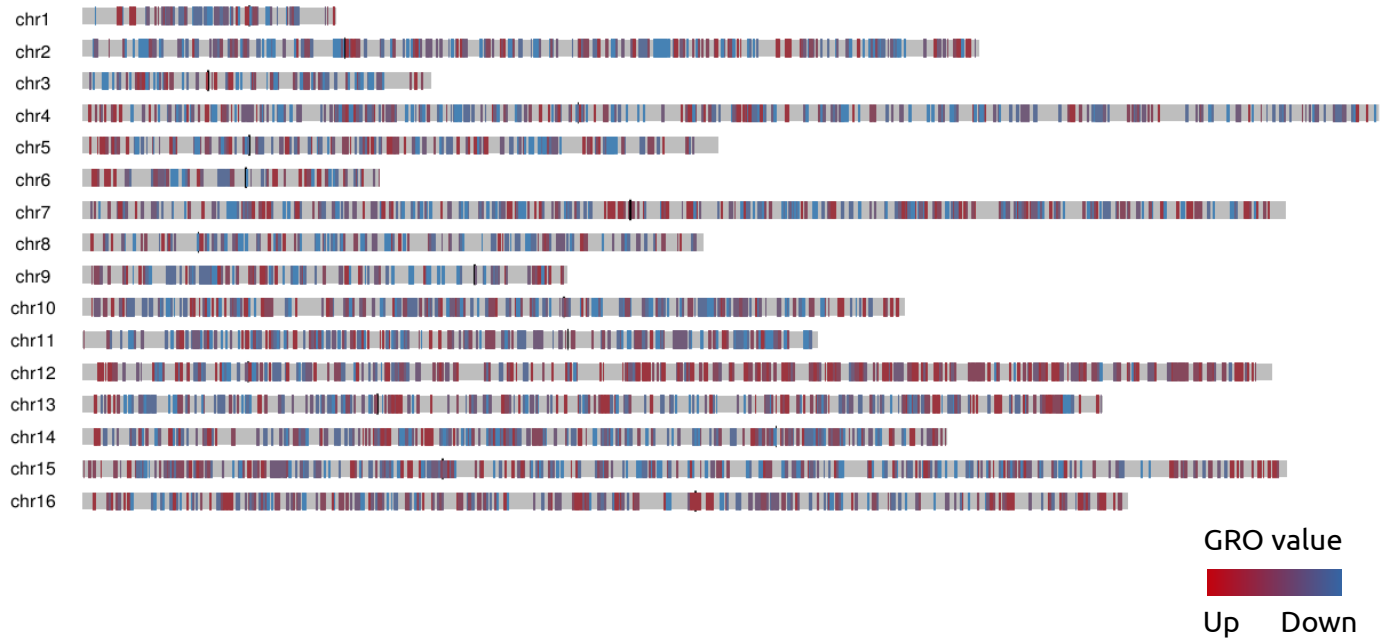
- Ancient, conserved genes
- Constitutive functions
- Simple regulation
- Small intergenic space
- Complex directionality

The "Suburban Genome"

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

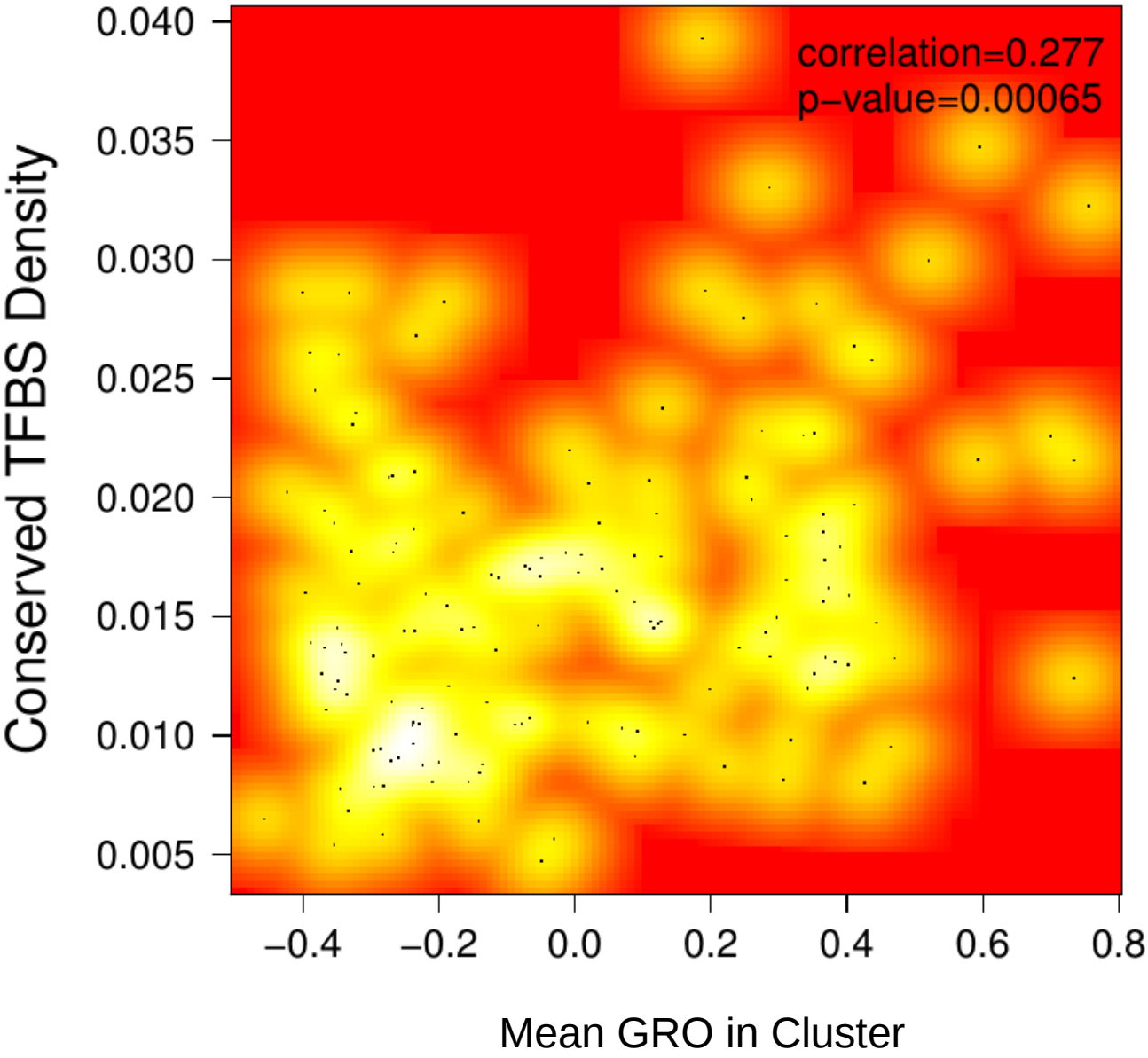
bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

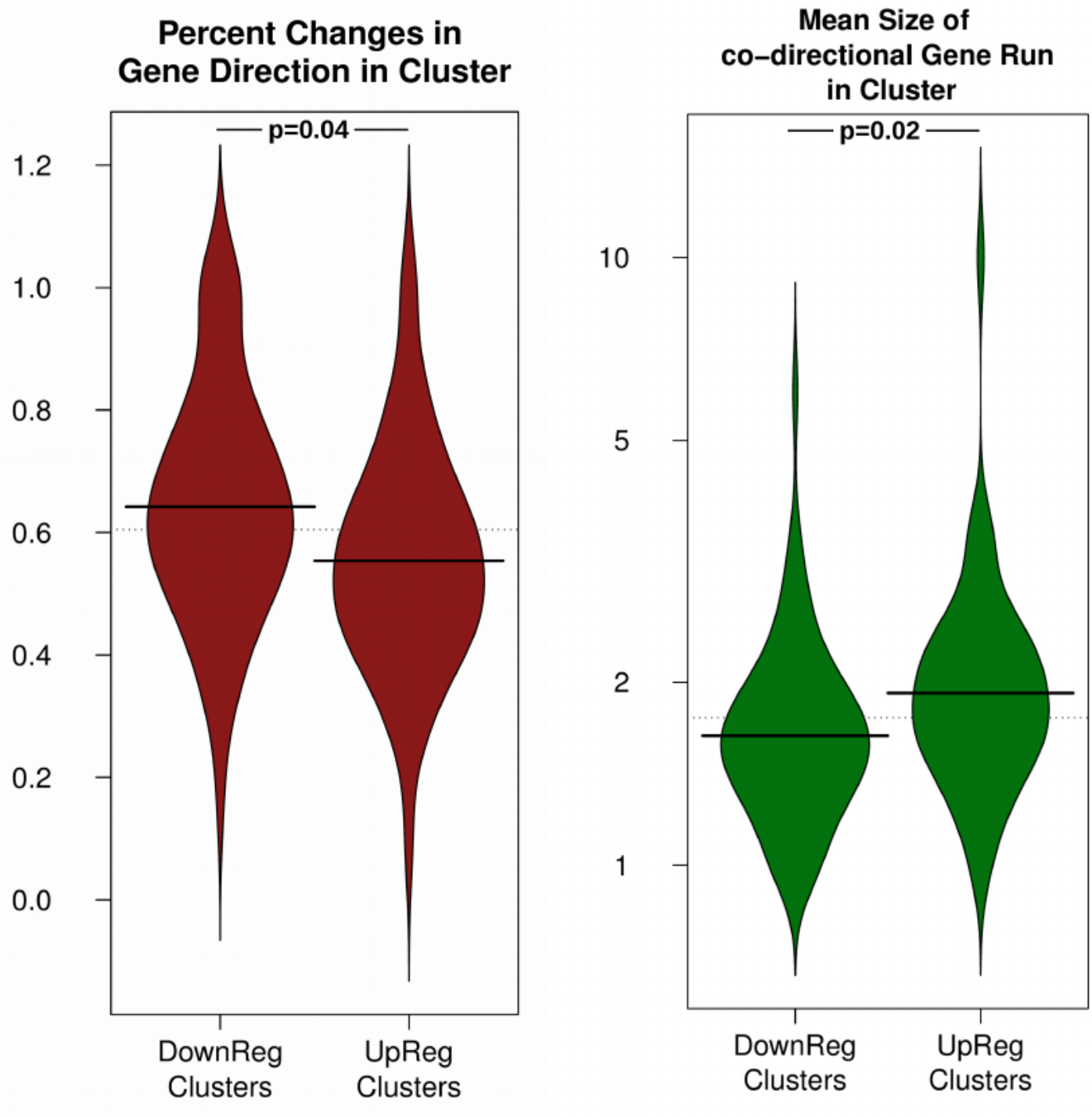
Supplementary Figure 1



bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/064667; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Figure 3