Large-Scale Functional Organization of Long-Range Chromatin Interaction Networks

Kuljeet Singh Sandhu[1,2,*], **Guoliang Li**[1], **Huay Mei Poh**[1], **Yu Ling Kelly Quek**[1,3], **Yee Yen Sia**[1], **Su Qin Peh**[1], **Fabianus Hendriyan Mulawadi**[1], **Joanne Lim**[1], **Mile Sikic**[4,5], **Francesca Menghi**[1], **Anbupalam Thalamuthu**[1], **Wing Kin Sung**[1,6], **Xiaoan Ruan**[1], **Melissa Jane Fullwood**[1,7], **Edison Liu**[1,8], **Peter Csermely**[9], and **Yijun Ruan**[1,8,*]

[1]Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672

[2]Department of Biological Sciences, Indian Institute of Science Education and Research (IISER), Knowledge City, Sector 81, Mohali 140306, India

[3]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St. Lucia 4072, Australia

[4]Bioinformatics Institute, 30 Biopolis Street, Singapore 138671

[5]Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR 10000 Zagreb, Croatia

[6]School of Computing, National University of Singapore, Singapore 117417

[7]A*STAR-Duke-NUS Neuroscience Research Partnership, 8 College Road, Singapore 169857

*Correspondence: sandhuks@iisermohali.ac.in (K.S.S.), ruanyj@gis.a-star.edu.sg (Y.R.) http://dx.doi.org/10.1016/j.celrep.2012.09.022.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.09.022.

WEB RESOURCES

The URLs for data presented herein are as follows:

BiNGO, http://www.psb.ugent.be/cbd/papers/BiNGO/Home.html

Cinteny, http://cinteny.cchmc.org/

Cytoscape, http://www.cytoscape.org/

ECRbase, http://ecrbase.dcode.org/

ENCODE, http://www.genome.gov/10005107

GEO, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33664

Gephi, http://gephi.org/

igraph, http://igraph.sourceforge.net/

Genome Institute of Singapore ENCODE Datasets, http://chiapet.gis.a-star.edu.sg/chiapet/downloads/encode-datasets

Lotka, http://cybermetrics.cindoc.csic.es/articles/lotka102.exe

MGI, http://www.informatics.jax.org/

LINK-group ModuLand, http://www.linkgroup.hu/modules.php

Network Ontology Analysis, http://app.aporc.org/NOA/

PANTHER, http://www.pantherdb.org/

The R Project for Statistical Computing, http://www.r-project.org/

UCSC, http://genome.ucsc.edu/

USC liftOver, http://genome.ucsc.edu/cgi-bin/hgLiftOver

[8]The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

[9]Department of Medical Chemistry, School of Medicine, Semmelweis University, Tuzolto Street 37-47, Budapest 1094, Hungary

SUMMARY

Chromatin interactions play important roles in transcription regulation. To better understand the underlying evolutionary and functional constraints of these interactions, we implemented a systems approach to examine RNA polymerase-II-associated chromatin interactions in human cells. We found that 40% of the total genomic elements involved in chromatin interactions converged to a giant, scale-free-like, hierarchical network organized into chromatin communities. The communities were enriched in specific functions and were syntenic through evolution. Disease-associated SNPs from genome-wide association studies were enriched among the nodes with fewer interactions, implying their selection against deleterious interactions by limiting the total number of interactions, a model that we further reconciled using somatic and germline cancer mutation data. The hubs lacked disease-associated SNPs, constituted a nonrandomly interconnected core of key cellular functions, and exhibited lethality in mouse mutants, supporting an evolutionary selection that favored the nonrandom spatial clustering of the least-evolving key genomic domains against random genetic or transcriptional errors in the genome. Altogether, our analyses reveal a systems-level evolutionary framework that shapes functionally compartmentalized and error-tolerant transcriptional regulation of human genome in three dimensions.

INTRODUCTION

Long-range chromatin interactions are pervasive in the human genome and serve to regulate gene expression (Göndör and Ohlsson, 2009; Schoenfelder et al., 2010). Proximity ligation in combination with next-generation sequencing has recently enabled us to explore genome-wide spatial crosstalk in the chromatin (Fullwood et al., 2009; Lieberman-Aiden et al., 2009). By implementing Chromatin Interaction Analysis using Paired End Tags (ChIA-PET) (Fullwood et al., 2009), we recently mapped all-to-all chromatin interactions associated with RNA polymerase II (RNAPII) at base-pair resolution. In addition to widespread promoter-enhancer chromatin interactions, our analysis revealed a range of distinct types of chromatin cross-wirings, including promoter-enhancer, enhancer-enhancer, promoter-terminator, and, intriguingly, promoter-promoter interactions. These interactions constitute a basic topological template for transcriptional coordination (Li et al., 2012). The observation of most interest was that interacting promoters not only correlate with gene coexpression, but can also regulate each other's transcriptional states, which blurs the traditional definitions of gene-regulatory elements in the genome. These observations support the notion of a chromatin interactome encompassing a dense repertoire of regulatory elements for transcriptional regulation.

Whole-genome chromatin interaction data sets are too complex to analyze by conventional approaches. To gain a better understanding of these interactions, we performed a complex network analysis by integrating chromatin interactions and several other genomic data sets

(Table S1). Network analysis has emerged as a powerful tool for obtaining novel insights into complex systems. The nonrandom topological properties of most real-world networks are strongly associated with their robustness and functional organization (Albert et al., 2000; Barabási and Albert, 1999; Barabási and Oltvai, 2004), which has motivated molecular biologists to explore cellular regulation using a systems approach. Although most cellular networks, such as gene-regulatory, metabolic, protein-protein interaction, and signaling networks, are being widely studied, the extensive communications among regulatory elements in the genome have not been viewed in a complex-network context (Singh Sandhu et al., 2011).

We show that a large proportion of the human genome converges to a complex hierarchical network to orchestrate transcription in functionally compartmentalized and evolutionarily constrained chromatin communities. We demonstrate that the hubs (i.e., nodes with a disproportionately high number of interactions) and spokes (i.e., nodes with fewer interactions) of the network exhibit distinct functional and etiological properties. Together, our findings present a chromatin-level explanation for how disease-associated mutations are tolerated during development and how the key cellular genes maintain their consistent and error-free expression.

RESULTS

Transcription-Associated Chromatin Interactions Form a Complex Hierarchical Network

ChIA-PET is a logical extension of proximity-ligation-based techniques such as chromosomal conformation capture (3C) and circularized 3C (4C). In brief, the chromatin is crosslinked with the use of 1% paraformaldehyde and sonicated, and complexes are pulled down using a specific antibody against a particular protein factor (in this case, 8WG16 antibody against RNAPII). Specific linkers are added to the open ends and the complexes are ligated in the diluted conditions. The ligated material is then subjected to PET extraction and next-generation sequencing (Figure 1A). Using K562 and/or MCF7 ChIA-PET data sets (Li et al., 2012), we constructed an RNAPII-associated chromatin interaction network (ChIN) by denoting the distinct genomic sites as vertices (nodes) and statistically significant (false discovery rate [FDR] < 0.05; Extended Experimental Procedures) chromatin interactions among those sites as edges (links) (Figures 1B and S1A; Extended Experimental Procedures). To remove redundancy from the ChIA-PET data, we merged the neighboring overlapping sites as illustrated in Figures 1B (left panel) and S1A. Several randomly selected intra- (*cis*) and interchromosomal (*trans*) interactions had been validated with Chromosomal Conformation Capture (3C) and DNA fluorescence in situ hybridization (FISH) assays in our earlier study (Li et al., 2012). The topological and functional properties of the ChIN presented in this study were also scrutinized against the artifacts of genomic rearrangements in the MCF7 and K562 cell lines (Figures S1C, S1D, S3C, S4C, S5C, and S6A).

The strategy elaborated in Figure 1 and Extended Experimental Procedures yielded a comprehensive network map of chromatin interactions with ~10,000 connected network components. Surprisingly, however, ~40% of the total nodes formed a giant network component of 36,748 nodes sharing 55,039 links among them (Figure S1E; Tables S2 and

S3), suggesting that the vast majority of the transcriptionally active genome displays widespread communication, and implying an unprecedented level of regulatory influence among genes and their associated genetic elements. This could lead to common pleiotropic gene effects. It is obvious that most of these interactions would not occur at the same time in the same cell due to spatial constraints, but rather represent highly dynamic interactions across a large population of cells (Sandhu et al., 2009).

Except for some of the properties concerning the topology of the network, which were determined using the giant network component, most analyses presented in this study were performed on the whole network map, including the smaller components. Functional analyses were performed on the K562 and/or MCF7 data sets, depending on the availability of other related genomic data sets, although the overall properties were coherent between the two cell lines as shown in Figure S1.

The giant network component of ChIN followed a scale-free-like degree distribution, according to which very few nodes would have a disproportionately large number of interactions, and most others would be weakly connected (Barabási and Albert, 1999) (Figure 2A, top panel). We confirmed this observation for the complete networks for both cell lines, as well as for the publicly available Hi-C data set (Lieberman-Aiden et al., 2009) (Figure S1F). Most real-world networks exhibit scale-free-like behavior (Albert et al., 2000), and the property ascribes error tolerance against random malfunctions, indicating that the ChINs allow for robust systems.

Furthermore, the ChIN displayed a hallmark of hierarchical network topology, characterized by a strong negative correlation between the degree (i.e., number of interactions per locus) and the clustering coefficient (i.e., the tendency of a node to form triangles; Pearson correlation coefficient [PCC] = −0.81; Figure 2A, lower panel). A hierarchical network exhibits high modularity in addition to scale-freeness and is an inherent property of biological networks that governs functional organization (Barabási and Oltvai, 2004). We further illustrated the hierarchical nature of the ChIN in an example in which the long-range (<200 Kb) interaction clusters, namely the HIST1 gene clusters, further converge via super-long-range (>500 Kb) interactions in a hierarchical manner (Figure 2B). The convergence of these three HIST1 clusters correlates significantly with their coexpression (Li et al., 2012), suggesting that super-long-range interactions are important for the global coordination of distant gene clusters. Indeed, we observed that the super-long-range and *trans* chromatin interactions are critical for maintaining the overall ChIN topology, despite having a lower frequency of interactions and accounting for a smaller proportion of all chromatin interactions (Figure S1G). Abolishment of these interactions would break the ChIN into smaller disconnected components and consequently alter the global coordination among distant genes, as in the case of HIST1 clusters.

Strong Regulatory Marks Govern the Modular Topology of the ChIN

To further investigate the modular nature of the ChIN, we used the ModuLand algorithm (Kovács et al., 2010) and mapped the network modules, which we refer to as chromatin communities. A chromatin community can include loci from different parts of the genome and, therefore, represents an extension of the multigene complex (Li et al., 2012) that was

methodically restricted in the genomic range of 1 Mb. The ModuLand algorithm identified 1,173 communities in the giant component of the K562 ChIN (Figure 2C). Most of the intracommunity interactions were enriched with *cis* interactions, whereas intercommunity interactions were mostly mediated by superlong-range or *trans* chromatin interactions (Figure S1H), adhering to the fractal nature of chromatin folding (Lieberman-Aiden et al., 2009; Sexton et al., 2012). We then asked whether distinct genomic elements and chromatin types, as identified by Ernst et al. (2011), could contribute distinctly to the modular topology of the ChIN (Figures 2D and S2). We calculated a centrality score, called the community centrality score, which is a cumulative measure of the influence of the entire network to the given node, and is maximal at the central core of the network modules (Kovács et al., 2010). The active/weak promoter and the strong enhancer elements showed significantly greater community centrality scores than the other categories, suggesting that the modular structure of the ChIN is primarily shaped around these genomic elements (Figures 2D and S2J). Therefore, a single promoter can have multiple enhancers, and a single enhancer can have multiple target promoters, mounting the regulatory complexity of the genome. We further classified the enhancer nodes as private or public enhancers based on their attainment by one or multiple (2) gene promoters, respectively. Interestingly, >70% of the public enhancers were also strong enhancers ($p < 2.2e-16$), whereas private enhancers were equally represented by strong and weak enhancers (Figures 2E and S2K). We examined this observation against the possibility of differential enrichment of RNAPII at strong and weak enhancer sites by restricting the analysis to sites of similar levels of RNAPII enrichment (Figure S2L). A specific example is shown in Figure 2F. Three active promoters, three strong enhancers, and three weak enhancers converged to a network complex. Here again, the active promoters *b*, *d*, and *g* were central to the network segment, undergoing three, five, and five interactions, respectively. Similarly, the strong enhancer *c* interacted with all three active promoters in the locus, whereas the other strong enhancer, *h*, interacted with two of the three active promoters (Figure 2F). On the other hand, the weak enhancers *f*, *i*, and *j* were peripheral to the chromatin communities connecting to individual genes (Figure 2F). Therefore, strong enhancers not only have a greater enhancing effect on transcription, which is the original definition of strong enhancers (Ernst et al., 2011), but also have the potential to regulate multiple genes (pleiotropic regulation).

We then sought a possible explanation for the greater centrality of strong/public enhancers. We assessed the correlations with individual histone modifications (Figure S2E). Hyperacetylation of nodes was associated with higher degree, which is in line with the supposition that hyperacetylation endows greater chromatin mobility (Brown et al., 2008; Krajewski and Becker, 1998). We previously showed that the abundance of chromatin interactions correlates with genomic descriptors such as SINE and LINE densities (Li et al., 2012). Therefore, in our analyses we controlled for these genomic correlates. The partial correlations controlled for SINE and LINE densities clearly suggested that the correlations between degree and SINE/LINE densities do not account for the correlations observed between node degree and enrichment of chromatin marks (Figures S2H and S2I). Furthermore, the elements bound with chromatin remodeling factors such as BRG1 and INI1 were more interactive than the rest (Figure S2M). We hypothesize a prominent role for chromatin-remodeling factors in determining the ChIN topology, which is also in line with

earlier reports on individual loci (Kim et al., 2009; Ni et al., 2008; Zhang et al., 2006). Surprisingly, contrary to the prevailing view on the role of CTCF in chromatin architecture, we did not observe strong association between CTCF binding and the number of RNAPII-associated chromatin interactions (Figures 2D and S2E), suggesting that CTCF orchestrates the genome architecture in a manner reasonably distinct from that of RNAPII, possibly by enclosing the chromatin communities in large chromatin compartments and thus ascribing a basic chromatin skeleton for transcription-associated complex connectivity (G.L. et al., unpublished data). Therefore, the notion that CTCF is the "master-weaver of the genome" needs to be reconciled by taking into account the role of other factors in three-dimensional genome organization.

Chromatin Communities Organize Functional Compartmentalization

The modular nature of the ChIN raises the possibility of functional compartmentalization of chromatin in the nucleus. To assess the functional enrichment in chromatin communities using network-based ontology tools (see Experimental Procedures), we focused only on promoter-promoter interactions. This resulted in the decomposition of the giant network into several smaller network components (Figures S3A and S3B). We analyzed the enrichment of gene ontology (GO) terms among the top 30 network components, containing at least 20 genes each (Extended Experimental Procedures; Table S4). Out of 30 such subnetworks, we observed the enrichment (FDR < 0.01) of one or more functions in 18 (60%; Figure 3A, left panel). Using an example of a network component, we further showed that the enrichments of multiple functions were localized in distinct chromatin communities within a network component (Figure 3A, right panel; Table S4). Figure S3C illustrates that the observed functional organization is not an artifact of genomic rearrangements. We further validated two interesting examples using DNA FISH experiments: (1) A common enhancer interacted with two brain-related proteases, both expressed in MCF7, in *cis* and *trans*. Interestingly, the enhancer locus was specifically conserved among primates, hinting at the possibility of primate-specific gene expression regulated via long-range chromatin interactions (Figure S3D). (2) Two small nuclear noncoding RNA loci were found to be interacting in *trans* (Figure S3E). DNA FISH experiments confirmed the significant interaction frequencies among the loci involved (p = 6.8e-07 and 2.2e-16, respectively; binomial test; Figures S3D and S3E).

Often, not all of the genes in a community served the same function, suggesting that the chromatin communities were not absolutely dedicated to a particular function and often incorporated overlapping secondary functions, which might be indirectly related to the primary function. For instance, a protease (*SIPA1*), a protease inhibitor (*CST6*), and a DNA repair-related factor (*MUS81*) were embedded in a community significantly enriched in defense/immunity-related genes (FDR < 0.01; Figure 3B). We hypothesize that such interactions might orchestrate a coordinated response to external stimuli. The overlapping functional enrichments in chromatin communities could also help in efficient reconfiguration of community function in response to external signals, as proposed earlier (Mihalik and Csermely, 2011; Pál et al., 2006).

Importantly, the community structures were largely conserved between MCF7 and K562 cell lines. Out of 1,783 total gene communities with at least three genes each, 1,279 (71%) showed >75% overlap in MCF7 and K562 cell lines (Figure S3F). Upon closer examination, we observed that the nodes with K562- and MCF7-specific interactions were often embedded in the communities enriched with genes common to both cell lines (Figure S3G). Therefore, the cell-line specificity is defined either by individual long-range transient interactions that do not converge to the same community or by fine level differences in chromatin looping within communities. Such fine differences in chromatin architectures have also been observed by others in different contexts (Filion and van Steensel, 2010; Lienert et al., 2011; Peric-Hupkes et al., 2010).

Chromatin Communities Are Evolutionarily Constrained

We speculated that evolutionary constraints may have shaped the functional compartmentalization of chromatin. To test this, we analyzed the density of interactions (i.e., the number of interactions per Mb) within genomic blocks that were syntenic or nonsyntenic to chimp and mouse genomes. We used a moderate level of coarse graining to map syntenic blocks using the *Cinteny* algorithm (Sinha and Meller, 2007) (Table S1), which revealed human-chimp and human-mouse syntenic blocks covering 48% and 45% of the human genome, respectively. Subsequent analysis revealed a nonrandomly higher density of chromatin interactions within the syntenic blocks than in the nonsyntenic blocks (Figures 4A–4C). Moreover, the frequency of loops connecting syntenic and nonsyntenic blocks was also very low (four and seven loops per megabase for human genomic blocks syntenic to chimp and mouse genomes, respectively). Because syntenic regions are expected to show higher gene density and expression, we performed some control analyses by selecting syntenic and nonsyntenic regions of similar gene density and expression. The analysis consistently showed a significantly higher density of chromatin interactions in the syntenic blocks (Figure S4A). Furthermore, a brief analysis on some of our unpublished ChIA-PET data for mouse embryonic fibroblasts (MEFs) suggests that RNAPII-mediated chromatin interactions tend to accumulate in mouse regions that are syntenic to the human genome (Figure S4B), reconciling the evolutionarily constrained nature of chromatin communities. The observed synteny of chromatin communities was also robust against the possibility of artifacts due to genomic rearrangements in human cancer cell lines (Figure S4C). The above observations were also supported by our analyses of human-mouse orthologous genes, conservation of genomic neighborhood, asynonymous sequence divergence between human-chimp genomes, and mammalian phastCons conservation scores (Figures S4D–S4G).

To obtain further details, we plotted the genomic distance between interacting loci of human genome against that of corresponding sites (UCSC's liftOver; 95% sequence similarity) in the mouse genome (Figure 4B). We had three key observations: (1) Paired coordinates for 32% and 38% of total interactions in K562 and MCF7 cell lines, respectively, could be directly mapped to the mouse genome as compared with 13% of randomly selected coordinates with the same span distribution (p < 2.2e-16). (2) The correlation between locus distances in human and mouse genomes was significantly higher for interacting loci than for randomly selected pairs of loci of the same span distribution, supporting the higher conservation of synteny for interacting pairs (PCC = 0.90, FDR = 0.004; Figures 4B and

S4H). (3) There were very few instances in which proximal mouse genomic sites were rearranged to distant sites in the human genome (Figure 4B), and in contrast, there was nonrandomly higher representation (FDR = 0.009; Figure S4H) of sites that were distant in the mouse genome but were rearranged to proximal domains in the human genome. This suggests two possibilities: (1) The physical interactions, if any, among distant genomic sites in the ancestor genome may have served as an evolutionary mechanism to translocate the interacting loci to proximal regions in the human genome (Figure 4B). Indeed, spatial proximity has been shown to mediate genome rearrangements associated with cancer genomes (Lin et al., 2009). Therefore, the evolution of gene clusters may have been mediated by long-range chromatin interactions. (2) If the distant loci had no interaction in the ancestor genome, then the newly acquired linear proximity of loci through the process of translocation may have been the sole driving force behind chromatin interactions. A detailed analysis of RNAPII-associated chromatin interaction data from other lower species would allow further examination of such observations in the future.

We previously demonstrated a nonrandomly higher correlation among expression profiles of interacting genes across several gene-expression data sets (Li et al., 2012). Along similar lines, we now asked whether the expression of genes with promoter-promoter interactions is evolutionarily more conserved than the rest. To address this issue, we obtained an expression-divergence data set for multiple human and chimpanzee tissues from the literature (Khaitovich et al., 2005). Indeed, the genes that had promoter-promoter interactions showed a significantly lower divergence of gene expression and sequence during the evolutionary split of chimpanzees and humans (Figures 4D and S4F). These observations highlight the strong evolutionary selection of advantageous chromatin communities for functional coordination of related genes.

Disease-Associated Genetic Errors Are Enriched Among Spokes

Genetic errors in distal noncoding elements could influence the expressivity of the genome (Freedman et al., 2011; Mu et al., 2011). One way in which genetic errors could influence gene activity and, consequently, the phenotype is via long-range chromatin interactions (Ahmadiyeh et al., 2010; Ferrai and Pombo, 2009; Sandhu et al., 2009; Steidl et al., 2007; Visel et al., 2009). Therefore, we focused on disease-associated SNP data obtained from the genome-wide Association Studies (GWAS) catalog (Hindorff et al., 2009). The representation of GWAS SNPs among genic and intergenic sites did not differ from that of the overall representation of these sites in the giant ChIN (Figures S5A and S5B), suggesting that disease-associated SNPs are equally probable for genic or intergenic regions in the ChIN. Although the target genes of intergenic GWAS SNPs determined by ChIA-PET showed good correspondence (~70%) with the targets reported in the GWAS catalogue, there were SNPs that had different or additional targets compared with the known ones (Figures S5D and S5E), suggesting that the chromatin interaction data can help one determine the precise targets of noncoding SNPs. Because >95% of total genes engaged in RNAPII-associated long-range interactions were related to the transcriptionally active status of the genes (Li et al., 2012), we reason that if the gene is expressed in the tissue for which the GWAS study was performed, most likely it will also have the corresponding chromatin looping to regulatory elements. This is also supported by the observation that 79% of total

interactions involving genes commonly expressed in the MCF7 and K562 cell lines were conserved among cell lines. Therefore, the cell-lineage discrepancy of integrated GWAS data and the ChIA-PET data sets might not be entirely incoherent in this context.

By mapping the GWAS SNPs onto the ChIN, we further showed that genomic elements with at least one disease-associated SNP were enriched to a lower degree (3–6; spokes) and that the ChIN hubs were devoid of such SNPs (FDR = 0.001; Figure 5A; for examples, see Figures 5C, 5D, S5D, and S5E, and Li et al. [2012]). This observation was also true for distinct types of promoters or enhancer loci (Figure S5F). A relatively weaker second dip in the FDR curve in Figure 5A appeared to be due to a locus having an abnormally high copy number. Therefore, we assessed the representation of all of the nodes with GWAS SNPs in the normal- and abnormal-copy-number regions. More than 90% of the disease-associated SNPs were found to be in the normal-copy-number regions (Figure S5C). Thus, we argue that our observation is not an artifact of the genomic abnormalities of cancer genomes. Furthermore, 80% of the chromatin interactions mediated by the nodes having GWAS SNP were generally restricted within the chromatin community (p = 1.07e-07, Fisher's exact test), suggesting that in general, the chromatin interactions of disease-linked SNPs associate with the spatially localized dysregulation of a limited number of genes.

Similar observations were reported for most disease genes from the morbid entries in the Online Mendelian Inheritance in Man (OMIM) database (Hamosh et al., 2000), where hubs were mostly devoid of genes associated with disease phenotypes (Figure S5G). Therefore, we asked whether disease-associated regulatory loci were selected against the possibility of erroneous interactions, like those reported elsewhere (De and Michor, 2011; Lin et al., 2009) or otherwise, by restricting themselves to fewer interactions. To test this, we compared the chromatin interactions of loci associated with germline and somatic cancer mutations (Futreal et al., 2004). Comparison revealed that the loci harboring cancer-associated germline mutations are less interactive than the ones with somatic mutations, despite the least difference in RNAPII enrichment between the two types of loci (Figures 5B and S5H). The data suggest that disease-associated regulatory elements generally function locally and have a rather limited repertoire of interactions. We propose that the possibility of erroneous genomic interactions and consequently functional dysregulation is minimized by means of restricting the total number of interactions of loci that are important for normal organism development. The germline transmission of genetic lesions having fewer interactions can thus be better tolerated. By contrast, somatic mutations are not under any selection pressure and thus could have a relatively wider exposure to the regulatory cross-wirings in the chromatin. Therefore, based on our analysis, we propose that disease-causing SNPs may generally be trapped in local chromatin communities that affect rather limited phenotypic traits, such as those shown in Figures 5C, 5D, S5D, and S5E.

Hubs Conform to a "Rich-Club" Organization of Key Cellular Functions

Given that the hubs had different characteristics as compared with spokes, we focused on the hubs (degree   60) to determine whether the hubs showed any particular behavior. Interestingly, we found that the top hubs had a preferential link structure, i.e., a rich-club, among themselves (Figure 6A and 6B; total 385 promoters and 2,386 other elements

centered on 25 hubs). By analyzing our in-house data sets of genomic rearrangements uncovered by the genomic DNA paired end tag (G-PET or DNA-PET) sequencing approach (Hillmer et al., 2011), we ensured that the rich-club is not an artifact of genomic rearrangements (Figure S6A).

Rich-clubs in several real-world networks were previously reported and were proposed to contribute to the greater robustness of the network against random hub failures (Colizza et al., 2006; Shi and Mondragon, 2004) and to enhance global cooperation in several biological systems (Bastolla et al., 2009; Saavedra et al., 2011). To test this hypothesis, we disrupted rich-club connectivity and performed a network-resilience analysis of the ChIN. We observed that the ChIN with an intact rich-club had greater topological robustness than the one with a disrupted rich-club (p = 0.004; Figures S6B and S6C).

Functionally, the rich-club genes (n = 385) were enriched in essential cellular functions, including chromatin assembly (e.g., HIST1 genes, TTF2, MTA2, TAF6, and BRD2), cellular organization (e.g., ACTB, ACTG1, CIT, KIF1B, KIF2C, and TRIP6), and primary metabolic processes (e.g., SLC17A7, SlC3A2, ITPA, ATP1A2, DHX29, MAP4K4, EEF1A2, and PLEC1), when compared with nonhub nodes (degree 1–6) or the target genes of GWAS SNPs, which were relatively enriched in development-related functions (FDR < 0.05; Figures 6C, S6D, and S7A). More importantly, 62% of the rich-club genes, and the genes interacting with them, that could be mapped to mouse phenotype information available from the Mouse Genome Informatics (MGI) database (Shaw, 2009) had a "lethal," "death," or "mortality" phenotype in mouse (red-colored nodes in Figure 6D and Table S5), whereas only ~23% of the mouse genes had shown these phenotypes (Shaw, 2009; p = 6.34e-08). A specific example of rich-club organization across an ~14 Mb region on chr1 is shown in Figure 6E. The hubs centered on JARID2 (chromatin modifying), E2F3 (transcription factor), and c6orf62 and HIST1 (chromatin assembly) genes converge via super-long-range interactions. These observations indicate that the hubs collectively perform essential cellular functions by conforming to a rich-club. Nonrandom spatial clustering of essential genomic loci might also relate to an evolutionary strategy to reduce expression noise by locating the essential loci to site(s) of abundant transcription, as discussed below.

DISCUSSION

In brief, we have demonstrated that chromatin interactions form a giant, interconnected network organized into three key interrelated structures: communities, hubs, and peripheral spokes. Communities are primarily centered on hyperacetylated, strong regulatory marks and organize the genome into distinct functional compartments. Hubs conform to an interconnected core, or rich-club, of key cellular functions, whereas spokes are relatively enriched in development-related and lineage-specific genes (Figure 7). This peculiar nonrandom functional organization of hubs and spokes might have evolved to keep the interactome healthy and robust against random deleterious genetic or transcriptional errors in the genome.

Functional enrichment of chromatin communities could be a potent constraint that ties together transcription-associated chromatin in the nucleus. Our observation strongly

supports the notion of specialized transcription factories (Pombo et al., 2000; Xu and Cook, 2008), wherein promoters with common properties, such as binding sites for a particular transcription factor, share the same transcription factory. We reason that the enrichment of secondary functions in the same community might relate to (1) the limitation of available GO knowledge, (2) the transcription of genes that need to be expressed as a coordinated response to external stimuli, or (3) neutral coexpression clusters, which were previously proposed to be a result of neutral coevolution (Sémon and Duret, 2006). Interestingly, the chromatin communities are weakly interconnected with each other through *trans* or super-long-range *cis* interactions, which are generally transient in nature, suggesting that the *trans* chromatin interactions might be critical for cross-functional communication of genes to enable a coordinated response to external signals and allow the genome to easily reconfigure under environmental changes.

Evolutionary conservation of genomic neighborhood, sequence, and gene expression clearly supports the evolutionary constraints of transcription-associated chromatin proximity. This is also in line with a recent Hi-C study on human and mouse embryonic stem cells (Dixon et al., 2012), which appeared when this work was in communication. Interestingly, we observed a population of interacting loci that were distant in the mouse genome but proximal in the human genome, hinting at the possibility of evolution of gene clusters by means of long-range chromatin interactions. Loci that are to be transcribed in a cooperative manner may have been located at genomically distant but spatially proximal sites in the ancestor genome, and may have translocated to proximal genomic sites in higher primates by a similar mechanism that mediates the genomic rearrangements in cancer genomes (Lin et al., 2009). Such hypotheses can be tested further when high-resolution RNAPII-associated chromatin interaction data become available for other mammalian and vertebrate species in the future.

The nonrandom enrichment of disease SNPs and germline mutations among the spokes hints at their selection against diverse chromatin interactions that may be important for controlled regulation of development-related genes. Highly diverse interactions at these loci might increase their susceptibility to erroneous interactions and eventually to transcriptional dysregulation. Moreover, it was previously shown that chromatin interactions can mediate mutations (De and Michor, 2011; Lin et al., 2009). Therefore, the genomic loci with disease-associated mutations would survive through development only if their interactions were limited. The mutations that occur at the hub loci would be lethal and would not be observed in the population.

Nevertheless, the ChIN had hubs that were enriched in key cellular functions. How do these hubs escape random malfunctions? Based on our analysis, we reason that (1) the ChIN follows a scale-free-like distribution of node degrees, which means that the number of hubs would be very low in the network and hence the probability of an error hitting a hub would also be very low; and (2) hubs are not randomly distributed in the ChIN and instead are arranged nonrandomly into an interconnected core or rich-club, which further reduces the probability of being hit by random malfunctioning.

A rich-club of key cellular functions implies two things: (1) In addition to the partial or complete loss of the known protein function, which may or may not explain the lethality, genetic or epigenetic errors in the top hub loci in the ChIN may have other consequences, such as alteration of transcription of other interacting genes through promoter-promoter interactions (Li et al., 2012), followed by a cascading dysregulation of the downstream gene regulatory network, and eventually contribute to lethality. Because the top hubs are directly interconnected through promoter-promoter interactions, we hypothesize that it is this core, rather than a single gene, that becomes transcriptionally dysregulated to cause lethality. 2) Nonrandom rich-club organization in the chromatin interactome may have evolved to shield the genes with key biological functions from random malfunctioning and also ensure their robust, high, and synchronized transcription through promoter-promoter interactions (Li et al., 2012; Figure S2G, degree correlation with gene expression). Indeed, nonrandom linear genomic clustering of essential genes was previously proposed to be associated with lower expression noise (Batada and Hurst, 2007). Along similar lines, non-random three-dimensional (3D) clustering of essential cellular genes at nuclear sites of abundant transcription may regulate their lower expression noise. This clustering can be attributed to 3D organization of gene-dense regions (~70% of hubs were located in regions with >20 genes per Mb) in the nuclear core, which was previously shown to be evolutionarily conserved (Neusser et al., 2007; Tanabe et al., 2002). Therefore, an interconnected core of housekeeping genes might suggest a selection mechanism that evolved to reduce the variation in gene expression at essential gene loci associated with core cellular functions. In contrast, such variations in expression in the peripheral, nonhub nodes associated with lineage-specific and developmental functions may have been relatively tolerant in the context of cell survival. Nonetheless, certain type of genetic errors and dysregulated expression levels at these loci could be lethal in the context of organism survival.

Interestingly, the rich-club remains intact after the genomically rearranged regions are removed from the network, hinting at two possibilities: (1) either the loci in the rich-club are protected against DNA breaks, possibly via efficient DNA repair and a protection mechanism, or (2) the genomic rearrangements at these loci are deleterious in the cell-survival context and hence not observed in the cancer cell lines. The former possibility is also supported by the fact that these loci are hyperacetylated, which also allows for efficient DNA repair (Ikura et al., 2007), and the hub loci generally locate to early replicating domains (Figure S2N), which are less susceptible to genetic errors (Stamatoyannopoulos et al., 2009) due to lower accumulation of single-stranded DNA.

Taken together, our results obtained via a network approach suggest that evolutionary and functional constraints may have shaped the 3D organization of the human genome. We propose that the human genome exhibits a robust systems organization of chromatin interactions to regulate transcription by compartmentalizing biological functions into distinct chromatin communities, and by ensuring the robust and consistent transcription of key essential genes in the interconnected dense core. The modular topology of the chromatin interactome may also guide GWAS studies to prioritize the SNPs for genotype-phenotype associations.

This work also suggests several future directions. First, by integrating the gene regulatory circuitry into the ChIN, investigators may be able to study and predict the erroneous waves a genetic or epigenetic lesion might radiate in a diseased genome. Second, for a full exploration of the emergent properties of chromatin interactome networks that arise over time, the dynamics of chromatin interactions during normal cell-lineage specification and evolution will need to be examined.

EXPERIMENTAL PROCEDURES

Data Sets

We used our previously published RNAPII ChIA-PET data sets (Li et al., 2012) to perform the comprehensive network analysis. These data sets are available from the Gene Expression Omnibus (GEO, GSE33664; Edgar et al., 2002) and from our in-house server. Other genomic data sets were taken from the resources listed in Table S1.

Network Construction

The ChIN was constructed using nonoverlapping distant genomic sites present in our RNAPII ChIA-PET libraries. The detailed strategy for network construction is elaborated in the Extended Experimental Procedures. Nodes were then demarcated as TSS, TES, GBD, and IGN, and as distinct chromatin types using genome annotations from the University of California Santa Cruz (UCSC) and ENCODE (Ernst et al., 2011).

Network Analysis

We used the *igraph* library on the R platform to analyze topological descriptors, such as the node degree (k), average degree of nearest neighbors (knn), average path length, clustering coefficient, and various node or edge centralities of ChIN. To assign communities and their centralities, we used the ModuLand algorithm (Kovács et al., 2010). We detected the rich-club using the recently proposed rich-club coefficient (Colizza et al., 2006). We performed a network resilience analysis by progressively deleting random nodes from the ChIN and measuring the network destruction as a function of the average path length or the number of disconnected network components (Albert et al., 2000).

GO Analysis

We used network ontology analysis (NOA; Wang et al., 2011), BiNGO (Maere et al., 2005), and PANTHER (Mi et al., 2010) to assess the enrichment of specific functions in chromatin communities.

Visualization

Networks were visualized on Cytoscape (Kohl et al., 2011) and Gephi. Spring-embedded layouts were used throughout the analysis. Chromatin loops and associated genomic features were browsed and analyzed on an advance genomic browser developed in-house (F.H.M. et al., unpublished data). Most of the plots were made on the R platform.

Statistics

Statistical tests of significance (i.e., Wilcoxon's rank sum test, Fisher's exact test, and binomial tests) were performed on the R platform. The FDR, when applicable, was calculated by randomizing the data sets several thousand times.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Ahmadiyeh N, Pomerantz MM, Grisanzio C, Herman P, Jia L, Almendro V, He HH, Brown M, Liu XS, Davis M, et al. 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. Proc Natl Acad Sci USA. 2010; 107:9742–9746. [PubMed: 20453196]

Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. Nature. 2000; 406:378–382. [PubMed: 10935628]

Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999; 286:509–512. [PubMed: 10521342]

Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004; 5:101–113. [PubMed: 14735121]

Bastolla U, Fortuna MA, Pascual-García A, Ferrera A, Luque B, Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. Nature. 2009; 458:1018–1020. [PubMed: 19396144]

Batada NN, Hurst LD. Evolution of chromosome organization driven by selection for reduced gene expression noise. Nat Genet. 2007; 39:945–949. [PubMed: 17660811]

Brown CR, Kennedy CJ, Delmar VA, Forbes DJ, Silver PA. Global histone acetylation induces functional genomic reorganization at mammalian nuclear pore complexes. Genes Dev. 2008; 22:627–639. [PubMed: 18316479]

Colizza V, Flammini A, Serrano MA, Vespignani A. Detecting rich-club ordering in complex networks. Nat Phys. 2006; 2:110–115.

De S, Michor F. DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nat Biotechnol. 2011; 29:1103–1108. [PubMed: 22101487]

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012; 485:376–380. [PubMed: 22495300]

Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:207–210. [PubMed: 11752295]

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43–49. [PubMed: 21441907]

Ferrai C, Pombo A. 3D chromatin regulation of Sonic hedgehog in the limb buds. Dev Cell. 2009; 16:9–11. [PubMed: 19154714]

Filion GJ, van Steensel B. Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells. Nat Genet. 2010; 42:4. author reply 5–6. [PubMed: 20037608]

Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, et al. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43:513–518. [PubMed: 21614091]

Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009; 462:58–64. [PubMed: 19890323]

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. A census of human cancer genes. Nat Rev Cancer. 2004; 4:177–183. [PubMed: 14993899]

Göndör A, Ohlsson R. Chromosome crosstalk in three dimensions. Nature. 2009; 461:212–217. [PubMed: 19741702]

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. Cell. 2011; 145:622–634. [PubMed: 21549415]

Hamosh A, Scott AF, Amberger J, Valle D, McKusick VA. Online Mendelian Inheritance in Man (OMIM). Hum Mutat. 2000; 15:57–61. [PubMed: 10612823]

Hillmer AM, Yao F, Inaki K, Lee WH, Ariyaratne PN, Teo AS, Woo XY, Zhang Z, Zhao H, Ukil L, et al. Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes. Genome Res. 2011; 21:665–675. [PubMed: 21467267]

Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci USA. 2009; 106:9362–9367. [PubMed: 19474294]

Ikura T, Tashiro S, Kakino A, Shima H, Jacob N, Amunugama R, Yoder K, Izumi S, Kuraoka I, Tanaka K, et al. DNA damage-dependent acetylation and ubiquitination of H2AX enhances chromatin dynamics. Mol Cell Biol. 2007; 27:7028–7040. [PubMed: 17709392]

Khaitovich P, Hellmann I, Enard W, Nowick K, Leinweber M, Franz H, Weiss G, Lachmann M, Pääbo S. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. Science. 2005; 309:1850–1854. [PubMed: 16141373]

Kim SI, Bresnick EH, Bultman SJ. BRG1 directly regulates nucleosome structure and chromatin looping of the alpha globin locus to activate transcription. Nucleic Acids Res. 2009; 37:6019–6027. [PubMed: 19696073]

Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. Methods Mol Biol. 2011; 696:291–303. [PubMed: 21063955]

Kovács IA, Palotai R, Szalay MS, Csermely P. Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PLoS ONE. 2010; 5:5.

Krajewski WA, Becker PB. Reconstitution of hyperacetylated, DNase I-sensitive chromatin characterized by high conformational flexibility of nucleosomal DNA. Proc Natl Acad Sci USA. 1998; 95:1540–1545. [PubMed: 9465051]

Li G, Fullwood MJ, Xu H, Mulawadi FH, Velkov S, Vega V, Ariyaratne PN, Mohamed YB, Ooi HS, Tennakoon C, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 2010; 11:R22. [PubMed: 20181287]

Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. [PubMed: 22265404]

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]

Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schübeler D. Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS Genet. 2011; 7:e1002090. [PubMed: 21655081]

Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK, Rosenfeld MG. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. Cell. 2009; 139:1069–1083. [PubMed: 19962179]

Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics. 2005; 21:3448–3449. [PubMed: 15972284]

Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res. 2010; 38(Database issue):D204–D210. [PubMed: 20015972]

Mihalik A, Csermely P. Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. PLoS Comput Biol. 2011; 7:e1002187. [PubMed: 22022244]

Mu XJ, Lu ZJ, Kong Y, Lam HY, Gerstein MB. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. Nucleic Acids Res. 2011; 39:7058–7076. [PubMed: 21596777]

Neusser M, Schubel V, Koch A, Cremer T, Müller S. Evolutionarily conserved, cell type and species-specific higher order chromatin arrangements in interphase nuclei of primates. Chromosoma. 2007; 116:307–320. [PubMed: 17318634]

Ni Z, Abou El Hassan M, Xu Z, Yu T, Bremner R. The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers. Nat Immunol. 2008; 9:785–793. [PubMed: 18500344]

Pál C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD. Chance and necessity in the evolution of minimal metabolic networks. Nature. 2006; 440:667–670. [PubMed: 16572170]

Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell. 2010; 38:603–613. [PubMed: 20513434]

Pombo A, Jones E, Iborra FJ, Kimura H, Sugaya K, Cook PR, Jackson DA. Specialized transcription factories within mammalian nuclei. Crit Rev Eukaryot Gene Expr. 2000; 10:21–29. [PubMed: 10813391]

Saavedra S, Stouffer DB, Uzzi B, Bascompte J. Strong contributors to network persistence are the most vulnerable to extinction. Nature. 2011; 478:233–235. [PubMed: 21918515]

Sandhu KS, Shi C, Sjölinder M, Zhao Z, Göndör A, Liu L, Tiwari VK, Guibert S, Emilsson L, Imreh MP, Ohlsson R. Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. Genes Dev. 2009; 23:2598–2603. [PubMed: 19933149]

Schoenfelder S, Clay I, Fraser P. The transcriptional interactome: gene expression in 3D. Curr Opin Genet Dev. 2010; 20:127–133. [PubMed: 20211559]

Sémon M, Duret L. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Mol Biol Evol. 2006; 23:1715–1723. [PubMed: 16757654]

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell. 2012; 148:458–472. [PubMed: 22265598]

Shaw DR. Searching the Mouse Genome Informatics (MGI) resources for information on mouse biology from genotype to phenotype. Curr Protoc Bioinformatics. 2009; Chapter 1(Unit 1.7)

Shi Z, Mondragon RJ. Redundancy and robustness of AS-level Internet topology and its models. Electron Lett. 2004; 40:151–152.

Singh Sandhu K, Li G, Sung WK, Ruan Y. Chromatin interaction networks and higher order architectures of eukaryotic genomes. J Cell Biochem. 2011; 112:2218–2221. [PubMed: 21520242]

Sinha AU, Meller J. Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. BMC Bio-informatics. 2007; 8:82.

Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009; 41:393–395. [PubMed: 19287383]

Steidl U, Steidl C, Ebralidze A, Chapuy B, Han HJ, Will B, Rosenbauer F, Becker A, Wagner K, Koschmieder S, et al. A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. J Clin Invest. 2007; 117:2611–2620. [PubMed: 17694175]

Tanabe H, Müller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci USA. 2002; 99:4424–4429. [PubMed: 11930003]

Visel A, Rubin EM, Pennacchio LA. Genomic views of distant-acting enhancers. Nature. 2009; 461:199–205. [PubMed: 19741700]

Wang J, Huang Q, Liu ZP, Wang Y, Wu LY, Chen L, Zhang XS. NOA: a novel Network Ontology Analysis method. Nucleic Acids Res. 2011; 39:e87. [PubMed: 21543451]

Xu M, Cook PR. Similar active genes cluster in specialized transcription factories. J Cell Biol. 2008; 181:615–623. [PubMed: 18490511]

Zhang Y, Smith CL, Saha A, Grill SW, Mihardja S, Smith SB, Cairns BR, Peterson CL, Bustamante C. DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. Mol Cell. 2006; 24:559–568. [PubMed: 17188033]
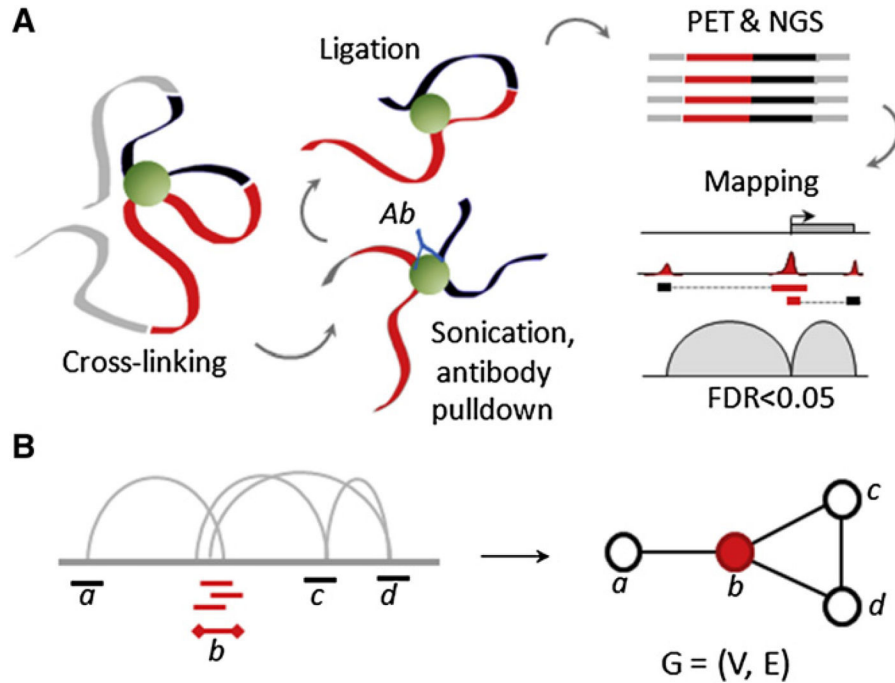
Figure 1. Simplified Illustration of the ChIA-PET Technique and Network Construction
(A) ChIA-PET technique.
(B) Network construction.
Formaldehyde crosslinked chromatin was fragmented by sonication and chromatin complexes bound with RNAPII (green), and pulled down using 8WG16 antibody (blue). Specific linkers were added to the open ends and subsequently ligated in the diluted conditions. After the chromatin complexes were decrosslinked, DNA material was subjected to PET extraction and next-generation sequencing using the Illumina GAIIx platform. Unique PETs were mapped back to the reference genome (Hg19) and statistically significant interactions were called at FDR 0.05 using the ChIA-PET tool (Li et al., 2010). To construct the network, the redundancy in the data were removed by merging the overlapping interaction sites. The cutoff range for overlap is shown in Figure S1A.
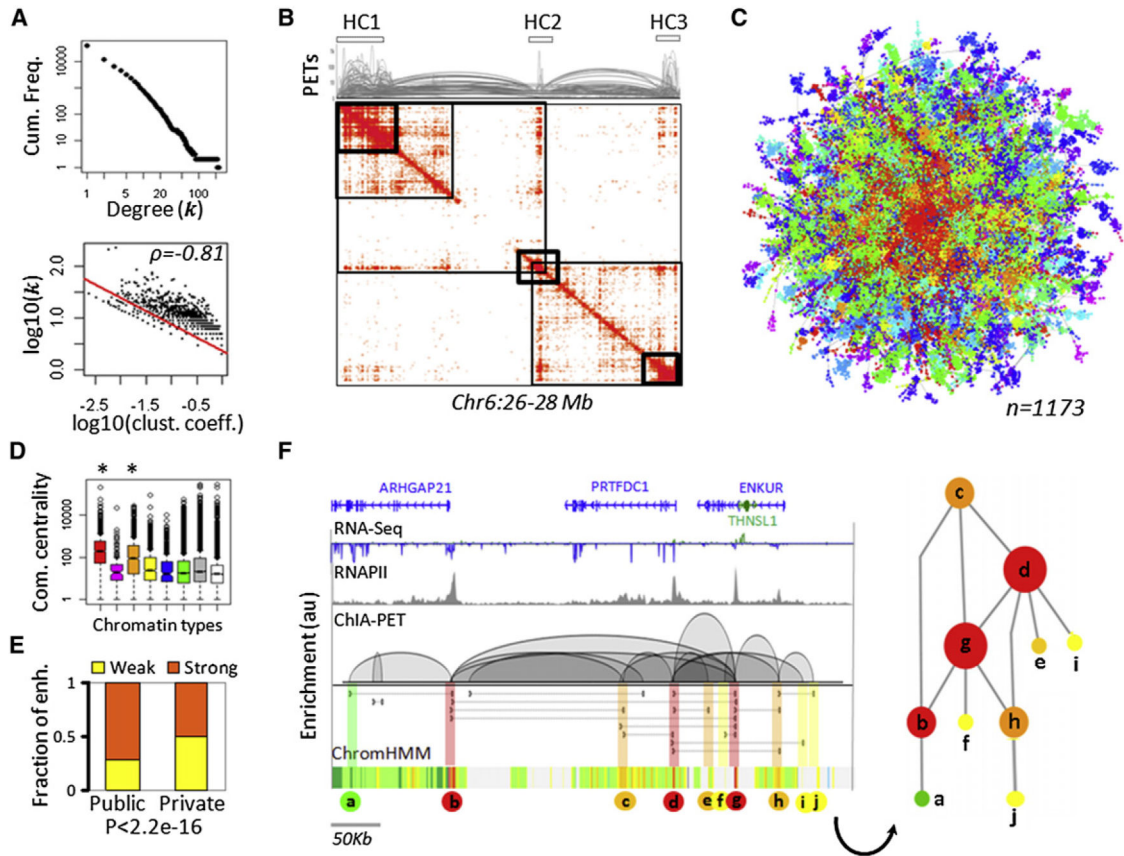
Figure 2. Topological Properties of Transcription-Associated ChIN

(A) Top panel: Log-log plot of the node degree distribution for ChIN constructed from the K562 ChIA-PET dataset (plots for other data sets are given in Figure S1F). The plot shows heavy tailed distribution atypical of scale-free-like networks. Bottom panel: Log-log scatter plot of node clustering coefficients and degree. The strong inverse correlation is atypical of hierarchical (scale-free + modular) networks.

(B) An example of hierarchical chromatin organization on chromosome 6 is shown in the right panel. Three distant HIST1 gene clusters (HC1, HC2, and HC3) converge in a hierarchical manner, as shown in the heat-map representation of ChIA-PET data.

(C) Hierarchical topology of K562 ChIN. The color gradient represents the hierarchical organization of chromatin communities (#1173); red and blue indicate the most central and most peripheral chromatin communities, respectively, as identified by the ModuLand algorithm. Other network properties for K562 and MCF7 ChINs are given in Figures S1 and S2.

(D) Community centralities of nodes having distinct chromHMM profiles (Ernst et al., 2011) in the K562 cell line. Red, active/weak promoter; magenta, poised promoter; orange, strong enhancer; yellow, weak enhancer; blue, insulator; green, transcribed region; gray, repressed region; white, all. Asterisks indicate the chromatin types for which the community centralities were significantly greater when compared with all nodes. The Mann-Whitney U test was used to calculate the p values (<2.2e-16 for each asterisk). See also Figure S2J.

(E) Public and private nature of enhancers. Shown is the bar plot for the proportion of strong and weak enhancers in private and public enhancers. The p value was calculated using Fisher's exact test. A box-plot representation of overall promoter interactions of strong and weak enhancers is given in Figure S2K.

(F) High-resolution example of physical interactions among distinct chromatin types. Shown are the tracks for UCSC known genes, RNA-Seq, RNAPII enrichment, ChIA-PET, and chromatin types (chromHMM) in the K562 cell line. The network constructed from this locus is shown in the right panel. Nodes are colored according to their chromatin types and their size is scaled to their degree.
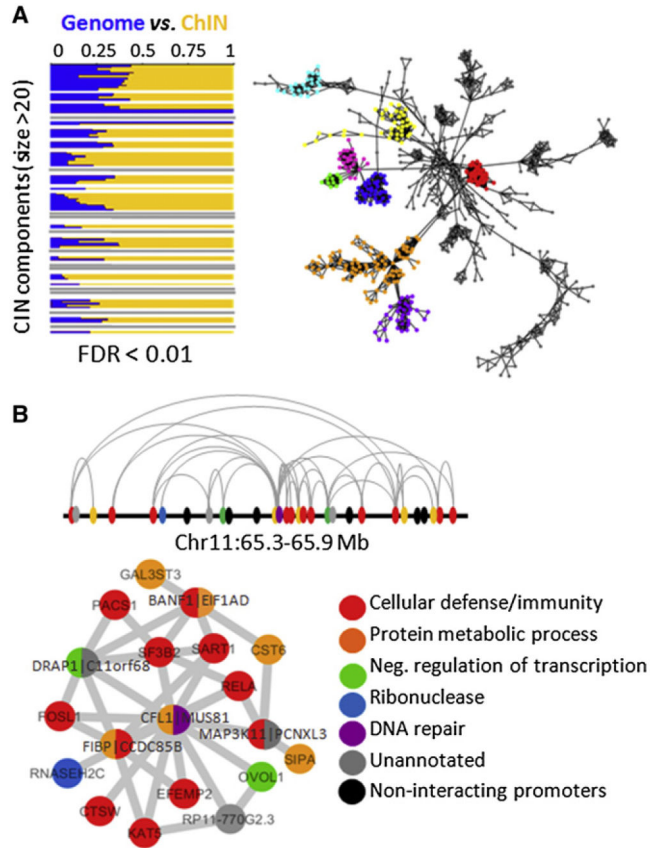
Figure 3. Functional Compartmentalization of Promoter-Promoter Interactions in ChIN

(A) Enrichment of GO process terms in the top 30 network components (size > 20 genes) in MCF7 ChIN compared with the whole genome. Each row separated by a gap represents a network component. Enrichment of GO terms is represented as scaled proportions of the observed number of hits in a ChIN component (orange) versus the expected number of hits in the genome (blue). Only significant terms (FDR < 0.01) are shown. Gray-colored bars represent the network components that had no significant GO enrichment. The network topology shown in the right panel represents a ChIN component (n = ~600) with distinct chromatin communities as determined by the Modu-Land algorithm. Nearly 50% of the communities were enriched with distinct functions (colored modules). Red, chromatin assembly; blue, response to stimuli and RNA processing; orange, lymphocyte-mediated immunity; magenta, fatty acid biosynthesis; pink, antigen processing and presenting; green, brain development; yellow, muscle filament sliding; cyan, proline biosynthesis; gray, no functional enrichment or only one representative gene. Edges are weighted by the PCC of the interacting genes across estrogen-induced time-course GRO-seq experiments (Hah et al., 2011).

(B) Example of chromatin interactions among cellular defense/immunity related genes on chromosome 11. Noninteracting genes (black) have unrelated functions such as testes-specific function and intracellular trafficking. Nodes with dual colors represent neighboring genes with bidirectional promoters. DNA FISH validations of a few interesting examples are given in Figures S3D and S3E.
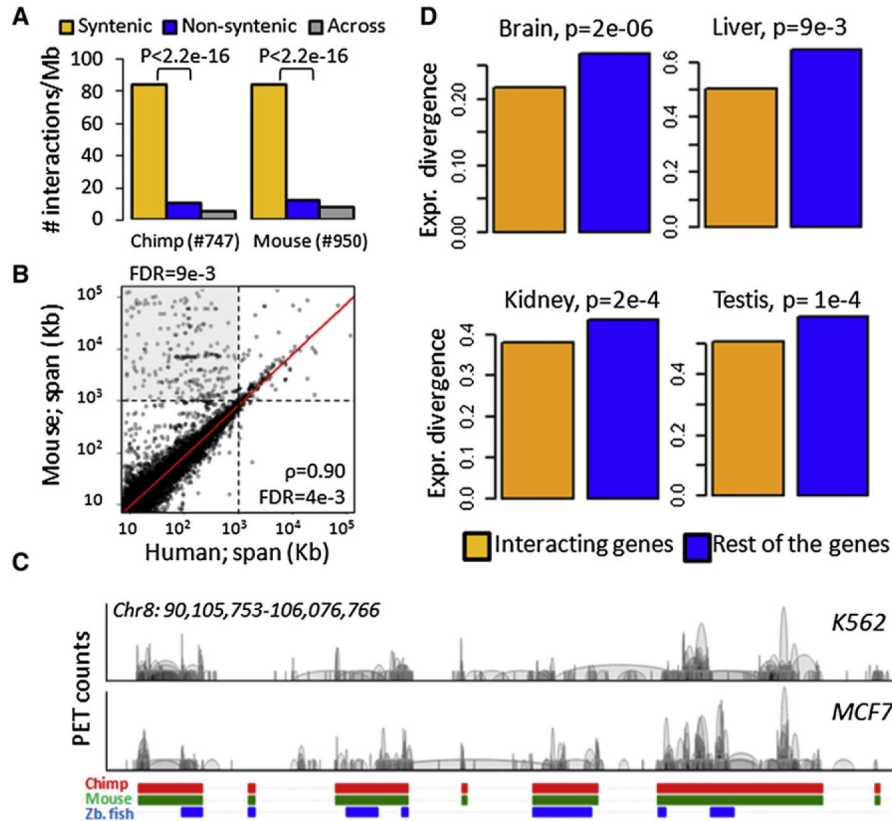
Figure 4. Evolutionary Constraints of Chromatin Communities

(A) Enrichment of chromatin interactions (K562) in the genomic blocks that were syntenic between human and chimp/mouse genomes. The p values for the difference between syntenic and non-syntenic blocks were calculated using binomial tests. The data for MCF7 are given in Figure S4C.

(B) Scatter plot of genomic spans between interacting sites in human and corresponding sites in the mouse genome. The red line represents the linear regression (PCC = 0.90) and dashed lines mark the 1 Mb span. The top-left quadrant, highlighted in gray, represents the genomic sites that are distant in the mouse genome but proximal in the human genome.

(C) Example of an ~10.5 Mb region on chromosome 8, illustrating the preferred interactions within syntenic blocks. Red, green, and blue bars represent blocks in the human genome that are syntenic with chimp, mouse, and zebra fish genomes, respectively. Only those interactions that span the 10.5 Mb region are shown.

(D) Mean divergence in expression (human versus chimp) across different tissues for genes having promoter-promoter interactions (orange) and the rest of the genes (blue). Additional supporting data are given in Figure S4. The p values were calculated using the Mann-Whitney U test.
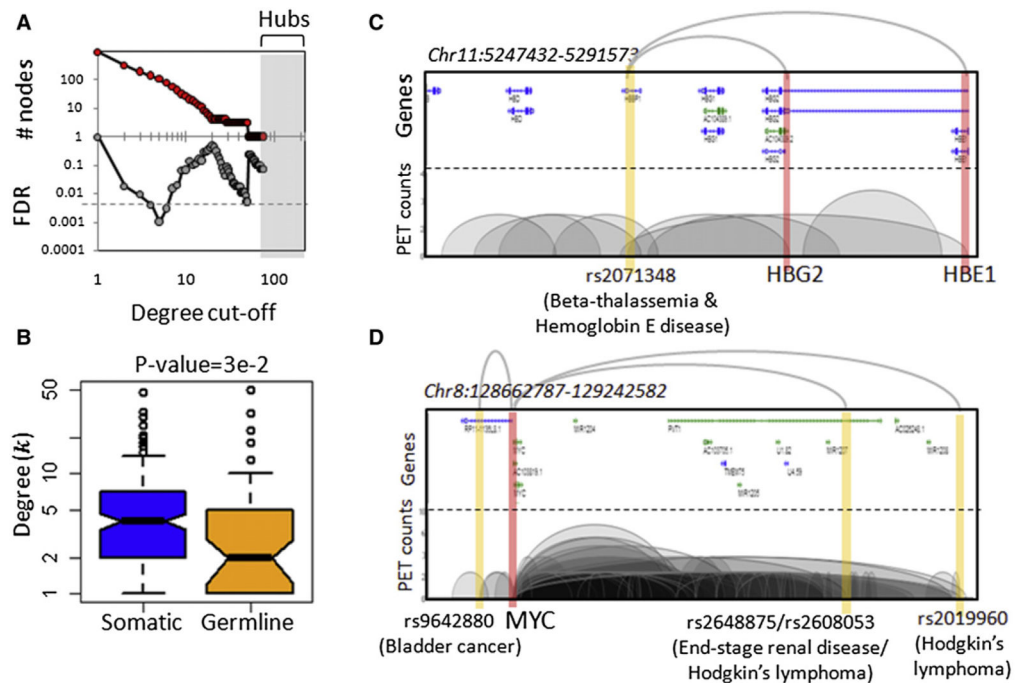
Figure 5. Disease-Associated Mutations in the ChIN

(A) Number of nodes (red) containing at least one GWAS SNP with respect to the degree cutoff. FDRs (gray) were calculated by randomizing the SNP position in the network 10^3 times at each degree cutoff.

(B) Degree distribution of TSS nodes for genes having somatic and germline mutations in cancer phenotypes. The p value was calculated using the Mann-Whitney U test.

(C and D) Examples of "within-community" chromatin interactions of disease-associated noncoding SNPs at the (C) beta-hemoglobin and (D) MYC locus in the K562 cell line. Yellow bars highlight the noncoding SNP positions, and red color highlights the locations of interacting promoters. Phenotypes associated with SNPs: rs2071348, beta-thalassemia/hemoglobin E disease; rs9642880, bladder cancer; rs2648875, end-stage renal disease; and rs2608053/rs2019960, Hodgkin lymphoma. Both of the example loci fall in the normal-copy-number range in the K562 cell line (as illustrated in Figure S1C).
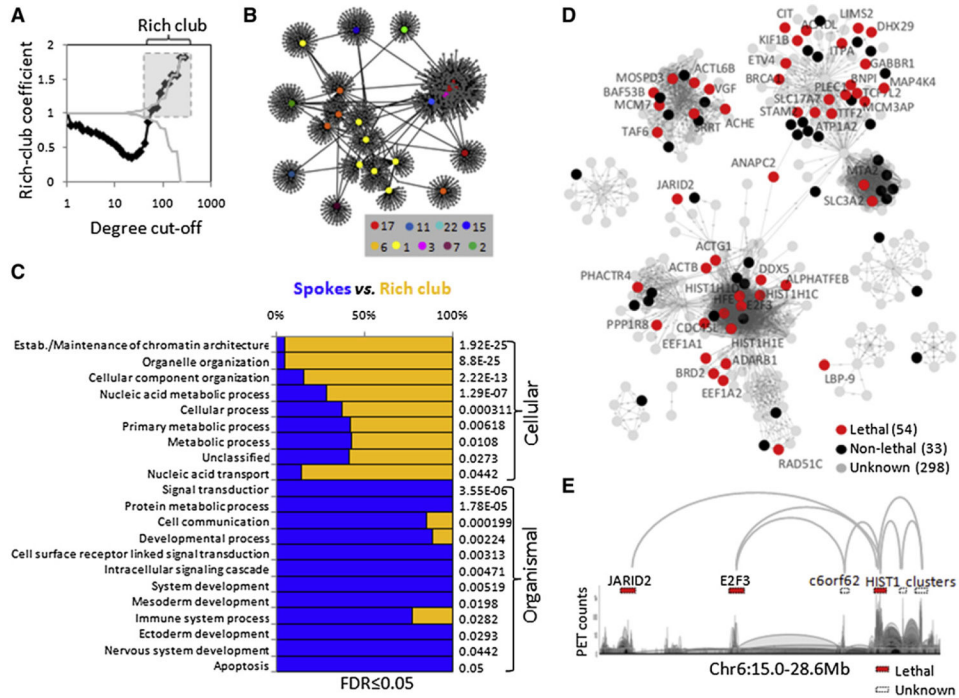
Figure 6. Rich-Club Organization of Key Cellular Functions in the ChIN

(A) XY plot of rich-club coefficient versus degree. Coefficient values > 1, for high degree nodes, signify the presence of a rich-club, i.e., a network core of interconnected hubs. The black curve represents the rich-club coefficient ($\sigma(k)/\sigma_{ran}(k)$; Extended Experimental Procedures), and gray curves represent the 95% confidence interval based on the distribution of rich-club coefficient values of 10^3 randomly rewired networks.

(B) Network representation of rich-club (total 25 hub loci linked to 385 promoters and 2,386 other genomic elements). Distinct colors of the hub nodes signify distinct chromosomes. For simplicity, only the links that connect to hub loci are shown.

(C) Overrepresented GO terms among the rich-club loci (orange) with respect to spoke loci (blue). The p values are corrected for multiple-hypothesis testing using the Benjamini-Hochberg method. GO comparisons with genes associated with GWAS SNPs in genic and intergenic regions are shown in Figures S6D and S7A.

(D) Promoter-promoter interactions (n = 3,933) among all of the genes (n = 385) in the rich-club. The red color represents the genes that showed a lethal/death/mortality phenotype (n = 54) in the MGI database. Black nodes depict the genes that had a nonlethal phenotype (n = 33) in MGI. Gray-colored nodes are genes for which mutation information was not available in MGI (n = 298). (E) An example locus (~14 Mb genomic spans) showing super-long-range chromatin interactions among hubs. Red and white bars represent lethal and unknown phenotypes of the representative hub loci (within the local cluster), respectively. Only those interactions that span the 14 Mb region are shown.
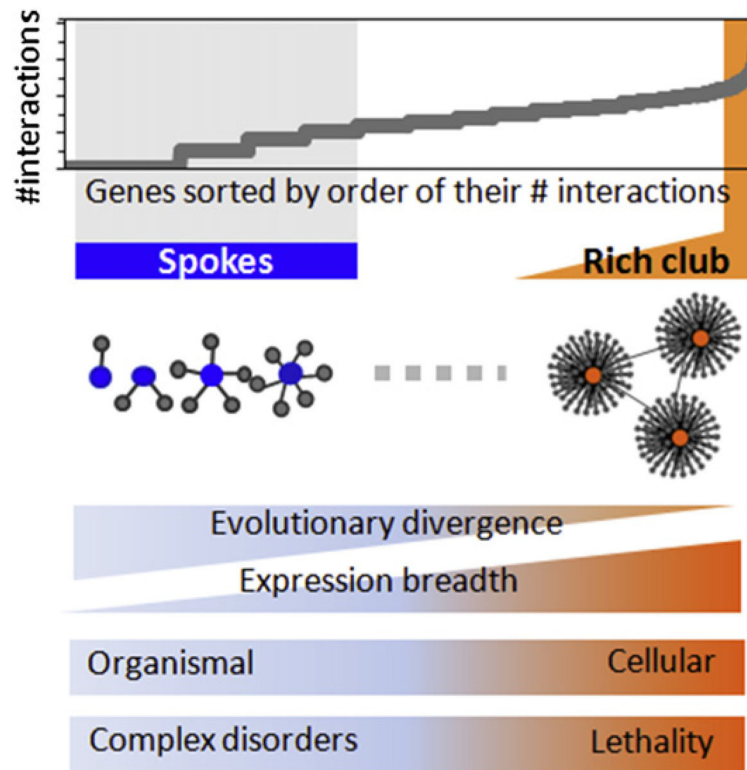
Figure 7. Hubs and Spokes Demarcation of Transcription-Associated ChIN

The grey curve shows the nodes (genes) sorted in order of their number of interactions. The blue bar represents the interaction range of spokes (degrees 1–6) and the orange bar represents the rich-club. Relative functional interpretations are given at the bottom. The original plot for expression breadth analysis is given in Figure S2O. Plots for evolutionary divergence of sequence and gene expression are given in Figures S7B and S7C.