Published online 20 January 2012 Nucleic Acids Research, 2012, Vol. 40, No. 9 3849–3855
doi:10.1093/nar/gks012

On the assessment of statistical significance of three-dimensional colocalization of sets of genomic elements

Daniela M. Witten[1,]* and William Stafford Noble[2,3,]*

[1]Department of Biostatistics, [2]Department of Genome Sciences and [3]Department of Computer Science and Engineering, University of Washington, Seattle, WA 98109

Received November 15, 2011; Revised December 23, 2011; Accepted December 26, 2011

ABSTRACT

A growing body of experimental evidence supports the hypothesis that the 3D structure of chromatin in the nucleus is closely linked to important functional processes, including DNA replication and gene regulation. In support of this hypothesis, several research groups have examined sets of functionally associated genomic loci, with the aim of determining whether those loci are statistically significantly colocalized. This work presents a critical assessment of two previously reported analyses, both of which used genome-wide DNA–DNA interaction data from the yeast *Saccharomyces cerevisiae*, and both of which rely upon a simple notion of the statistical significance of colocalization. We show that these previous analyses rely upon a faulty assumption, and we propose a correct non-parametric resampling approach to the same problem. Applying this approach to the same data set does not support the hypothesis that transcriptionally coregulated genes tend to colocalize, but strongly supports the colocalization of centromeres, and provides some evidence of colocalization of origins of early DNA replication, chromosomal breakpoints and transfer RNAs.

INTRODUCTION

Recently, three published studies have used generalizations of chromosome conformation capture (3C) (1) to obtain genome-wide DNA–DNA interaction data for the genomes of human (2), budding yeast (3) and fission yeast (4). Such methods, coupled with complementary experimental assays such as fluorescence *in situ* hybridization (FISH) (5), DNA adenine methyltransferase identification (DamID) (6) and chromatin interaction analysis by paired end tag sequencing (ChIA-PET) (7), promise to provide an increasingly detailed picture of the 3D structure of chromatin *in vivo*.

Ultimately, the widespread and growing interest in the experimental characterization of chromatin structure is driven by the underlying hypothesis that the structure of DNA in the nucleus is tightly related to DNA function. Experimental evidence supports the existence of a variety of well-defined nuclear substructures, including the nuclear lamina, nucleoli, PML and Cajal bodies and nuclear speckles (8). Furthermore, in some genomes, extensive evidence suggests the existence of relatively well-defined chromosome territories, as well as the systematic orientation of gene-poor, heterochromatic regions near the nuclear periphery and gene-dense, euchromatic regions in the nuclear interior (9). Strikingly, the overall pattern of nuclear architecture varies systematically among cell types yet shows evidence of evolutionary conservation (10). Finally, increasing evidence couples the dynamic repositioning of genomic regions with the regulation of gene expression [reviewed in (8)].

In this article, we do not argue against the hypothesis that chromatin structure is coupled with genome function. However, we do present a cautionary tale illustrating a potential statistical pitfall in the search for connections between gene function and genome structure. In particular, we investigate two recent claims about nuclear colocalization of functional elements in the budding yeast *Saccharomyces cerevisiae*. The first article, published in *Nature* and coauthored by one of us (Noble), claims that there are extensive interchromosomal interactions between transfer RNA genes, centromeres, chromosomal breakpoints, origins of early DNA replication and sites where chromosomal breakpoints occur (3). The second, published in *Nucleic Acids Research*, claims that many transcription factors regulate genes that are colocalized

*To whom correspondence should be addressed. Tel: +1 206 543 8930; Fax: +1 206 685 7301; Email: noble@gs.washington.edu
Correspondence may also be addressed to Daniela M. Witten. Tel: +206 616 7182; Fax: +206 543 3286; Email: dwitten@u.washington.edu

in the nucleus (11). We show here that the statistical test employed by both sets of authors rests upon a faulty assumption, and we illustrate the effect of this faulty assumption via simulations and via reanalysis of the yeast data. Finally, we propose a correct resampling approach to the same problem, and we apply this non-parametric procedure to the same data. Our reanalysis moderately impacts the conclusions from Duan *et al.* (3): contrary to the initial analysis, we do not observe evidence of telomere colocalization; however, we do observe strong evidence for the colocalization of centromeres, as well as statistical support for the colocalization of chromosomal breakpoints, transfer RNAs and origins of early DNA replication. In contrast, the resampling analysis does not provide support for the central claim in the Dai and Dai (11) paper, namely, that transcriptionally coregulated genes tend to colocalize. The statistical analysis of three-dimensional genome structure data sets must be performed with care in order to avoid being misled by the inherent structure of such data.

MATERIALS AND METHODS

Yeast interaction data and functional elements

We obtained from a recent study (3) a list of yeast interchromosomal interactions observed at a false discovery rate below 0.01, obtained by measuring interactions among 3991 segments (chromosomal loci flanked by pre-defined restriction enzyme sites) distributed throughout the yeast genome. We also obtained the genomic coordinates of centromeres, telomeres, transfer RNAs, chromosomal breakpoints and origins of early DNA replication, all of which were studied in Duan *et al.* (3). In addition, we obtained 174 gene sets, each of which contains at least 20 yeast genes coregulated by a single transcription factor, that were recently tested for colocalization using the yeast interaction data (11).

The hypergeometric approach for assessing gene set colocalization

Duan *et al.* (3) and Dai and Dai (11) take the following approach for assessing the extent to which various genomic functional groups (e.g. centromeres, telomeres, genes coregulated by a single transcription factor) colocalize in the nucleus. For simplicity, we will refer to the elements of a genomic functional group as 'genes', though this need not be the case. Suppose that there are a total of N genes, of which n belong to the gene set of interest. Let M denote the number of all possible interchromosomal interactions between the N genes, and let K denote the actual number of experimentally observed interchromosomal interactions between the N genes. Let m denote the number of all possible interchromosomal interactions between the n genes in the gene set of interest, and let k denote the actual number of experimentally observed interchromosomal interactions between the n genes in the gene set of interest. Then, the authors claim that the probability of observing k interchromosomal interactions among the genes in the gene set is derived from a hypergeometric distribution; that is, the probability is

$$\frac{\binom{m}{k}\binom{M-m}{K-k}}{\binom{M}{K}}. \tag{1}$$

Hence, they conclude that the probability associated with observing at least k interchromosomal interactions among the genes in the gene set is

$$1-\sum_{x=0}^{k-1}\frac{\binom{m}{x}\binom{M-m}{K-x}}{\binom{M}{K}}. \tag{2}$$

Applying Equation 2 to a candidate set of n genes yields a P-value indicating whether the n genes colocalize in the nucleus.

Though both Duan *et al.* (3) and Dai and Dai (11) used a hypergeometric test to assess colocalization of genomic elements and gene sets, there is a slight discrepancy in the way that the two sets of authors defined the concept of an 'interchromosomal interaction'. In order to illustrate the difference we discuss the definition of K, the actual number of observed interchromosomal interactions. Duan *et al.* (3) computed K by summing, for each pair of genomic elements that lie on different chromosomes, the number of segments in the first genomic element that interacted with a segment in the second genomic element at a false discovery threshold below 0.01. On the other hand, Dai and Dai (11) computed K by counting the number of pairs of genes on different chromosomes for which at least one segment in the first gene interacted with at least one segment in the second gene at a false discovery threshold below 0.01. In reassessing the evidence for colocalization of the genomic elements and gene sets studied by the two sets of authors, we defined the concept of interchromosomal interactions as did each set of authors.

A resampling method for assessing gene set colocalization

Let $n_1,\ldots,n_I$ denote the number of genes in the gene set of interest that belong to each of the I chromosomes. Note that $\sum_{i=1}^{I} n_i = n$. We propose the following non-parametric resampling approach for assessing gene set colocalization:

(1) Compute k, the number of experimentally observed interchromosomal interactions among the genes in the gene set of interest.
(2) Compute m, the number of possible interchromosomal interactions among the genes in the gene set of interest.
(3) For $b = 1,\ldots, B$, where B is a large integer, such as 1000:
    (a) For the i-th chromosome, draw n_i genes uniformly at random, without replacement, from the genes on this chromosome. Repeat for each chromosome, so that $\sum_{i=1}^{I} n_i = n$ genes have been drawn. This is a 'random gene set'.
    (b) Compute k^{*b}, the number of experimentally observed interchromosomal interactions among the genes in the random gene set.

(c) Compute m^{*b}, the number of possible interchromosomal interactions among the genes in the random gene set.

(4) The P-value for the gene set of interest is given by

$$\frac{1}{B}\sum_{b=1}^{B} 1_{\left(\frac{k^{*b}}{m^{*b}} \geq \frac{k}{m}\right)}, \tag{3}$$

where $1_{(k^{*b}/m^{*b} \geq k/m)}$ is an indicator variable that equals 1 if $k^{*b}/m^{*b} \geq k/m$, and 0 otherwise.

Note that in Step 3(a), we ensure that the number of genes on each chromosome in our random gene set is the same as the number of genes on each chromosome in the gene set of interest. Essentially, our resampling approach computes a P-value by comparing the k/m ratio observed for the gene set of interest to the ratios obtained on arbitrary sets of genes. This is a P-value for the null hypothesis that the given gene set of interest shows no more colocalization than a randomly-chosen set of genes. We note that this approach for P-value calculation is not inherently novel, and indeed a similar approach was taken in a recent paper (4). The use of resampling approaches for hypothesis testing is discussed in more general terms in Efron and Tibshirani (12).

Generation of random gene sets in yeast interaction data

To assess the characteristics of the hypergeometric and resampling-based P-values on data generated under the null hypothesis of no gene set colocalization, we generated 1000 random gene sets. Each random gene set was obtained by selecting one of the 174 gene sets from the Dai and Dai paper and drawing that number of genes, without replacement, from the full set of genes. That is, each random gene set contained the same number of genes as one of the real gene sets known to be coregulated by a single transcription factor.

Generation of random interaction data

We repeated the following experiment 100 times, in order to generate 100 random interaction data sets. We generated the 3D positions of 1000 'genes' independently from the uniform distribution in the unit cube. We computed a 1000×1000 interaction matrix between these genes, where interactions were declared between each pair of genes whose Euclidean distance was among the smallest 10% of observed Euclidean distances. We then created 250 random gene sets, each of which was obtained by drawing 100 genes without replacement from the full set of 1000 genes.

RESULTS

Hypergeometric *P*-values are inappropriate for assessing gene set colocalization

If the hypergeometric P-values derived from Equation 2 are valid, then the P-value associated with an arbitrary selection of n genes should be drawn from a uniform distribution. We used the interaction data described in Duan *et al.* (3) to assess whether P-values obtained in this way are indeed uniform. We generated 1000 arbitrary gene sets (details in 'Materials and Methods' section) and computed hypergeometric P-values using Equation 2. A histogram of these P-values is displayed in Figure 1(a). These P-values are decidedly non-uniform—there are far too many extreme P-values.

To further investigate the properties of the hypergeometric P-values, we generated a simple simulated data set consisting of 1000 randomly generated observations in the 3D unit cube, which we used to generate an interaction matrix, as described in 'Materials and Methods' section. Gene sets were selected at random, and the histogram of hypergeometric P-values computed according to Equation 2 is given in Figure 1(b). Once again, the P-values are far from uniform.

What is wrong with using a hypergeometric P-value to assess colocalization of a given gene set? Such a P-value is based upon a 2×2 contingency table, shown in Table 1. The units that contribute to the contingency table are *gene pairs*; in Table 1, there are a total of $M = a+b+c+d$ gene pairs. A fundamental assumption that underlies the use of the hypergeometric distribution is that each gene pair in the contingency table is independent from all other gene pairs. That is, we can think of each gene pair as having two associated pieces of information: whether or not an interaction was observed for that gene pair (a binary variable, x_i for the i-th gene pair), and whether or not it is in the gene set of interest (a binary variable, z_i for the i-th gene pair). For the hypergeometric distribution to be valid, we need $(x_1, z_1), (x_2, z_2),\ldots,(x_M, z_M)$ to be independent and identically distributed (13).

But it is not hard to see that the assumption of independence is grossly violated in at least two ways. First, to see that $x_1,\ldots, x_M$ are not independent, note that if there is an interaction between the gene pair (i, j), and also between the gene pair (i, k), then the likelihood that there also is an interaction between the gene pair (j, k) is higher than the likelihood of interaction for a randomly selected pair of genes. This is because if the i-th gene is located near the j-th gene in 3D space, and the i-th gene is located near the k-th gene in 3D space, then the j-th and k-th genes must also be located near each other in 3D space. Second, to see that $z_1,\ldots,z_M$ are not independent, note that if the gene pair (i, j) is in the gene set, and also the gene pair (i, k) is in the gene set, then it must be the case that (j, k) is in the gene set. Given that the independence assumption underlying the hypergeometric distribution is violated in the context of assessing gene set colocalization, it should come as no surprise that the hypergeometric P-values are invalid. It is for this reason that the P-values observed in Figures 1a and 1b are non-uniform.

A valid, resampling approach for calculating *P*-values for gene set colocalization

In order to obtain a valid P-value for the extent of colocalization of a set of n genes, we can compare the number of experimentally observed interchromosomal

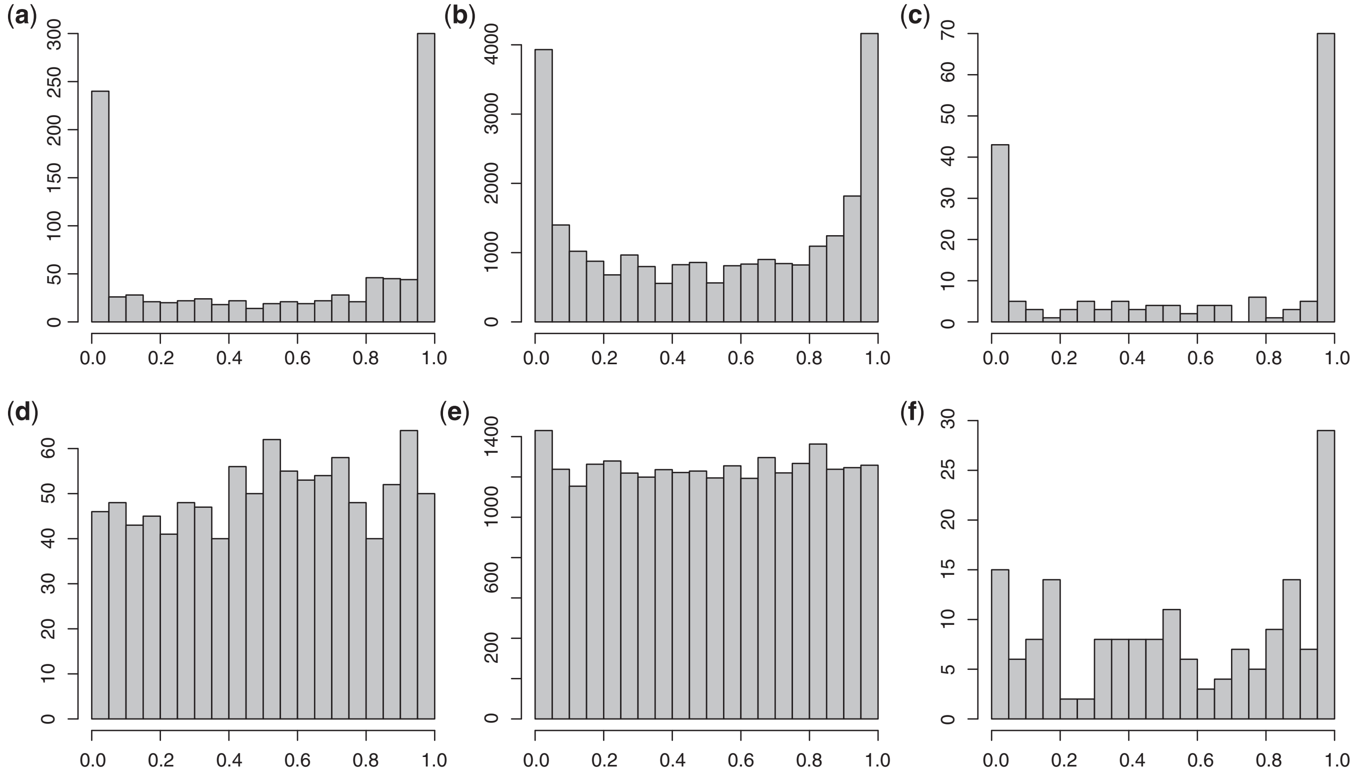

Figure 1. (**a**)–(**c**) show histograms of hypergeometric *P*-values. In panel (**a**), the *P*-values are computed for 1000 random gene sets with respect to the yeast interaction data set of Duan *et al.* (3). In panel (**b**) the *P*-values are computed with respect to a simulated data set for 250 random sets of 100 genes. In (c), the *P*-values correspond to 174 gene sets regulated by a single transcription factor and studied in (11), computed with respect to the yeast interaction data set of (3). Panels (**d**)–(**f**) are analogous to panels (a)–(c), but the *P*-values are computed using the resampling approach. In each case, the resampling-based *P*-values provide no evidence of colocalization of gene sets.

Table 1. The hypergeometric test is based upon a 2 × 2 contingency table of gene pairs

	Interaction	No interaction
In gene set	*a*	*b*
Not in gene set	*c*	*d*

Each element in the contingency table indicates the number of gene pairs corresponding to the associated row and column. For instance, there are *a* gene pairs in the gene set for which an interaction was observed, and *d* gene pairs not in the gene set for which no interaction was observed.

interactions among genes in this gene set to the distribution of the number of experimentally observed interchromosomal interactions that results from a set of *n* genes drawn at random from the full set of *N* genes. We cannot, unfortunately, compute the corresponding *P*-value analytically, as was the case for the hypergeometric *P*-value; however, it is straightforward to calculate the *P*-value using a resampling approach. Details of our proposed procedure are given in 'Materials and Methods' section.

To assess the validity of our approach, we first computed resampling-based *P*-values (Equation 3) on randomly selected gene sets of the yeast interaction dat of Duan *et al.* (3), as described in 'Materials and Methods' section. The resulting *P*-values are shown in Figure 1d. As expected, because the gene sets were chosen at random, the resulting *P*-values have a uniform distribution. We repeatedly generated such random gene sets and found that the quantiles of the *P*-values obtained very closely matched the quantiles of a uniform distribution.

We next computed resampling-based *P*-values for the simulated data set consisting of 1000 'genes' uniformly distributed in the unit cube (details given in Materials and Methods). The *P*-values that result almost perfectly match the quantiles of a uniform distribution, as expected (Figure 1e; the apparent excess of *P*-values in the left-most bin is due to the discreteness of the *P*-values).

Of course, the fact that the resampling-based *P*-values are uniformly distributed under the null hypothesis does not provide sufficient evidence of their adequacy: it must also be shown that they are small in the presence of colocalization, i.e. under the alternative. To assess this, we generated gene sets under the alternative by choosing a gene at random, and then selecting the 100 genes nearest to it in terms of Euclidean distance. Each of the resulting gene sets had an extremely small *P*-value, indicating that the resampling-based *P*-values have power to reject the null hypothesis when there is indeed evidence of colocalization.

Reanalysis of the colocalization of transcriptionally regulated yeast genes

Dai and Dai (11) recently examined the yeast interaction data of Duan *et al.* (3) in order to determine the extent to which sets of genes coregulated by a single transcription factor tend to colocalize. They identified 174 transcription factors, each of which regulated at least 20 genes. They applied the hypergeometric *P*-values (Equation 2) in order to assess the colocalization of each set of coregulated genes, and found that 34 sets of coregulated genes had *P*-values below 0.01. We repeated the analysis of Dai and Dai, and confirmed their finding that a substantial number of the coregulated gene sets had very small hypergeometric *P*-values (Figure 1c). However, on the basis of resampling-based *P*-values, there is essentially no evidence of colocalization of sets of coregulated genes (Figure 1f). None of the 174 gene sets had a resampling-based *P*-value below 0.01 after Bonferroni correction. The hypergeometric and resampling-based *P*-values displayed in Figures 1(c) and 1(f), as well as the transcription factors regulating each of the 174 gene sets, can be found in Supplementary Table S1.

Not surprisingly, because both the hypergeometric and resampling-based *P*-values are based upon the number of experimentally observed interchromosomal interactions in a given gene set of interest, these two types of *P*-values are highly correlated with each other: their Spearman correlation is 0.969. Therefore, the choice of *P*-value does not affect the relative ranking of evidence for gene set colocalization as much as it does the absolute amount of evidence for gene set colocalization.

Reanalysis of the colocalization of functional elements in the yeast genome

Duan *et al.* (3) assessed the extent to which certain genomic functional groups—centromeres, telomeres, transfer RNAs, chromosomal breakpoints and origins of early DNA replication—tend to colocalize in the nucleus. Of 14 such functional groups, they found that 10 colocalize in the nucleus, as evidenced by a hypergeometric $P< 0.01$ after Bonferroni correction. (In a related analysis, no evidence was found for colocalization of genes that share various Gene Ontology terms.) We computed the resampling-based *P*-values for each of these 14 functional groups, as described in the 'Materials and Methods' section, with the following approach for drawing random functional elements: for each of the n functional elements of interest (e.g. $n = 32$ telomeres) we repeatedly drew a 'random' functional element from the corresponding chromosome, of the same length as the true functional element of interest, uniformly at random along the length of the chromosome. (In contrast, in our reanalysis of the Dai and Dai data (11) we generated 'random' genes by drawing genes, without replacement, from the full set of genes on a given chromosome. Unfortunately, it is not possible to take that approach in our reanalysis of the Duan *et al.* (3) paper, due to the nature of the functional elements considered. The approach that we took instead is a natural alternative.)

In the Duan *et al.* (3) study, 10 groups of functional elements showed significant evidence of colocalization according to the Bonferroni adjusted hypergeometric test. In our reanalysis (Figure 2), three of these groups are no longer significant after Bonferonni adjustment: the complete set of telomeres, one of the two sets of early-firing origins, and one of the two sets of chromosomal breakpoints. Thus, our results suggest that (i) Duan *et al.* (3) incorrectly concluded that telomeres exhibit colocalization, and (ii) the evidence for colocalization of early-firing origins and for chromosomal breakpoints is weaker than initially reported.

DISCUSSION

In two recent papers, Duan *et al.* (3) and Dai and Dai (11) assessed the extent to which certain functional genomic elements colocalize, using *P*-values derived from a hypergeometric distribution. We have shown here that such hypergeometric *P*-values are flawed. The assumptions of the hypergeometric distribution are inappropriate in this setting, and consequently hypergeometric *P*-values computed on random gene sets are far from uniform. We then presented an alternative, resampling-based *P*-value calculation approach that is suitable for this setting. These resampling-based *P*-values indicate a complete lack of evidence that the 174 coregulated gene sets studied in Dai and Dai (11) colocalize in the nucleus. However, they do support the hypothesis that centromeres colocalize, and provide some evidence in support of colocalization of other functional genomic elements.

In the current study, we reassessed the extent to which target genes of 174 TFs, considered by Dai and Dai (11), exhibit colocalization. We did not investigate several other results in that study that were also based on a hypergeometric test: that only one TF shows significant colocalization based on intrachromosomal interactions, that 5 of 158 TFs measured via ChIP-chip show evidence of colocalization of their targets, and that various classes of chromatin regulatory genes—histone modification regulated genes, genes whose promoters exhibit high chromatin remodeler occupancy, genes that show expression changes in response to chromatin remodeler perturbation, genes whose promoters are occupied by nucleosomes, genes containing histone variant H2A.Z, and genes with high *trans* effects on gene expression divergence—are colocalized. We are not claiming that coregulated gene sets do not colocalize in the nucleus; we are simply stating that there does not appear to be evidence in the Duan *et al.* (3) data set of colocalization of the 174 gene sets studied by Dai and Dai (11).

The implications of our reanalysis for the claims made in the Duan *et al.* paper are relatively minor. The primary colocalization claims in that paper—regarding centromeres, telomeres, tRNAs, breakpoints and origins of replication—were based primarily upon the qualitative assessment of a set of receiver operating characteristic curves (Figure 4d of that paper). This analysis was augmented by a set of hypergeometric tests, reported in their

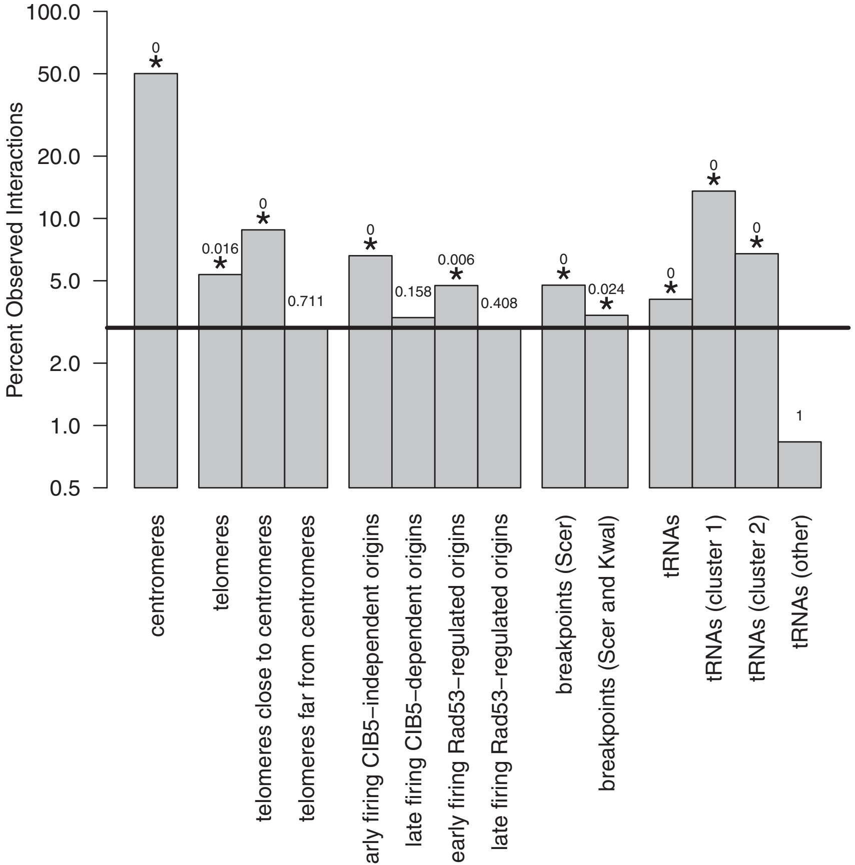

Figure 2. Based on the yeast interaction data of Duan *et al.*, hypergeometric and resampling-based *P*-values were computed to assess the extent to which certain functional groups colocalize. The height of each bar indicates enrichment or depletion of observed interchromosomal interactions relative to the percent (black line) of all possible interactions that were observed at a false discovery rate below 0.01. Above each bar, the resampling-based *P*-value is reported (without correction for multiple testing), and an asterisk indicates that the hypergeometric *P*-value was below 0.01 after Bonferroni correction. Additional information about the fourteen sets of functional elements can be found in Duan *et al.* (3).

Supplementary Figure 11. Our analysis suggests that three of the asterisks in that figure (indicating Bonferroni adjusted significance of 0.01) were erroneous. These three changes imply weaker statistical support for the colocalization of early-firing origins of replication and chromosomal breakpoints, and no support for the colocalization of telomeres.

We have shown that using a hypergeometric test to assess colocalization of a gene set is invalid, since the gene pairs underlying the hypergeometric test calculation are not independent. Goeman and Buhlmann (13) showed that for a similar reason, it is incorrect to use a hypergeometric test to assess the extent to which genes associated with a particular Gene Ontology term are differentially expressed. The problem of assessing colocalization of gene sets is inherently a difficult one, since it is unclear what one would expect 3D interaction data to look like under the null hypothesis, i.e. in the absence of colocalization. To overcome this difficulty, we have proposed a resampling-based approach for assessing colocalization. This approach suffers from some drawbacks that are shared with the hypergeometric test. Using the terminology of Goeman and Buhlmann (13), both the resampling-based and hypergeometric *P*-values test a *competitive null hypothesis*, which posits that genes in a given gene set colocalize no more than the genes not in the gene set. Both are *gene sampling methods*, and hence do not provide any information about whether a given gene set will colocalize on a new interaction matrix derived from a future experiment. (Indeed, on the basis of a single interaction matrix, one cannot make claims about future experiments.) Instead, these *P*-values tell us whether or not, if one were to obtain more genes corresponding to a given gene set, one would expect those new genes to colocalize. Though both the hypergeometric and resampling-based *P*-values are gene sampling methods for testing the competitive null, the resampling-based *P*-values do not rely on the untenable assumption of independence of gene pairs. Consequently, unlike the hypergeometric *P*-values, our proposed *P*-values follow a uniform distribution under the null hypothesis of no colocalization.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table S1.

ACKNOWLEDGEMENTS

We thank Zhiming Dai and Mirela Andronescu for helpful responses to our inquiries and for making data available. Ali Shojaie and three reviewers provided suggestions that improved the quality of this manuscript.

FUNDING

National Institutes of Health (DP5OD009145 to D.W., R01 GM098039 to W.S.N.). Funding for open access charge: NIH award P41 RR0011823.

Conflict of interest statement. None declared.

REFERENCES

1. Dekker,J., Rippe,K., Dekker,M. and Kleckner,N. (2002) Capturing chromosome conformation. *Science*, **295**, 1306–1311.
2. Lieberman-Aiden,E., van Berkum,N.L., Williams,L., Imakaev,M., Ragoczy,T., Telling,A., Amit,I., Lajoie,B.R., Sabo,P.J., Dorschner,M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–293.
3. Duan,Z., Andronescu,M., Schutz,K., McIlwain,S., Kim,Y., Lee,C., Shendure,J., Fields,S., Blau,C. and Noble,W. (2010) A three-dimensional model of the yeast genome. *Nature*, **465**, 363–367.
4. Tanizawa,H., Iwasaki,O., Tanaka,A., Capizzi,J.R., Wickramasignhe,P., Lee,M., Fu,Z. and Noma,K. (2010) Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.*, **38**, 8164–8177.
5. Bolzer,A., Kreth,G., Solovei,I., Koehler,D., Saracoglu,K., Fauth,C., Müller,S., Eils,R., Cremer,C., Speicher,M.R. *et al.* (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.*, **3**, e157.
6. Guelen,L., Pagie,L., Brasset,E., Meuleman,W., Faza,M.B., Talhout,W., Eussen,B.H., de Klein,A., Wessels,L., de Laat,W. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*, **453**, 948–951.
7. Pan,Y.F., Lin,J., Xu,H., Mohamed,Y.B., Orlov,Y.L., Ho,A., Mei,P.H., Chew,E.G. and Huang,P.Y. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, **462**, 58–64.
8. Lanctôt,C., Cheutin,T., Cremer,M., Cavalli,G. and Cremer,T. (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Gene.*, **8**, 104–115.
9. Foster,H.A. and Bridger,J.M. (2005) The genome and the nucleus: a marriage made by evolution. *Chromosoma*, **114**, 212–229.
10. Tanabe,H., Müller,S., Neusser,M., von Hase,J., Calcagno,E., Cremer,M., Solovei,I., Cremer,C. and Cremer,T. (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Nat Acad. Sci. USA*, **99**, 4424–4429.
11. Dai,Z. and Dai,X. (2012) Nuclear colocalization of transcription factor target genes strengthens coregulation in yeast. *Nucleic Acids Res.*, **40**, 27–36.
12. Efron,B. and Tibshirani,R. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, London.
13. Goeman,J. and Buhlmann,P. (2007) Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics*, **23**, 980–987.