



# Gene Repositioning Is Under Constraints After Evolutionary Conserved Gene Neighborhood Separate

Zhiming Dai<sup>1,2\*</sup>

<sup>1</sup> School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China, <sup>2</sup> Guangdong Province Key Laboratory of Big Data Analysis and Processing, Sun Yat-Sen University, Guangzhou, China

## OPEN ACCESS

### Edited by:

Matteo Brilli,  
University of Milan,  
Italy

### Reviewed by:

Michael Eric Schranz,  
Wageningen University & Research,  
Netherlands  
Ivan Junier,  
UMR5525 Techniques de l'Ingénierie  
Médicale et de la Complexité  
Informatique, Mathématiques et  
Applications, Grenoble (TIMC-IMAG),  
France

### \*Correspondence:

Zhiming Dai  
daizhim@mail.sysu.edu.cn

### Specialty section:

This article was submitted to  
Bioinformatics and  
Computational Biology,  
a section of the journal  
Frontiers in Genetics

**Received:** 12 June 2019

**Accepted:** 25 September 2019

**Published:** 03 October 2019

### Citation:

Dai Z (2019) Gene Repositioning  
Is Under Constraints After  
Evolutionary Conserved Gene  
Neighborhood Separate.  
*Front. Genet.* 10:1030.  
doi: 10.3389/fgene.2019.01030

Genes are not randomly distributed on eukaryotic chromosomes. Some neighboring genes show order conservation among species, while some neighboring genes separate during evolution even though their neighborhoods are conserved in some species. Here, I investigated whether after-separation gene repositioning is under natural selection for evolutionary conserved gene neighborhoods compared with nonconserved neighborhoods. After separation, genes with conserved neighborhoods show low-expression divergence between the after-separation species and the before-separation species. After genes separate from their conserved gene neighbors, their after-separation gene neighbors tend to show coexpression and coprotein complex with their before-separation gene neighbors. These results indicate evolutionary constraints on the selection of neighboring genes after evolutionary conserved gene neighborhoods separate.

**Keywords:** genomic neighbors, chromatin regulation, gene expression, yeast, bioinformatics

## INTRODUCTION

Gene order in genomes is not random (Hurst et al., 2004). Genes that tend to remain proximal along DNA (synteny property) show strong transcriptional coexpression in bacterial genomes (Junier and Rivoire, 2016). Neighboring gene pairs tend to show higher coexpression than random gene pairs in yeast *Saccharomyces cerevisiae* (Kruglyak and Tang, 2000; Fukuoka et al., 2004; Lercher and Hurst, 2006), nematode *Caenorhabditis elegans* (Lercher et al., 2003), fruit fly *Drosophila melanogaster* (Boutanaev et al., 2002; Spellman and Rubin, 2002; Bailey et al., 2004; Kalmykova et al., 2005), and mammal human (Singer et al., 2005; Li et al., 2006; Semon and Duret, 2006; Ghanbarian and Hurst, 2015). This coexpression spans up to tens of neighboring genes in *S. cerevisiae* and *D. melanogaster* (Lercher and Hurst, 2006; Weber and Hurst, 2011). Essential genes tend to be genomic neighborhoods in *C. elegans*, *S. cerevisiae* and mouse (Kamath et al., 2003; Pal and Hurst, 2003; Hentges et al., 2007). Genes in specific functional gene groups tend to be neighbors within chromosomes of the human genome (Teichmann and Veitia, 2004; Thevenin et al., 2014). Genes encoding subunits of the same complex tend to aggregate on chromosomes in *S. cerevisiae* (Teichmann and Veitia, 2004). Genes from the same metabolic pathway also show clustering in eukaryotic genomes (Lee and Sonnhammer, 2003).

A range of mechanisms have been proposed to be responsible for the nonrandom gene order in genomes. While proximal genes tend to be coexpressed in bacteria by facilitated coexpression mechanisms independently of the action of transcription factor (TF) (Junier and Rivoire, 2016), TF

regulation constrains gene order on yeast chromosomes (Janga et al., 2008), as neighboring genes tend to be coregulated by the same TF (Hershberg et al., 2005; Michalak, 2008). Adjacent genes also show similarity in nucleosome occupancy, chromatin remodeling program, and histone modification pattern (Batada et al., 2007; Chen et al., 2010; Deng et al., 2010). These chromatin structures might influence gene expression and, hence, the conservation of gene order.

Coexpression of neighboring genes still persists after their separation (Wang et al., 2011). Moreover, neighboring genes show interchromosomal colocalization after their separation, and these gene pairs tend to be bound by the same TF (Veron et al., 2011; Dai et al., 2014). These results suggest that gene pairs that were genomic neighbors in the evolutionary past, but currently separated, might be still under similar regulatory programs.

A considerable number of genes show gene neighborhood conservation over long evolutionary periods (Hurst et al., 2002; Arnone et al., 2012), suggesting that these gene neighborhoods are selectively favorable. If the conservation has functional consequence, conserved gene neighborhoods should maintain links after their separation. Regarding three-dimensional nuclear links, we have previously found that genes showing neighborhood conservation in more species show more interchromosomal colocalization after their separation (Dai et al., 2014). Regarding one-dimensional links, after evolutionary-conserved neighboring genes separate, I asked whether their new positions are under selective constraints (Figure 1A). In this study, I focused on yeast species including *S. cerevisiae* to test this possibility since *S. cerevisiae* has been the main model in the studies of gene order, as tremendous amounts of molecular knowledge and data are available.

## MATERIALS AND METHODS

### Data Preparation

The genome sequences, gene and chromosome annotations of the yeast species, and high-quality predicted gene orthology map across yeast species were obtained from a previous compilation (Wapinski et al., 2007). Gene expression data of yeast species were taken from Thompson et al. (2013). In this study, I used 11 yeast species whose gene annotation data and gene expression data are both available (Figure 1B). These include *S. cerevisiae* and its close relatives (*sensu stricto* clade), two other species who have diverged after the whole-genome duplication (*Saccharomyces castellii*, and the human pathogen *Candida glabrata*), two members of the *Kluyveromyces* clade (*Kluyveromyces waltii* and *Kluyveromyces lactis*), two members of the *Candida* clade (the human pathogen *C. albicans* and the halophile *Debaryomyces hansenii*), and *Yarrowia lipolytica*.

Gene expression divergence data in yeast species (including *S. cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, and *Saccharomyces kudriavzevii*) were taken from Tirosh et al. (2006). The measure of gene expression divergence was based on the interspecies correlations in the gene expression patterns. For each gene, the expression divergence between each two species was corrected by subtracting its value with the average

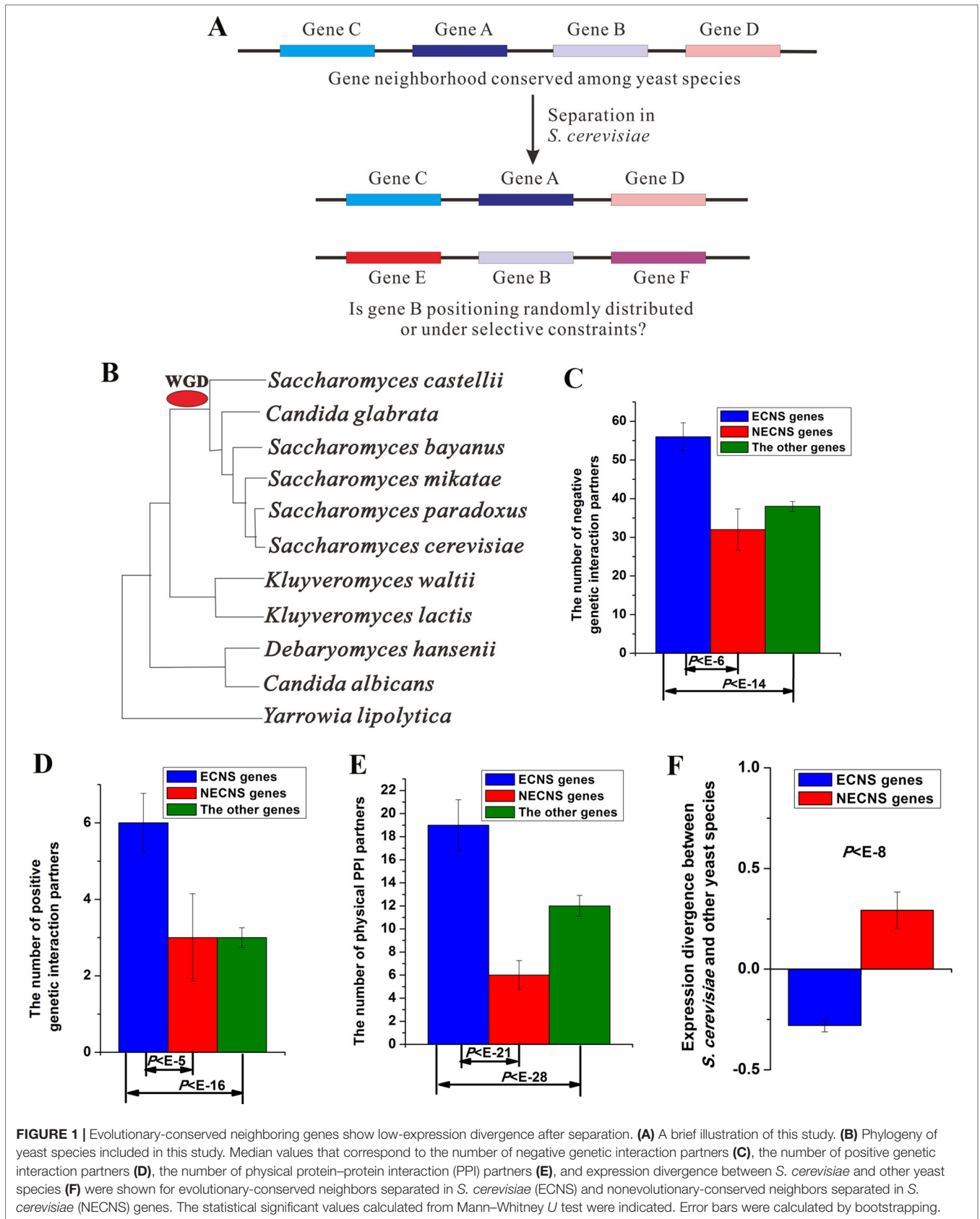
intraspecies divergence of the two corresponding species. The corrected expression divergence for each pairwise species comparison was normalized by subtracting its mean and dividing by its standard deviation, and the normalized values were then averaged over the all pairwise species comparisons; thus, a single value of expression divergence was given for each gene (Tirosh et al., 2006). Genome-wide binding data corresponding to 203 TFs were taken from Harbison et al. (2004). A P value cutoff of 0.005 was used to define the set of genes bound by one particular TF. Protein-protein interaction (PPI) data were taken from Stark et al. (2006). Genetic interaction data were taken from Costanzo et al. (2016). The member genes of cellular component gene sets from *S. cerevisiae* were downloaded from the Gene Ontology repository (Harris et al., 2004). P values for Gene Ontology process terms were derived using "GOTermFinder" (<https://go.princeton.edu/cgi-bin/GOTermFinder>). The list of essential genes was taken from Giaever et al. (2002).

### Identification of Two Gene Groups

For each gene in *S. cerevisiae*, I identified its orthologous counterparts in the other 10 yeast species according to a yeast orthology map (Wapinski et al., 2007). I focused on *S. cerevisiae* genes having orthologous counterparts in at least 60% (i.e., 6–10 species) of the other 10 species, resulting in 4,093 genes. I identified *S. cerevisiae* gene pairs whose orthologous counterparts show neighborhoods in more than half ( $\geq 50\%$ ) of their appeared yeast species, meaning that for gene pairs respectively having orthologous counterparts in 6, 7, 8, 9, and 10 species, their orthologous counterparts are gene neighbors in at least 3, 4, 4, 5, and 5 species. The resulting gene pairs were referred to as evolutionary-conserved neighboring gene pairs. I searched for *S. cerevisiae* gene pairs that met two criteria: 1) They are evolutionary-conserved neighboring gene pairs, and 2) they are either on different chromosomes or flanked by at least 15 genes if on the same chromosome in *S. cerevisiae*, that is, they are separated in *S. cerevisiae*. These criteria resulted in 902 gene pairs, which included 1,367 genes. These gene pairs were referred to as yeast evolutionary-conserved neighboring but *S. cerevisiae* separated gene pairs (Table S1). Genes in most (853) of the 902 gene pairs are on different chromosomes in *S. cerevisiae*. Orthologous counterparts of the resulting gene pairs tend to remain neighbors in seven yeast species (Figure S1). The numbers in the seven species were similar.

Note that the resulting genes might show evolutionary-neighboring conservation with either *S. cerevisiae* neighboring genes, that is, they might not separate from their other evolutionary-conserved neighbors in *S. cerevisiae*, which might cause confusion. Among the 1,367 genes involved in the identified yeast evolutionary-conserved neighboring but *S. cerevisiae* separated gene pairs, I excluded genes in evolutionary-conserved neighboring gene pairs identified above, resulting in 990 genes, and referred to them as genes with evolutionary-conserved neighbors separated in *S. cerevisiae* (ECNS) (Table S2).

To control for the evolutionary-conserved neighborhood, I searched for *S. cerevisiae* gene pairs that met two criteria: (1) Their orthologous counterpart are gene neighbors in only 1 of the 10 yeast species, and (2) they are either on different chromosomes or flanked by at least 15 genes if on the same chromosome in *S. cerevisiae*, that is, they are separated in *S. cerevisiae*. I



referred to the resulting gene pairs as yeast nonevolutionary-conserved neighboring but *S. cerevisiae* separated gene pairs. To avoid confusion, I excluded genes in evolutionary-conserved neighboring gene pairs from the resulting genes. The refining genes were referred to as genes with nonevolutionary-conserved neighbors separated in *S. cerevisiae* (NECNS).

## Randomized Experiments

I used randomized experiments to test the statistical significance of frequencies of cobound TF, pairwise genetic interactions, pairwise PPI, and cocellular component for gene pairs between before- and after-separation neighboring genes of ECNS genes. For each gene in all gene pairs, I randomly permuted it with one gene that met two criteria: 1) It was on the same chromosomes as the actual gene because some properties like TF binding show preference for some specific chromosomes, and 2) it was not the same gene as the actual gene. In this way, I generated random genes of which the number was the same as that of genes in actual gene pairs. I replaced the actual genes with their corresponding random genes. In this way, I generated random gene pairs of which the number was the same as that of actual gene pairs. I calculated frequencies of cobound TF, pairwise genetic interactions, pairwise PPI, and cocellular component for these random pairs. If the cobound TF, genetic interactions, PPI, and cocellular component were not features of actual gene pairs, the random gene pairs should show a similar degree of frequencies as actual gene pairs. I repeated the randomized experiment 100,000 times to calculate the frequency of experiments having higher degrees than actual gene pairs and referred to this value as *P* value.

## RESULTS

### Evolutionary-Conserved Neighboring Genes Show Low-Expression Divergence After Separation

I studied 11 yeast species whose gene annotation data and gene expression data are both available. I identified 902 *S. cerevisiae* gene pairs, orthologous counterparts of which were gene neighbors in most of their appeared yeast species, but separated in *S. cerevisiae* (see details in Materials and methods section, **Table S1**). However, some genes in my identified gene pairs show evolutionary-neighboring conservation with their *S. cerevisiae* neighboring genes. I excluded these genes to avoid confusion, resulting in 990 genes, and referred to them as ECNS genes (see details in Materials and methods section, **Table S2**). I also identified *S. cerevisiae* gene pairs, orthologous counterparts of which were gene neighbors in only one yeast species, but separated in *S. cerevisiae*. I referred to the resulting genes as NECNS genes. Both ECNS and NECNS genes separated from neighbors of their orthologous counterparts in *S. cerevisiae*; these two gene groups differ in whether their orthologous counterparts show evolutionary conservation in neighborhoods.

I examined the enrichment of ECNS genes for genetic interaction in *S. cerevisiae*. ECNS genes have more genetic interaction partners than NECNS and the other genes (negative interaction:  $P < 10^{-6}$  for NECNS genes,  $P < 10^{-14}$  for the other genes; positive interaction:

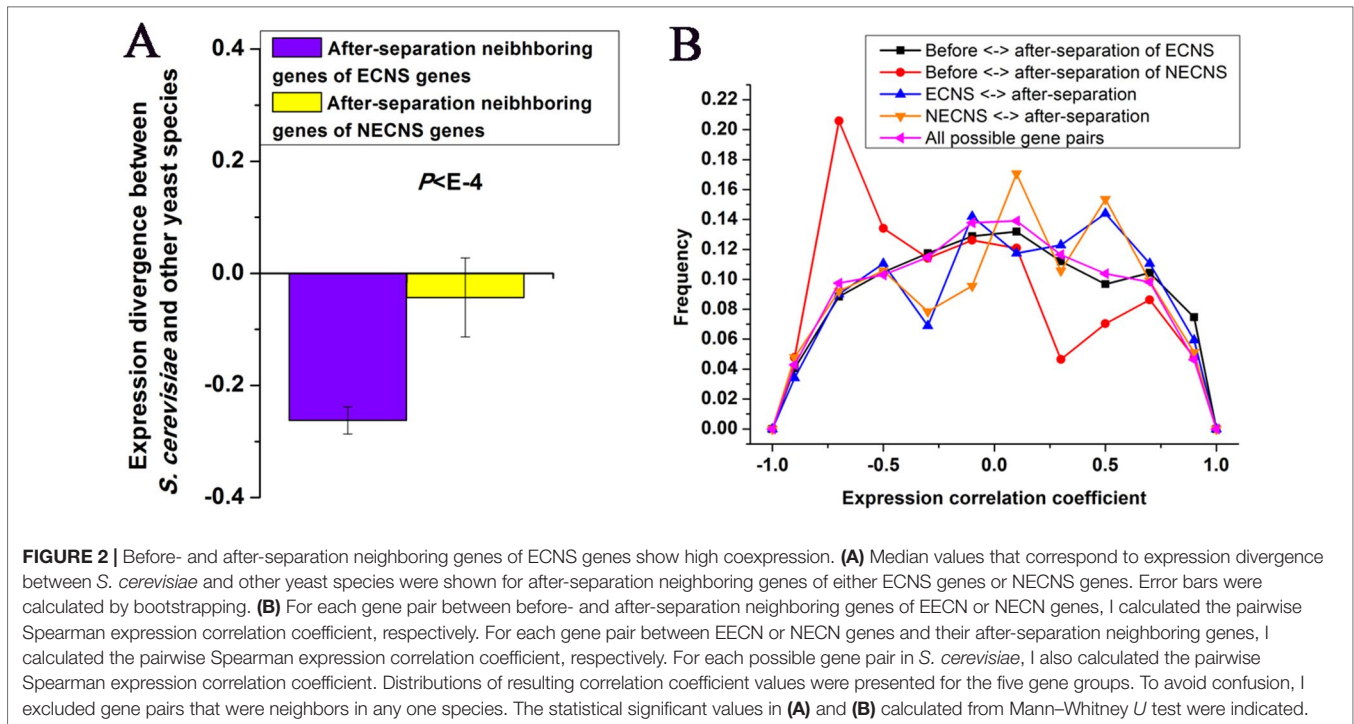
$P < 10^{-5}$  for NECNS genes,  $P < 10^{-16}$  for the other genes; Mann–Whitney *U* test, **Figure 1C, D**). Moreover, ECNS genes have more PPI partners than NECNS and the other genes ( $P < 10^{-21}$  for NECNS genes,  $P < 10^{-28}$  for the other genes, Mann–Whitney *U* test, **Figure 1E**). ECNS genes tend to be essential genes compared with NECNS genes (~22 vs. ~7%) (Giaever et al., 2002). However, ~16% of all *S. cerevisiae* genes are essential genes, indicating that NECNS are relatively depleted in essential genes. ECNS genes tend to be involved in housekeeping processes, including cellular process, metabolic process, and biogenesis as indicated by Gene Ontology, whereas NECNS genes tend to be involved in transport process (**Table S3**).

I examined whether ECNS and NECNS genes show similar expression divergence between *S. cerevisiae* and other yeast species (Tirosh et al., 2006). Interestingly, I found that ECNS genes show lower expression divergence than NECNS genes ( $P < 10^{-8}$ , Mann–Whitney *U* test, **Figure 1F**). Moreover, before-separation neighboring genes of ECNS genes also show lower expression divergence than those of NECNS genes ( $P < 10^{-13}$ , Mann–Whitney *U* test, **Figure S2**). These results are consistent with a previous human finding that grouping human genes together along the genome is associated with reduced expression noise (Kustatscher et al., 2017).

### After-Separation Neighboring Genes of ECNS Genes Show Low-Expression Divergence

I examined whether after-separation neighboring genes of ECNS genes in *S. cerevisiae* are under selective constraints to maintain low-expression divergence of ECNS genes. I found that after-separation neighboring genes of ECNS genes show lower expression divergence than those of NECNS genes ( $P < 10^{-4}$ , Mann–Whitney *U* test, **Figure 2A**). I asked whether ECNS genes show high coexpression with their after-separation neighboring genes to maintain their low-expression divergence. ECNS genes and their after-separation neighboring genes show higher coexpression than all possible gene pairs in *S. cerevisiae* (Thompson et al., 2013) (average coexpression level: 0.07 vs. 0.004,  $P < 10^{-6}$ , Mann–Whitney *U* test, **Figure 2B**) but show similar coexpression levels compared with those between NECNS genes and their after-separation neighboring genes (average coexpression level: 0.07 vs. 0.04,  $P = 0.59$ , Mann–Whitney *U* test).

Considering the low-expression divergence of ECNS genes, if their evolutionary-conserved neighboring genes strongly influence their gene expression, after-separation gene neighbors that show high coexpression with their before-separation neighbors should be selectively favorable. Indeed, before- and after-separation neighboring genes of ECNS genes show higher coexpression than all possible gene pairs (average coexpression level: 0.04 vs. 0.004,  $P < 10^{-11}$ , Mann–Whitney *U* test, **Figure 2B**). Before- and after-separation neighboring genes of ECNS genes also show higher coexpression than before- and after-separation neighboring genes of NECNS genes (average coexpression level: 0.04 vs. -0.14,  $P < 10^{-20}$ , Mann–Whitney *U* test, **Figure 2B**). However, before- and after-separation neighboring genes of ECNS genes show lower coexpression than ECNS genes and their after-separation neighboring genes ( $P = 0.01$ , Mann–Whitney *U* test, **Figure 2B**). These results together implied that there is selective



pressure for genes to insert in genomic neighborhoods that have similar expression noise (i.e., low-expression divergence), independent of whether genes are coexpressed. On the other hand, ECNS genes and their before-separation neighboring genes show comparable coexpression levels in *S. cerevisiae* with all possible gene pairs (average coexpression level: 0.01 vs. 0.004,  $P = 0.07$ , Mann–Whitney *U* test).

### Before- and After-Separation Neighboring Genes of ECNS Genes Tend to be in the Same Cellular Component

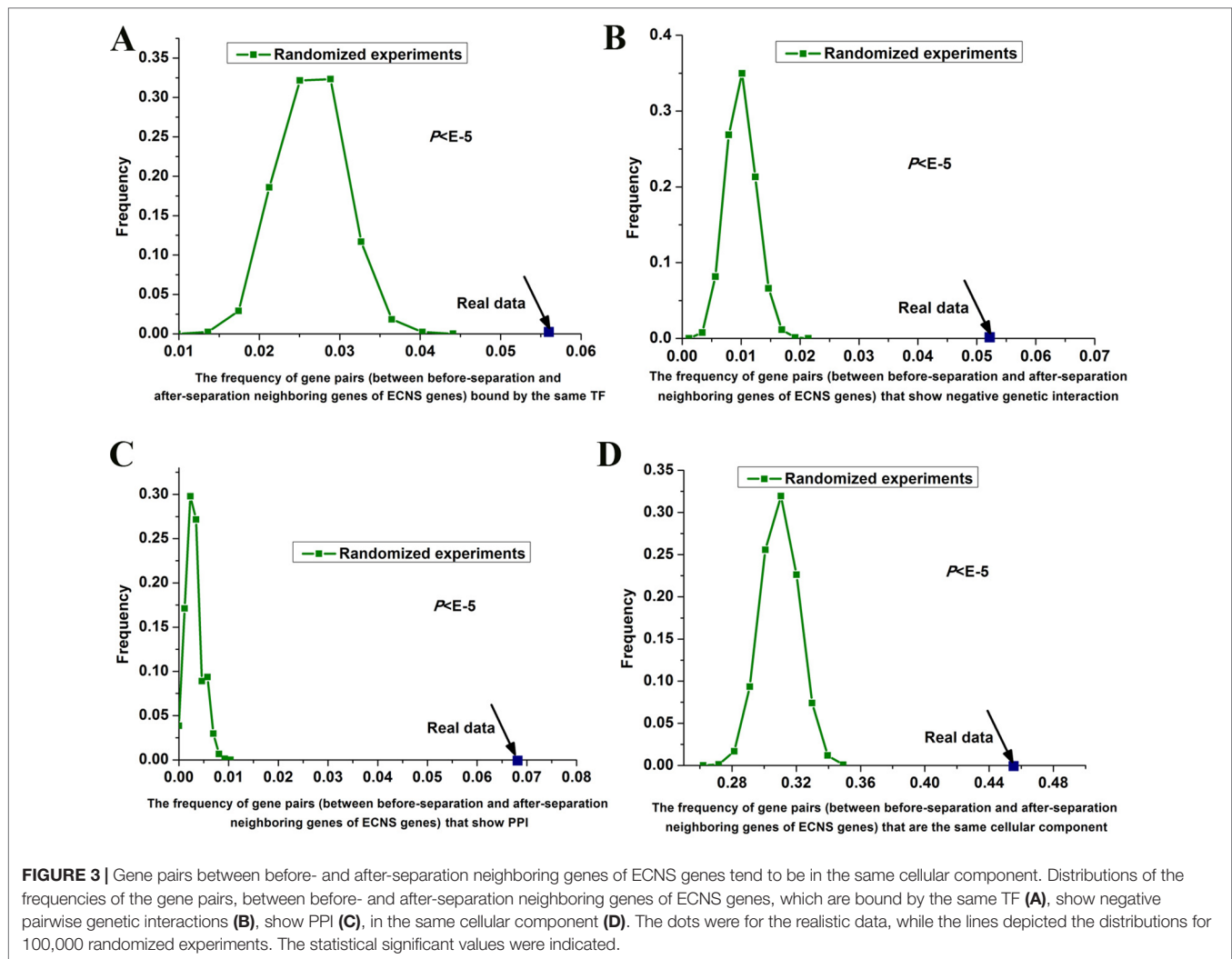
I tested whether before- and after-separation neighboring genes of ECNS genes are regulated by similar transcriptional programs. To address this, I used genome-wide 203 TF binding data in *S. cerevisiae* (Harbison et al., 2004). I generated random gene pairs to test the statistical significance (see details in Materials and methods section). I repeated the randomized experiment 100,000 times to calculate *P* value. I found that before- and after-separation neighboring genes of ECNS genes tend to be bound by the same TF: the frequencies of gene pairs bound by the same TF for all these randomized experiments were lower than that of the actual gene pairs ( $P < 10^{-5}$ , **Figure 3A**).

Genetic interaction network plays an important role in the evolution of gene order (Yang et al., 2017). Using genome-wide genetic interaction network data in *S. cerevisiae* (Costanzo et al., 2016) and 100,000 times of randomized experiments, I found that gene pairs between before- and after-separation neighboring genes of ECNS genes tend to have pairwise genetic interactions ( $P < 10^{-5}$  for negative genetic interaction, **Figure 3B**,  $P = 0.02$  for positive genetic interaction). Gene pairs between before- and after-separation neighboring genes of NECNS genes do not

show these properties ( $P = 0.37$  for negative genetic interaction,  $P = 0.64$  for positive genetic interaction). As genes showing negative genetic interactions tend to encode proteins in the same protein complex (Costanzo et al., 2016), I tested whether it holds true for gene pairs between before- and after-separation neighboring genes of ECNS genes. Using genome-wide physical PPI data in *S. cerevisiae* (Stark et al., 2006) and 100,000 times of randomized experiments, I found that encoding proteins of gene pairs between before- and after-separation neighboring genes of ECNS genes tend to show pairwise PPI ( $P < 10^{-5}$ , **Figure 3C**). Using Gene Ontology data in *S. cerevisiae* (Harris et al., 2004) and 100,000 times of randomized experiments, I found that gene pairs between before- and after-separation neighboring genes of ECNS genes tend to belong to the same cellular component ( $P < 10^{-5}$ , **Figure 3D**).

## DISCUSSION

Neighboring genes are frequently coexpressed. Whether this coexpression has consequent function or is just transcriptional noise remains debated (Xu et al., 2012; Kustatscher et al., 2017). If it is just the by-product of transcriptional noise, after genes separated from their evolutionary conserved neighboring genes, their after-separation gene neighbors should not be under selective constraints to show coexpression with their before-separation gene neighbors. Our results show that before- and after-separation neighboring genes of ECNS genes show statistical significantly higher coexpression than before- and after-separation neighboring genes of NECNS genes, although the mean coexpression level between before- and



after-separation neighboring genes of ECNS genes is relatively low (0.04).

Using randomized experiments, I found that a statistically significant number of gene pairs between before- and after-separation neighboring genes of ECNS genes are bound by the same TF, to be in the same cellular component, show pairwise genetic interactions and PPI. These cofeatures are generally associated with gene coregulation. These results indicate that genes showing some similar specific properties with before-separation neighboring genes of ECNS genes tend to be selected as after-separation neighboring genes of ECNS genes. On the other hand, the numbers of gene pairs showing cofeatures differ among these properties. Significantly, ~46% of gene pairs between before- and after-separation neighboring genes of ECNS genes are in the same cellular component (Figure 3D). It is likely that genes showing cofeature with before-separation neighboring genes of ECNS genes are most favorable as after-separation neighboring genes of ECNS genes.

One interesting finding in this study is that ECNS genes and their after-separation neighbor genes show comparable

coexpression levels with those between NECNS genes and their after-separation neighbor genes. I also identified nonseparation gene neighbors conserved from at least four yeast species to *S. cerevisiae*. These nonseparation gene neighbors show similar coexpression levels with those between ECNS genes and their after-separation neighbor genes (data not shown). The coexpression of neighboring genes is a phenomenon caused by many factors. The contribution of these factors varies with genes. Insights into mechanisms of regulation and gene order evolution from only neighboring coexpression information might be masked by some factors. Integrating neighboring coexpression information with gene neighborhood conservation information and before-separation neighboring information could provide deep insights into mechanisms of gene order evolution.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.broad.mit.edu/regev/orthogroups/>.

## AUTHOR CONTRIBUTIONS

ZD designed the study, implemented the algorithms, carried out the experiments, analyzed the results, and drafted the manuscript.

## FUNDING

This work was supported by National Natural Science Foundation of China (NSFC) (grant nos. 61872395 and U1611265),

## REFERENCES

- Arnone, J. T., Robbins-Pianka, A., Arace, J. R., Kass-Gergi, S., and Mclear, M. D. (2012). The adjacent positioning of co-regulated gene pairs is widely conserved across eukaryotes. *BMC Genomics* 13, 546. doi: 10.1186/1471-2164-13-546
- Bailey, J. A., Baertsch, R., Kent, W. J., Haussler, D., and Eichler, E. D. (2004). Hotspots of mammalian chromosomal evolution. *Genome Biol.* 5, R23. doi: 10.1186/gb-2004-5-4-r23
- Batada, N. N., Urrutia, A. O., and Hurst, L. D. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends Genet.* 23, 480–484. doi: 10.1016/j.tig.2007.08.003
- Boutanaev, A. M., Kalmykova, A. I., Shevelyov, Y. Y., and Nurminsky, D. D. (2002). Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* 420, 666–669. doi: 10.1038/nature01216
- Chen, W. H., De Meaux, J., and Lercher, M. D. (2010). Co-expression of neighbouring genes in *Arabidopsis*: separating chromatin effects from direct interactions. *BMC Genomics* 11, 178. doi: 10.1186/1471-2164-11-178
- Costanzo, M., Vandersluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353. doi: 10.1126/science.aaf1420
- Dai, Z., Xiong, Y., Dai, Z., and Pmid, D. (2014). Neighboring genes show interchromosomal colocalization after their separation. *Mol. Biol. Evol.* 31, 1166–1172. doi: 10.1093/molbev/msu065
- Deng, Y., Dai, X., Xiang, Q., Dai, Z., He, C., Wang, J., et al. (2010). Genome-wide analysis of the effect of histone modifications on the coexpression of neighboring genes in *Saccharomyces cerevisiae*. *BMC Genomics* 11, 550. doi: 10.1186/1471-2164-11-550
- Fukuoka, Y., Inaoka, H., and Kohane, I. D. (2004). Inter-species differences of co-expression of neighboring genes in eukaryotic genomes. *BMC Genomics* 5, 4. doi: 10.1186/1471-2164-5-4
- Ghanbarian, A. T., and Hurst, L. D. (2015). Neighboring Genes Show Correlated Evolution in Gene Expression. *Mol. Biol. Evol.* 32, 1748–1766. doi: 10.1093/molbev/msv053
- Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Veronneau, S., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391. doi: 10.1038/nature00935
- Harbison, C. T., Gordon, D. B., Lee, T. I., Rinaldi, N. J., Macisaac, K. D., Danford, T. W., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99–104. doi: 10.1038/nature02800
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi: 10.1093/nar/gkh036
- Hentges, K. E., Pollock, D. D., Liu, B., and Justice, M. D. (2007). Regional variation in the density of essential genes in mice. *PLoS Genet.* 3, e72. doi: 10.1371/journal.pgen.0030072
- Hershbegg, R., Yeger-Lotem, E., and Margalit, H. D. (2005). Chromosomal organization is shaped by the transcription regulatory network. *Trends Genet.* 21, 138–142. doi: 10.1016/j.tig.2005.01.003
- Hurst, L. D., Pal, C., and Lercher, M. D. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310. doi: 10.1038/nrg1319

by Natural Science Foundation of Guangdong Province (2018A030313285), and also by Pearl River Nova Program of Guangzhou (201710010044).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01030/full#supplementary-material>

- Hurst, L. D., Williams, E. J., and Pal, C. D. (2002). Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet.* 18, 604–606. doi: 10.1016/S0168-9525(02)02813-5
- Janga, S. C., Collado-Vides, J., and Babu, M. D. (2008). Transcriptional regulation constrains the organization of genes on eukaryotic chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 105, 15761–15766. doi: 10.1073/pnas.0806317105
- Junier, I., and Rivoire, O. (2016). Conserved units of co-expression in bacterial genomes: an evolutionary insight into transcriptional regulation. *PLoS One* 11, e0155740. doi: 10.1371/journal.pone.0155740
- Kalmykova, A. I., Nurminsky, D. I., Ryzhov, D. V., and Shevelyov, Y. D. (2005). Regulated chromatin domain comprising cluster of co-expressed genes in *Drosophila melanogaster*. *Nucleic Acids Res.* 33, 1435–1444. doi: 10.1093/nar/gki281
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421, 231–237. doi: 10.1038/nature01278
- Kruglyak, S., and Tang, H. (2000). Regulation of adjacent yeast genes. *Trends Genet.* 16, 109–111. doi: 10.1016/S0168-9525(99)01941-1
- Kustatscher, G., Grabowski, P., and Rappsilber, J. (2017). Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol Syst Biol* 13, 937. doi: 10.15252/msb.20177548
- Lee, J. M., and Sonnhammer, E. D. (2003). Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.* 13, 875–882. doi: 10.1101/gr.737703
- Lercher, M. J., and Hurst, L. D. (2006). Co-expressed yeast genes cluster over a long range but are not regularly spaced. *J. Mol. Biol.* 359, 825–831. doi: 10.1016/j.jmb.2006.03.051
- Lercher, M. J., Blumenthal, T., and Hurst, L. D. (2003). Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. *Genome Res.* 13, 238–243. doi: 10.1101/gr.553803
- Li, Y. Y., Yu, H., Guo, Z. M., Guo, T. Q., Tu, K., and Li, Y. D. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Comput. Biol.* 2, e74. doi: 10.1371/journal.pcbi.0020074
- Michalak, P. D. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91, 243–248. doi: 10.1016/j.ygeno.2007.11.002
- Pal, C., and Hurst, L. D. (2003). Evidence for co-evolution of gene order and recombination rate. *Nat. Genet.* 33, 392–395. doi: 10.1038/ng1111
- Semon, M., and Duret, L. D. (2006). Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol. Biol. Evol.* 23, 1715–1723. doi: 10.1093/molbev/msl034
- Singer, G. A., Lloyd, A. T., Huminiecki, L. B., and Wolfe, K. D. (2005). Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol. Biol. Evol.* 22, 767–775. doi: 10.1093/molbev/msi062
- Spellman, P. T., and Rubin, G. D. (2002). Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* 1, 5. doi: 10.1186/1475-4924-1-5
- Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. D. (2006). BioGRID: a general repository for interaction datasets. *Acids Res.* 34, D535–D539. doi: 10.1093/nar/gkj109
- Teichmann, S. A., and Veitia, R. A. (2004). Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation

- from a dosage balance perspective. *Genetics* 167, 2121–2125. doi: 10.1534/genetics.103.024505
- Thevenin, A., Ein-Dor, L., Ozery-Flato, M., and Shamir, R. (2014). Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome. *Nucleic Acids Res.* 42, 9854–9861. doi: 10.1093/nar/gku667
- Thompson, D. A., Roy, S., Chan, M., Styczynsky, M. P., Pfiffner, J., French, C., et al. (2013). Evolutionary principles of modular gene regulation in yeasts. *eLife* 2, e00603. doi: 10.7554/eLife.00603
- Tirosh, I., Weinberger, A., Carmi, M., and Barkai, N. D. (2006). A genetic signature of interspecies variations in gene expression. *Nat. Genet.* 38, 830–834. doi: 10.1038/ng1819
- Veron, A. S., Lemaitre, C., Gautier, C., Lacroix, V., and Sagot, M. D. (2011). Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny. *BMC Genomics* 12, 303. doi: 10.1186/1471-2164-12-303
- Wang, G. Z., Chen, W. H., and Lercher, M. D. (2011). Coexpression of linked gene pairs persists long after their separation. *Genome Biol. Evol.* 3, 565–570. doi: 10.1093/gbe/evr049
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A. (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449, 54–61. doi: 10.1038/nature06107
- Weber, C. C., and Hurst, L. D. (2011). Support for multiple classes of local expression clusters in *Drosophila melanogaster*, but no evidence for gene order conservation. *Genome Biol.* 12, R23. doi: 10.1186/gb-2011-12-3-r23
- Xu, C., Chen, J., and Shen, B. (2012). The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force? *BMC Syst. Biol.* 6 Suppl 1, S21. doi: 10.1186/1752-0509-6-S1-S21
- Yang, Y. F., Cao, W., Wu, S., and Qian, W. (2017). Genetic Interaction Network as an Important Determinant of Gene Order in Genome Evolution. *Mol. Biol. Evol.* 34, 3254–3266. doi: 10.1093/molbev/msx264

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.