# Title: Transcription factor binding dynamics shape noise across biological processes


## Authors

Lavisha Parab[1†], Sampriti Pal[1†], Riddhiman Dhar[1]*

## Affiliations

[1] Department of Biotechnology, Indian Institute of Technology (IIT) Kharagpur, Kharagpur, West Bengal-721302, India.

*email: riddhiman.dhar@iitkgp.ac.in

†These authors contributed equally to this work



## Summary

Cellular processes driven by coordinated actions of individual genes generate cellular phenotypes. Stochastic variations in these processes lead to phenotypic heterogeneity that often has important implications for antibiotic persistence, mutation penetrance, cancer growth and anti-cancer drug resistance. However, the architecture of noise in cellular processes has remained largely unexplored even though expression noise in individual genes have been widely studied. Here we quantify noise in biological processes in yeast and through an integrated quantitative model show that the number of regulating transcription factors and their binding dynamics are the primary drivers of noise. Specifically, binding dynamics arising from competition and cooperation among TFs for promoter binding can predict a large fraction of noise variation. Our work reveals a novel mechanism of noise regulation that arises out of the dynamic nature of gene regulation and is not dependent on specific transcription factor or specific promoter sequence.



## Keywords

Noise, Biological processes, Integrated model, Transcription factors, Competitive binding, Cooperative binding, Stochastic simulations, Noise balance

## Introduction

Stochastic fluctuations in molecular processes operating inside a cell generate variations in the expression levels of genes which leads to variation in activities of molecular pathways. Such variations ultimately impact cellular phenotypes and generate phenotypic heterogeneity among individual cells of an isogenic population under identical environmental condition. Phenotypic heterogeneity has many important biological implications. Heterogeneity in cell growth rate and gene expression has been shown to give rise to persister cells that are able to survive antibiotic treatment (Balaban *et al.*, 2004; Rotem *et al*., 2010; Wakamoto *et al*., 2013; Arnoldini *et al,* 2014; Page and Peti, 2016). Phenotypic heterogeneity can also facilitate emergence of antibiotic resistant populations (Levin and Rozen, 2006; Bergmiller *et al.*, 2017; Levin-Reisman *et al.*, 2017; Dewachter *et al.*, 2019). Stochastic variation in expression level of genes can give rise to variable fitness effect of a mutation among individuals of a population – a phenomenon known as incomplete penetrance (Eldar *et al.*, 2009; Raj *et al*., 2010; Burga *et al.*, 2011; Dickinson *et al.*, 2016; Taeubner *et al.*, 2018). Incomplete penetrance has important implications in development of multi-cellular organisms and in human diseases (Dickinson *et al.*, 2016; Taeubner *et al.*, 2018). In addition, phenotypic heterogeneity has important role in growth and evolution of cancers (Meacham and Morrison, 2013; Nguyen *et al.*, 2019; Sharma *et al.*, 2019). Further, heterogeneity has been associated with emergence of anti-cancer drug resistance (Pisco, A. O. *et al*, 2013; Sheng, S. *et al.*, 2018; Hammerlindl and Schaider, 2018; Gupta *et al.*, 2019; Farquhar *et al.*, 2019).

Even though stochastic variations in the expression level of all genes have been systematically quantified in microbial systems such as bacteria *E. coli* and yeast *Saccharomyces cerevisiae* (Newman *et al.*, 2006; Bar-Even, A. *et al.*, 2006; Taniguchi *et al.*, 2010; Silander *et al.*, 2013), the architecture of noise in cellular pathways has remained largely unexplored beyond the traditional gene-set enrichment analysis. A quantitative understanding of the variations that exist in cellular pathways is absolutely essential for a better understanding of the origin of phenotypic heterogeneity as the large majority of cellular traits and phenotypes are polygenic in nature. Thus, these traits and phenotypes are shaped by cellular pathways driven by a coordinated functioning of a group of genes. As an example, stochastic variation in growth rate that has important implication for evolution of cancer, is likely to originate from a coordinated action of multiple genes. Therefore, the noise profile of a cellular pathway is shaped by the overall combination of expression variations of all genes in that pathway.

Molecular origins of noise in expression of individual genes have been investigated over many years (Mcadams and Arkin, 1997; Elowitz *et al.*, 2002; Blake *et al.*, 2003; Raser and O'Shea, 2004; das Neves *et al.*, 2010; Sanchez *et al.*, 2011; Hornung *et al.*, 2012; Salari *et al.*, 2012; Sanchez *et al.*, 2013; Sharon *et al.*, 2014; Chen and Zhang, 2016; Wu *et al.*, 2017; Faure *et al.*, 2017; Baudrimont *et al*., 2019). These studies have shown a correlation between the presence of TATA box motif in the promoter sequence of a gene with higher expression noise (Bar-Even *et al.*, 2006; Tirosh *et al.*, 2006; Hornung *et al.*, 2012; Ravarani *et al.*, 2016). Further, promoter nucleosome occupancy, alone as well as in conjunction with presence/absence of TATA box motif, has also been associated with expression noise (Tirosh and Barkai, 2008; Choi and Kim, 2009; Small *et al.*, 2014). Recent studies have also associated histone modification patterns with expression noise (Weinberger *et al.*, 2012; Chen and Zhang, 2016; Faure *et al.*, 2017; Nicolas *et al.*, 2018). However, even after so many years, how well these molecular factors can predict noise remains completely unknown and a predictive model of noise remains elusive.

In the current work, we systematically quantify noise in biological processes in yeast. We also build an integrated model of noise by combining all possible molecular factors that can impact noise and also test predictive ability of each of these molecular factors. We uncover that the transcription factor binding dynamics has the highest predictive power and can explain the highest fraction of noise variation existing in biological processes. We show that the number of regulatory transcription factors (TFs) of genes in a cellular process impacts noise. Further, we show that an increase in number of TFs results in overlaps in TF binding sites and leads to competition between TFs for binding to the same genomic region. Through mathematical modeling and stochastic simulations, we demonstrate that competition between TFs having varying regulation strengths can increase noise. Our work through a combination of an integrated quantitative and predictive model of noise and mathematical modeling uncovers a general mechanism of noise regulation that is not dependent on the specific details of the biological system. This mechanism instead highlights a dynamic origin of noise and can thus be extended across biological organisms.

## Results

Stochastic variations in cellular processes can generate phenotypic heterogeneity in a cell population (Fig. 1A). To quantify noise in biological processes, we analyzed the overall noise levels of 181 biological processes in yeast according to Gene Ontology (GO) classification (Ashburner *et al.*, 2000; Carbon *et al.*, 2019). We achieved this by combining individual expression noise values

of all genes in a biological process and repeated this for all biological processes. Processes showed a wide range of distributions for median noise values and noise dispersion (Fig. S1). Majority of the biological processes (~83%), however, did not have any significant deviation towards high or low noise and were balanced in their content of high- and low-noise genes (Wilcoxon rank-sum test, FDR<0.1). Only a small number of processes (~17%) showed significantly high or low noise with 22 processes showing high noise and nine processes exhibiting low noise (Wilcoxon rank-sum test, FDR<0.1) (Fig 1B; Table S1, S2). Processes showing high noise included ATP biosynthesis, proton transport coupled to ATP synthesis, tricarboxylic acid cycle, amino acid biosynthesis process and cellular transport process (Fig. 1B, Table S1). Processes exhibiting low noise were associated with transcription and translation (Fig. 1B, Table S2).

Molecular factors that have been associated with expression noise in individual genes, such as presence of TATA box sequence in the promoter (Bar-Even *et al.*, 2006; Tirosh *et al.*, 2006; Hornung *et al.*, 2012; Ravarani *et al.*, 2016) and promoter nucleosome occupancy (Tirosh and Barkai, 2008; Choi and Kim, 2009; Small *et al.*, 2014) were poor predictors of noise in biological processes (Fig. 1C), even though they showed significant correlation with noise (Figure S2 A-F). Specifically, TATA box sequence, promoter nucleosome occupancy alone and in conjunction with TATA box sequence, could only explain ~11%, ~22% and ~24% of the noise variation in biological processes respectively and had low predictive power (predicted $R^2$=0.06±0.21 for TATA box alone, predicted $R^2$=0.09±0.22 for promoter nucleosome occupancy alone, predicted $R^2$=0.13±0.21 for TATA box +promoter nucleosome occupancy; Fig. 1C, Fig. S2G). This suggested that these factors were largely associated with noise and not predictive.

Thus, to systematically identify the key molecular factors that could predict noise in biological processes, we built an integrated model considering all factors that are known or are likely to influence gene expression. These factors included diverse aspects of gene regulation by transcription factors (TFs) such as TF binding site location, expression and noise levels of TFs, TF binding site histone modifications; tRNA adaptation index, histone levels and histone modification patterns across the promoter region as well as gene body, and three-dimensional configuration of the genome in addition to the factors such as presence of TATA-box sequence, nucleosome occupancy in the promoter and gene body (Fig. 1D).

## Genes in high-noise processes are regulated by higher number of transcription factors

Surprisingly, our integrated model revealed a strong correlation between number of regulating transcription factors and noise (r=0.54, p=3.1×10-15; Fig. 2A). The genes in high-noise processes on average had >50% more regulatory transcription factors compared to the genes in the low-noise processes (p=0.002; Fig. 2A). This increase in number of regulatory transcription factors resulted in an increase in number of TF binding sites upstream of the start codon (Fig. S3A). In addition, genes in high-noise processes showed an upstream shift in the point of lowest nucleosome occupancy (Fig. S3B) and this also corresponded with a ~15bp upstream shift of transcription start site (Fig. S3C), suggesting variation in TF binding patterns as a likely driver of noise.

The increase in number of regulatory TFs led to an increase in percentage of TFs that bound co-operatively to the promoter region (Yang *et al*., 2010; Chen *et al*., 2012), with ~30% of the TFs showing co-operative binding in genes of high-noise biological processes compared to ~20% of the TFs exhibiting co-operative binding in genes of low-noise processes (p=0.019; Fig. 2B; Fig. S4A). In addition, the increase in regulatory TFs increased the percentage of activators of genes in high-noise processes whereas the percentage of repressors remained unchanged (Fig. 2C, D). Genes belonging to high-noise processes also showed an increase in the percentage of co-expressing TFs that had positive expression correlations among themselves (Fig. S4B, C). These results suggested possible roles for co-operatively binding TFs, activators and co-expressing TFs in noise regulation.

## Overlaps in TF binding sites lead to higher noise

Interestingly, increase in number of regulatory TFs and a corresponding increase in their binding sites led to a substantial increase in overlap of TF binding sites in the promoter region, with the number of overlaps increasing from ~0.25 overlaps per TF binding site in genes of low-noise processes to ~0.44 overlaps per binding site in genes of high-noise processes (p=0.0004; Fig. 2E). Most of the TF binding site overlaps were observed up to 600 bp upstream of the start codon and genes in the high-noise processes showed significantly higher percentage of overlaps in the region between 300 bp to 600bp upstream of the start codon compared to the low-noise processes (Fig. 2F). Most interestingly, genes in high noise processes showed a two-fold increase in percentage of co-expressing TFs whose binding sites had overlap (p=0.0002; Fig. 2G) compared to genes in low-noise processes.

Overlap in binding sites of co-expressing TFs causes competition between these TF molecules for binding to the same region of the promoter. Binding competition can generate noise in two possible ways. First, TF binding is a dynamic process consisting of rapid binding and unbinding steps (Burger *et al.*, 2010; Das *et al.*, 2017). Thus, competition between TFs leads to a scenario where a gene is regulated by different subsets of TFs at different points of time, thus generating temporal variation in TF binding (Fig. 3A). In presence of TFs that differ in their strengths of regulation, such temporal variation in TF binding leads to variation in transcription rate. Second, asynchronous temporal variation in TF binding among individual cells in a population leads to inter-individual variation in TF binding (Fig. 3A). Again, in presence of TFs exhibiting variations in strength of gene regulation, this can generate inter-individual variation in transcription rate.

We compared the impact of co-operative and competitive binding of TFs to promoter region on noise through mathematical modeling and stochastic simulations (Fig. S5A). Co-operative binding required co-ordination between two or more TFs which led to a reduction in burst frequency even though burst size remained the same (Fig. 3B, Fig. S5B). In contrast, competitive TF binding did not alter burst frequency but changed average burst size as well as introduced higher variability in burst size when regulated by TFs of different strengths (Fig. 3B). At the protein level both co-operative and competitive binding of TFs led to substantially higher noise (Fig. 3C, Fig. S5B).

Interestingly, our data showed that the variation in strengths of regulation among activators was ~50% higher for genes in high-noise processes ($p=0.0001$, Fig 3D) in addition to higher mean strength of activation (Fig. S6A). Further, the average strength of regulation of activators that bound to overlapping sites and the difference in strength of regulation between these activators were higher in genes of high-noise processes (Fig. 3E, Fig. S6F). These observations in conjunction with our mathematical model suggested that the competition between TFs for binding and variations in their regulation strengths could generate noise.

Further, the response correlation between the dosage of regulating TFs and the dosage of the regulated gene was also different between genes from high- and low-noise processes. Genes of high-noise processes showed on average higher response correlation to TFs as well as exhibited higher variation in response correlation (Fig. 3F, Fig. S6C, D, G). Increased variation in response

correlation implicates that the regulated genes respond differently to dosages of different TFs and thereby, can increase noise.

Overlaps in binding sites for different TFs can also lead to degeneracy of TF binding sequences to accommodate different consensus binding motifs of different TFs. Such degeneracy can change binding affinity of TFs to the DNA and can lead to noisy transcription. However, we did not see any difference in binding site degeneracy among genes of high and low- noise processes (Fig. S6H).

**Genes in high-noise processes are regulated by noisy transcription factors**

Variation in the expression level of the TFs themselves can impact the strength of gene regulation among isogenic cells and in turn, can facilitate noise propagation through biological processes. Noise in biological processes was correlated with the median noise levels of the regulatory TFs (Fig. 4A) and genes in high-noise processes showed an increase in high-noise regulatory TFs compared to genes in low-noise processes (Fig. S7). Further, the activators of genes in high-noise processes also showed substantially higher median noise (+0.052) compared to that of low-noise processes (median noise of -0.125) ($p=0.012$, Fig. 4B), further highlighting the role of activators and variations in their expression levels on regulation of noise in cellular processes.

We quantified the abilities of all the aspects of transcription factor binding dynamics to predict noise (Fig. 4C). Among all the features, TF binding site overlap had the highest predictive ability with predicted $R^2=0.45\pm0.3$ and could explain ~51% (±4%) of the noise alone (Fig. 4C). All aspects of TF binding dynamics together exhibited good predictive power with predictive $R^2=0.51\pm0.25$ and explained ~73% (±5%) of the variation in noise across biological processes (Fig 4C).

We also assessed the predictive power of histone modification patterns and 3D genomic contacts. However, histone modification pattern could only explain ~43% of the noise variation with a predictive power given by predicted $R^2=0.22$ (Fig. S8A-D). The number of 3D genomic contacts, both intra-chromosomal and inter-chromosomal, did not show any variation among genes in high- and low- noise processes and was unable to predict noise (Fig. S8E, F)

We further assessed the extent to which each of the molecular factors individually could explain and predict noise in biological processes and ranked them according to their impact (Fig. S9A, B). Among the top factors that showed highest predictive power were response correlation between TFs and their regulated genes, difference in regulation strength of activators that bound to overlapping sites and the average regulation strength of these activators (Fig. S9A). In addition, the top factors that explained the largest fraction of noise variation also included response correlation of TFs and their regulated genes and regulation strength of activators that bound to overlapping sites (Fig. S9B). These results emphasized the importance of competition between TFs for binding to overlapping sites, their regulation strengths and the gene-TF response correlations for noise regulation.

The integrated model comprising of all factors was able to explain ~83% (±2%) of the variation in median noise as well as ~80% (±2%) of the variation in noise dispersion across biological processes (Fig. 4D). In addition, this model was able to predict the median noise across biological processes with predicted $R^2$ value of 0.65 (±0.26) and the noise dispersion with predicted $R^2$ value of 0.57 (±0.36) (Fig. 4D, E).

**Noise-balance in biological processes is achieved by properties of regulatory TFs**

Interestingly, cellular processes with higher median noise also showed higher noise dispersion with high median noise processes exhibiting approximately four-fold higher dispersion compared to low-noise processes (Fig. 4F). To test whether such a relationship between noise dispersion and median noise can arise out of chance alone, we performed a randomization and calculated the net noise dispersion for a process from the difference between dispersions of genes with positive noise and negative noise values (Fig. 4G). Interestingly, 90% of the processes that did not exhibit high or low noise showed very different actual net dispersion compared to the random expectation (Fig. 4G), suggesting that the distribution of noise values of genes in biological processes has been shaped by evolution. About 69% of the processes showed low net dispersion (LND) and these processes exhibited a balanced content of `positive and negative noise genes. Positive and negative noise genes in these processes showed no difference in the number of regulatory TFs, contained similar number of TF binding site overlaps and showed no difference in number of co-operative TFs (Fig. S10A-C). These results highlighted a pivotal role of the properties of the regulatory TFs of positive and negative noise genes for establishing noise balance in cellular processes (Fig. S10, S11).

## Discussion

In summary, through an integrated quantitative analysis we have shown that the noise in biological processes can be largely explained and predicted by the transcription factor binding dynamics. Noisy processes tend to be controlled by a larger number of TFs. These include a substantial fraction of TFs that bind co-operatively to the promoter region and the coordination required between these TFs for binding can drive noise. In addition, increase in number of TFs leads to overlapping TF binding sites which can lead to competition between TFs for binding to the same promoter region. This can give rise to a time-dependent variation in TF binding within a cell and can generate noise when TFs of different strengths of regulation are present. In addition, such competition can generate inter-individual differences in TF binding pattern and in presence of TFs of varied regulation strengths can generate noise. Although competition between specific interacting partners of TATA binding protein has been implicated in noise before (Ravarani *et al.*, 2016), we uncover a much-more general molecular mechanism of noise generation that is not dependent on any specific TF or any specific promoter sequence.

We also show that a large majority of the biological processes is noise-balanced in terms of containing genes of high and low- noise. This noise balance is achieved due to properties of regulatory TFs of positive and negative noise genes. Previous studies have reported several strategies for noise minimization in genes and cellular processes (Fraser *et al.*, 2004; Lehner, 2008; Battich *et al.*, 2015; Peng *et al.*, 2016). Our work reveals that noise can also be regulated at the level of biological processes by a balanced content of positive and negative noise genes. Such balance can attenuate overall noise in biological processes and can minimize detrimental effects of high noise.

Although our integrated model can predict a large fraction of noise variation, there is, however, still a substantial fraction of noise that cannot be predicted with our model. This could be because of two reasons. Firstly, it is possible that several other molecular factors which can generate noise have not been accounted for in our model. Some of these molecular factors may have not been discovered yet. Second, the inherent randomness of molecular processes might not allow us to completely predict the noise variation. This aspect of noise could thus limit our predictive capability.

Finally, our findings have major implications for noise prediction in cellular processes across biological organisms. Recent explosion of genomics data has led to genome-wide characterization of TF binding sites across a diverse range of organisms. Thus, the knowledge of TF binding sites maps and overlaps can enable us to predict noise in the regulated genes and subsequently in the associated cellular processes. In addition, with increasing availability of genome-wide nucleosome occupancy maps, histone modification patterns and three-dimensional genome configuration data, our study also provides a framework for building integrated model of gene expression and its noise across a diverse range of organisms. Stochastic variations in molecular processes are ubiquitous in cells across biological systems and have major implications in human diseases and their treatment including cancer evolution and drug resistance emergence. Thus, increased ability to predict variations in biological processes will be extremely useful in predicting the extent of heterogeneity in cellular traits and phenotypes. This will ultimately help us better predict disease progression and evolution and will enable us to devise better treatment strategies.

## Materials and Methods

### Noise analysis in biological processes

Biological processes were classified according to the Gene ontology classification for *S. cerevisiae* obtained from http://geneontology.org/ (data version: releases/2017-11-22) (Ashburner *et al.*, 2000; Carbon *et al.*, 2019). Gene names were mapped to GO terms using SGD yeast gene association map (https://www.yeastgenome.org/, GOC Validation Date: 11/10/2017) (Cherry *et al.*, 2012). Only terms associated with Biological Processes (BP) were included in our analysis and if gene expression noise data for at least 10 genes were available for that class. Noise values for individual genes in yeast was obtained from Newman *et al.*, 2006 and the DM values in the SD medium were used for all noise analysis. For each included GO term, median noise value and noise dispersion (through calculation of inter-quartile range) were calculated. To identify processes with high or low noise, one-sample Wilcoxon signed-rank test was performed to check whether the noise values of genes in a biological process differed significantly from zero followed by multiple hypothesis testing correction using Benjamini-Hochberg method (Benjamini and Hochberg, 1995). A False Discovery Rate (FDR) cut-off of 10% was used for identifying high and low-noise processes.

### Building an integrated model of noise in yeast

The integrated model of noise was generated by carefully considering all factors that can regulate gene expression. These included several aspects of TF binding dynamics, presence of TATA box

sequence in the promoter, absolute nucleosome occupancy levels in the promoter and the gene body, histone modification patterns in the promoter and the gene body, variations in transcription start sites (TSS), tRNA adaptation index (tAI) and three-dimensional contacts in the genome (Fig. 1D). Promoters with the presence of TATA box sequence were identified from the data of Rhee and Pugh, 2012. For each biological process, percentage of genes with TATA box sequence in their promoters was calculated. The genome-wide absolute nucleosome occupancy level for yeast genome was obtained from Oberbeckmann and Wolff *et al*., 2019 and average nucleosome occupancy level per nucleosome occupied site for all promoter regions and gene bodies were calculated. The region from 1000bp upstream to 10bp downstream of the start codon was considered to be the promoter region of a gene. Average nucleosome occupancy level per occupied site calculated between -1000bp to -900bp region of the start codon was shown as the nucleosome occupancy at -1000bp. Similarly, average nucleosome occupancy level per site was calculated from -900 to -800bp, -800 to -700bp, -700 to -600bp, -600 to -500bp, -500 to -400bp, -400 to -300bp, -300 to -200bp, -200 to -150bp, -150 to -100bp, -100 to -50bp and -50bp to +10bp regions.

Transcription start site (TSS) data for all yeast genes were obtained from Lu and Lin, 2019. Closest TSS site for each gene was obtained, and a spread of potential TSS sites for all genes were calculated. Genome-wide histone modification data for yeast was obtained from Pokholok *et al.*, 2006 and all modifications were mapped to Gene body, promoter region, and transcription factor binding sites. Data on three-dimensional model of yeast genome was obtained from Duan and Andronescu *et al*., 2010. Number of intra- and Inter- chromosomal contacts for all genes and promoter regions were quantified. tRNA adaptation index for all genes were calculated following the method of Tuller *et al*., 2010. For all molecular factors, the value of a factor in a biological process was obtained by calculating the median values of the factor for all genes in the process.

**Analysis of transcription factor binding**

The list of transcription factors for all yeast genes were obtained from Yeastract database (http://www.yeastract.com/) (Teixeira *et al.*, 2018) and for a gene, only those TFs for which experimental evidence for DNA binding has been obtained or the knockout of TF has been experimentally shown to impact expression of the gene (Teixeira *et al.*, 2018) were considered. To map the binding sites of all TFs to promoter regions of genes, the consensus motif sequences for all TFs were obtained from YeTFaSCo database (de Boer and Hughes, 2012). All position weighted matrices for all motifs were obtained and all possible combinations of bases were considered. Positions of all such motif sequences of all regulatory TFs of a gene were identified in the promoter

region (ranging from -1000bp to +10bp of the start codon) after allowing for maximum two mutations in the consensus sequence. Overlaps between TF binding sites in promoter regions were mapped. The list of co-operatively binding TFs in yeast were obtained from Yang *et al*., 2010 and Chen *et al*., 2012. If the noise value for a TF was more than 0.5 (DM-value), the TF was considered to be a high-noise TF. Similarly, a TF showing a noise value of less than -0.5 (DM-value) was considered to be a low-noise TF.

**Gene-transcription factor (TF) and TF-TF expression correlation analysis**

Gene expression data measured through RNA sequencing from Dhar *et al*., 2019 (NCBI GEO dataset id 104343) was used to calculate all pairwise gene-TF and all pairwise TF-TF (of a gene) expression correlations. Significant positive correlation ($p<0.05$) between a gene and a regulatory TF indicated that the TF acted as an activator for the gene since the expression of the gene increased with increase in expression of the TF. Similarly, Significant negative correlation between a gene and its regulatory TF indicated that the TF acted as a repressor for the gene. The value of correlation coefficient between a gene and a TF (if significant) was taken as the response correlation and the slope of the line was considered as the strength of regulation of the TF.

In addition, pairwise expression correlations between all TFs of a gene were calculated. If a TF showed significant positive correlation with at least three other TFs, the TF was considered to be a positively correlated (co-expressing) TF. Similarly, if a TF showed significant negative correlation with at least three other TFs, the TF was considered to be a negatively correlated TF.

**Modeling and stochastic simulation of transcription factor-DNA binding process**

The dynamics of TF binding to DNA was captured using a two-state model with consideration of rapid binding and unbinding of TF to DNA. The binding-unbinding of TF to DNA was considered to be a poisson process and thus, the time intervals between two successive bindings (or two successive unbindings) were exponentially distributed. The dynamics of co-operative binding and competitive binding of TFs was compared to the dynamics of the single TF regulation system. The co-operative binding was modeled by co-operative interactions between two TFs whereas competitive binding was modeled using competition between three TFs. The binding rate $K_a$ and dissociation rate $K_d$ were considered to be the same for all TFs across three scenarios. As the concentrations of TFs can impact the chances of binding, the concentration of TF in single TF binding scenario was considered to be same as the concentration of each of the co-operatively binding TFs. Further, the concentration of the TF in single TF binding scenario was considered to

be equal to the sum of the concentrations of three TFs in case of competitive binding scenario. Binding of TF led to ON state and resulted in transcription leading to accumulation of protein (Supplementary fig. 5A).

The dynamics of transcription and translation were modeled using the following equations.

$Changes\ in\ mRNA\ conc.\ over\ time$: $\frac{d[mRNA]}{dt} = \beta_m - \alpha_m \times [\mathrm{mRNA}]$

where $\beta_m$ denoted the transcription rate per unit time (or burst size) and $\alpha_m$ denoted the removal rate of mRNA due to degradation and dilution.

Similarly,

$Changes\ in\ protein\ conc.\ over\ time$: $\frac{d[P]}{dt} = \beta_p \times [\mathrm{mRNA}] - \alpha_p [\mathrm{P}]$

where $\beta_p$ denoted the protein production rate from mRNA and $\alpha_p$ denoted the protein removal rate. The parameters $\beta_p$ , $\alpha_m$ , $\alpha_p$ were considered to be same across all cases of TF binding. In case of co-operative TF binding, $\beta_{m,coop}$ was assumed to be the same as the $\beta_{m,single.}$ In case of competitive TF binding, the production rates varied between three TFs, with $\beta_{m1} = 1.4 \times \beta_{m,single}$ , $\beta_{m2} = \beta_{m,single}$ , and $\beta_{m3} = 0.7 \times \beta_{m,single}$ .

Stochastic simulations were performed using Gillespie's algorithm (Gillespie, 1977) to decipher the dynamics of TF-DNA binding in all three scenarios and to investigate the impact of co-operative and competitive TF binding on noise. The behavior of the system was tracked at small discrete time intervals $\Delta t$ from the initial time point t. These resulted in observations at n time points t, t+ $\Delta t$, t+2×$\Delta t$, …, t+n×$\Delta t$. Any event of binding or unbinding occurring within a time interval was noted and resulted in changes in transcription rate which eventually led to a change in protein concentration (Supplementary fig. 5B). Binding of TFs led to transcription and increase in mRNA and protein concentration according to the above equations. As the time interval $\Delta t$ was considered to be small, the equations modeling the behavior of the systems was simplified as

$$[\mathrm{mRNA}]_{t+\Delta t} = [\mathrm{mRNA}]_t + (\beta_m - \alpha_m \times [\mathrm{mRNA}]_t) \times \Delta t$$

and

$$[\mathrm{P}]_{t+\Delta t} = [\mathrm{P}]_t + (\beta_p \times [\mathrm{mRNA}]_t - \alpha_p \times [\mathrm{P}]_t) \times \Delta t$$

$\beta_m$ , $\alpha_m$ , $\beta_p$ and $\alpha_p$ were expressed in the units of $\Delta t$ for further simplification of these equations. The dynamics of transcription, variation in the mRNA concentration and variation in protein concentration with time were modeled across 10,000 cells (Supplementary fig. 5B). Noise was expressed as coefficient of variation (CV) from the calculation of mean and standard deviation in the protein level across these 10,000 cells and across all individual time points.

## Identification of noise-balanced processes

To identify noise-balanced processes, the following randomization was performed. For each biological process, genes with their individual expression noise values were randomly distributed to the biological process keeping the number of genes in the class constant. Next, the noise dispersions for genes with positive noise (DM) (Positive noise genes) and negative noise values (negative noise genes) were calculated through separate quantification of interquartile range for positive and negative noise genes. The difference between dispersion values for positive and negative noise genes was considered as the net noise dispersion value for each process. Processes that showed higher net dispersion value than the maximum net dispersion obtained from randomization were termed as high net dispersion (or HND) processes, whereas the processes that showed lower net dispersion value than the minimum net dispersion value obtained from randomization were termed as low net dispersion (LND) processes.

## Statistical analysis

For all correlation analysis, Pearson's correlation coefficient was calculated. For all statistical comparisons between high- and low- noise processes Mann Whitney U test was performed. For all comparison between positive and negative noise genes in LND and HND processes Mann Whitney U test was performed. Pearson's correlation coefficient values were reported in all correlation analysis.

## Linear regression analysis

To estimate the percentage of variation explained by each factor, linear regression analysis was performed on our noise dataset. To estimate the impact of each factor, the data was divided into training and test sets, the training set consisted of 80% of the whole dataset whereas the test set contained the remaining 20% of the dataset. Linear regression was performed on the training set to obtain the percentage of variation explained. The linear regression model thus built was applied on test dataset to obtain predicted $R^2$ value and RMSE values. This process of training and prediction was repeated 1000 times to obtain mean and standard deviation values for percentage of variation explained as well as to get predicted $R^2$ and RMSE values.

For estimating the predictive power of a class of factors on noise, a variable selection method using Ridge and Lasso regression was performed to minimize the problems of multi-collinearity and

overfitting. Ridge regression was performed with the R package 'ridge' (Cule *et al*., 2020), appropriate number of principal components was chosen and factors showing significant effect on noise were identified. Lasso regression was performed using the R package 'glmnet' (Friedman *et al*., 2010) and factors with non-zero coefficients were identified as factors influencing noise. Next, Linear regression with the selected factors was performed on the training set to obtain the percentage of variation explained. The linear regression model obtained was then applied on the test data to obtain predicted values along with predicted $R^2$ and RMSE values. This process of training and prediction was repeated 1000 times to obtain mean and standard deviation values for percentage of variation explained, predicted $R^2$ and RMSE values.

In addition to the original dataset, two filtered datasets were created with reduced number of variables – one filtered on correlation and another filtered on impact. The first filtered dataset was created by removing factors that did not show significant correlation ($p<0.05$) with process noise. To create the second filtered set, first, the impact of all individual factors on noise were obtained by linear regression as described above. Only the factors that showed significant impact (explained at least 5% of the noise variation or had predicted $R^2$ of more than 0.05) were retained in the filtered set. Linear regression was performed on these datasets as described above to obtain percentage of variation explained and $R^2$ values for prediction. The analysis showing best results for percentage of variation explained and predicted $R^2$ were reported.

## References


Arnoldini, M. *et al.* Bistable expression of virulence genes in salmonella leads to the formation of an antibiotic-tolerant subpopulation. *PLoS Biol.* **12**, (2014).

Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000)

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as phenotypic switch. *Science* **305**, 1622–1625 (2004)

Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nat. Genet.* **38**, 636–643 (2006).

Battich, N., Stoeger, T. & Pelkmans, L. Control of Transcript Variability in Single Mammalian Cells. *Cell* **163**, 1596–1610 (2015).

Baudrimont, A., Jaquet, V., Wallerich, S., Voegeli, S. & Becskei, A. Contribution of RNA Degradation to Intrinsic and Extrinsic Noise in Gene Expression. *Cell Rep.* **26**, 3752-3761.e5 (2019).

Bergmiller, T. *et al.* Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science* **356**, 311–315 (2017).

Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).

Burga, A., Casanueva, M. O. & Lehner, B. Predicting mutation outcome from early stochastic variation in genetic interaction partners. *Nature* **480**, 250–253 (2011).

Burger, A., Walczak, A. M. & Wolynes, P. G. Abduction and asylum in the lives of transcription factors. *Proc. Natl. Acad. Sci. USA* **107**:4016-21 (2010)

Carbon, S. *et al.* The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019)

Chen, M.J. *et al.* De novo motif discovery facilitates identification of interactions between transcription factors in Saccharomyces cerevisiae. *Bioinformatics* **28**, 701-8. (2012)

Chen, X. & Zhang, J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell Syst.* **2**, 347–54 (2016).

Cherry, J. M. *et al.* Saccharomyces Genome Database: The genomics resource of budding yeast. *Nucleic Acids Res.* **40**, (2012)

Choi, J. K. & Kim, Y. J. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat. Genet.* **41**, 498–503 (2009)

Cule, E., Moritz, S., & Frankowski, D. ridge: Ridge Regression with Automatic Selection of the Penalty Parameter. R package version 2.5, https://CRAN.R-project.org/package=ridge. (2020)

das Neves, R. P. *et al.* Connecting variability in global transcription rate to mitochondrial variability. *PLoS Biol.* **8**, e1000560 (2010).

Das, D., Dey, S., Brewster, R. C. & Choubey, S. Effect of transcription factor resource sharing on gene expression noise. *PLoS Comput. Biol.* **13**, e1005491 (2017).

de Boer, C. G. & Hughes, T. R. YeTFaSCo: a database of evaluated yeast transcription factor sequence specificities. *Nucleic Acids Res*. **40**, D169-79 (2012).

Dewachter, L., Fauvart, M. & Michiels, J. Bacterial Heterogeneity and Antibiotic Survival: Understanding and Combatting Persistence and Heteroresistance. *Molecular Cell* **76**, 255–267 (2019)

Dhar, R., Missarova, A. M., Lehner, B. & Carey, L. B. Single cell functional genomics reveals the importance of mitochondria in cell-to-cell phenotypic variation. *Elife* **8**, (2019)

Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).

Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).

Eldar, A. *et al.* Partial penetrance facilitates developmental evolution in bacteria. *Nature* **460**, 510–514 (2009).

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).

Farquhar, K. S. *et al.* Role of network-mediated stochasticity in mammalian drug resistance. *Nat. Commun.* **10**, 2766 (2019)

Faure, A. J., Schmiedel, J. M. & Lehner, B. Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst.* **5**, 471-484.e4 (2017)

Fraser, H. B., Hirsh, A. E., Giaever, G., Kumm, J. & Eisen, M. B. Noise Minimization in Eukaryotic Gene Expression. *PLoS Biol.* **2**, e137 (2004).

Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010)

Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977)

Gupta, P. B., Pastushenko, I., Skibinski, A., Blanpain, C. & Kuperwasser, C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* **24**, 65–78 (2019)

Hammerlindl, H. & Schaider, H. Tumor cell-intrinsic phenotypic plasticity facilitates adaptive cellular reprogramming driving acquired drug resistance. *J. Cell Commun. Signal.* **12**, 133–141 (2018)

Hornung, G. *et al.* Noise-mean relationship in mutated promoters. *Genome Res.* **22**, 2409–17 (2012).

Lehner, B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol. Syst. Biol*. **4**, 170 (2008).

Levin, B. R. & Rozen, D. E. Non-inherited antibiotic resistance. *Nature Reviews Microbiology* **4**, 556–562 (2006).

Levin-Reisman, I. *et al.* Antibiotic tolerance facilitates the evolution of resistance. *Science* **355**, 826–830 (2017).

Lu, Z. & Lin, Z. Pervasive and dynamic transcription initiation in Saccharomyces cerevisiae. *Genome Res.* **29**, 1198–1210 (2019)

Mcadams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 814–819 (1997)

Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013)

Newman, J. R. S. *et al.* Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840–846 (2006).

Nguyen, A., Yoshida, M., Goodarzi, H. & Tavazoie, S. F. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nat. Commun.* **7**, 11246 (2016)

Nicolas, D., Zoller, B., Suter, D. M. & Naef, F. Modulation of transcriptional burst frequency by histone acetylation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 7153–7158 (2018).

Oberbeckmann, E. *et al.* Absolute nucleosome occupancy map for the *Saccharomyces cerevisiae* genome. *Genome Res.* **29**, 1996–2009 (2019).

Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* **12**, 208–214 (2016).

Peng, W., Song, R. & Acar, M. Noise reduction facilitated by dosage compensation in gene networks. *Nat. Commun*. **7**, 1–8 (2016).

Pisco, A. O. *et al.* Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat. Commun.* **4**, 2467 (2013)

Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005)

Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–8 (2010).

Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science.* **304**, 1811–1814 (2004)

Ravarani, C. N. J., Chalancon, G., Breker, M., De Groot, N. S. & Babu, M. M. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat. Commun.* **7**, 10417 (2016).

Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).

Rotem, E. *et al.* Regulation of phenotypic variability by a threshold-based mechanism underlies bacterial persistence. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12541–6 (2010).

Salari, R. *et al.* Teasing Apart Translational and Transcriptional Components of Stochastic Variations in Eukaryotic Gene Expression. *PLoS Comput. Biol.* **8**, e1002644 (2012)

Sanchez, A., Garcia, H. G., Jones, D., Phillips, R. & Kondev, J. Effect of Promoter Architecture on the Cell-to-Cell Variability in Gene Expression. *PLoS Comput. Biol.* **7**, e1001100 (2011).

Sanchez, A., Choubey, S. & Kondev, J. Regulation of Noise in Gene Expression. *Annu. Rev. Biophys.* **42**, 469–491 (2013).

Silander, O. K. *et al.* A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli. *PLoS Genet.* **8**, e1002443 (2012).

Sharma, A. *et al.* Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors. *Cell Rep.* **29**, 2164-2174.e5 (2019)

Sharon, E. *et al.* Probing the effect of promoters on noise in gene expression using thousands of designed sequences. *Genome Res.* **24**, 1698–1706 (2014)

Sheng, S. *et al.* Tackling tumor heterogeneity and phenotypic plasticity in cancer precision medicine: our experience and a literature review. *Cancer Metastasis Rev.* **37**, 655–663 (2018)

Small, E. C., Xi, L., Wang, J. P., Widom, J. & Licht, J. D. Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2462–E2471 (2014)

Taeubner, J. *et al.* Penetrance and Expressivity in Inherited Cancer Predisposing Syndromes. Trends in Cancer **4**, 718–728 (2018)

Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science.* **329**, 533–538 (2010).

Teixeira, M. C. *et al.* YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae. *Nucleic Acids Res*. **46**, D348–D353 (2018).

Tirosh, I., Weinberger, A., Carmi, M. & Barkai, N. A genetic signature of interspecies variations in gene expression. *Nat. Genet.* **38**, 830–834 (2006)

Tirosh, I. & Barkai, N. Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* **18**, 1084–1091 (2008).

Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–354 (2010)

Wakamoto, Y. *et al.* Dynamic persistence of antibiotic-stressed mycobacteria. *Science* **339**, 91–95 (2013).

Weinberger, L. *et al.* Expression Noise and Acetylation Profiles Distinguish HDAC Functions. *Mol. Cell* **47**, 193–202 (2012).

Wu, S. *et al.* Independent regulation of gene expression level and noise by histone modifications. *PLoS Comput. Biol.* **13**, e1005585 (2017).

Yang, Y., Zhang, Z., Li, Y., Zhu, X. G. & Liu, Q. Identifying cooperative transcription factors by combining ChIP-chip data and knockout data. *Cell Research* **20**, 1276–1278 (2010)

## Acknowledgments

Work in the lab of RD was funded by an ISIRD grant from IIT Kharagpur and the Early career research (ECR) grant (ECR/2017/002328) from Science and Engineering research Board (SERB), India.

## Author contributions

RD conceived the study, RD, LP, and SP performed data analysis, RD wrote the manuscript with inputs from LP and SP. All authors read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

Figure 1

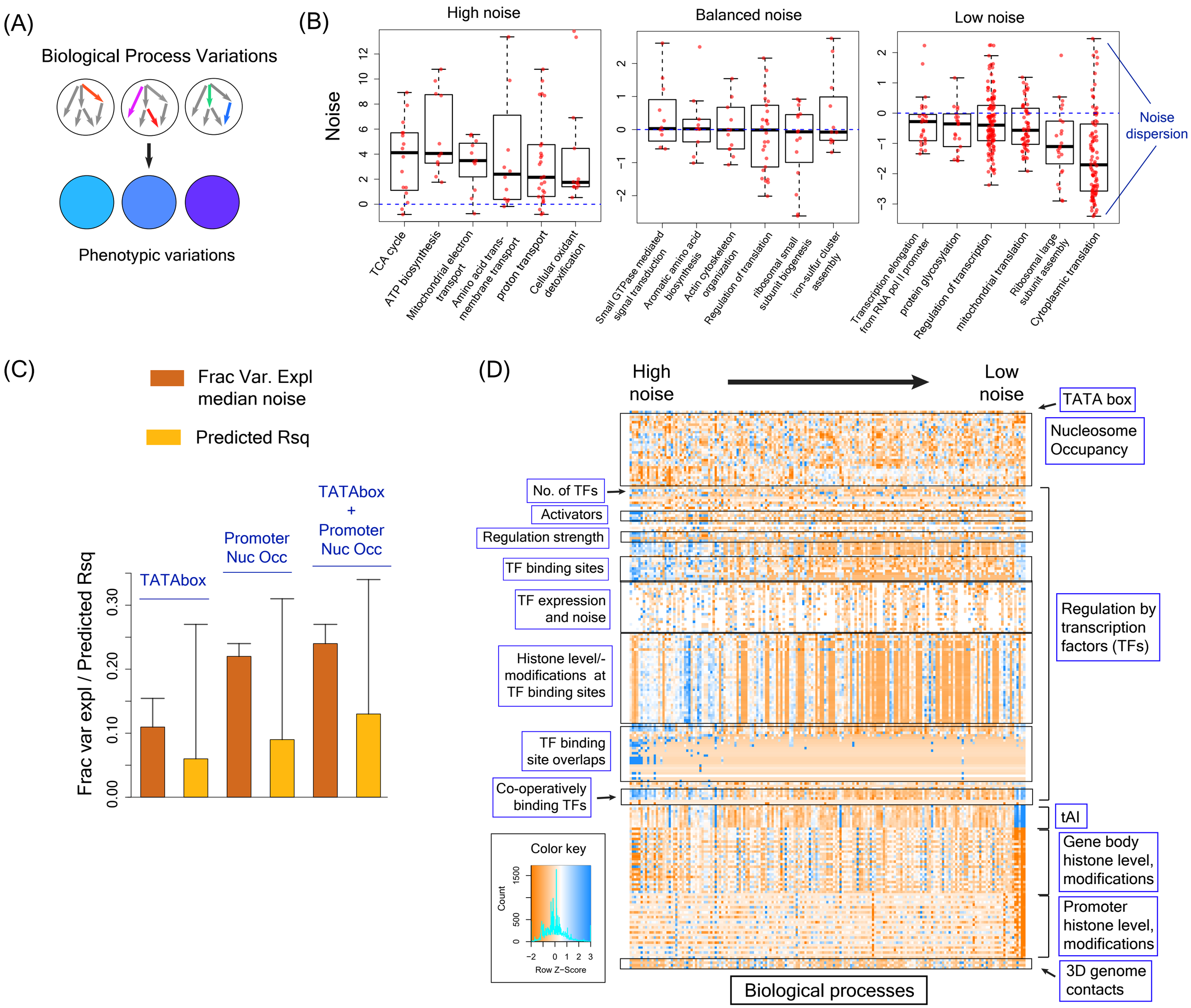



**Figure 1: An integrated model of noise in biological processes. (A)** Stochastic variations in biological processes lead to phenotypic heterogeneity in an isogenic population **(B)** Examples of biological processes with high, low and balanced noise with the underlying distribution of gene noise values. **(C)** Percentage of variation in median noise explained and predicted $R^2$ value by presence or absence of TATA box sequence, average promoter nucleosome occupancy per site and the combination of presence or absence of TATA box sequence with promoter nucleosome occupancy. **(D)** An integrated model of noise in biological processes constructed considering TATA box sequence, absolute nucleosome occupancy levels, gene regulation by transcription factors, tRNA adaptation index, histone modification pattern in genebody and promoter region and 3D genomic contacts. The heatmap shows values of all these factors (scaled and centered) in 181 biological processes (represented in the columns) sorted according to noise level.

Figure 2

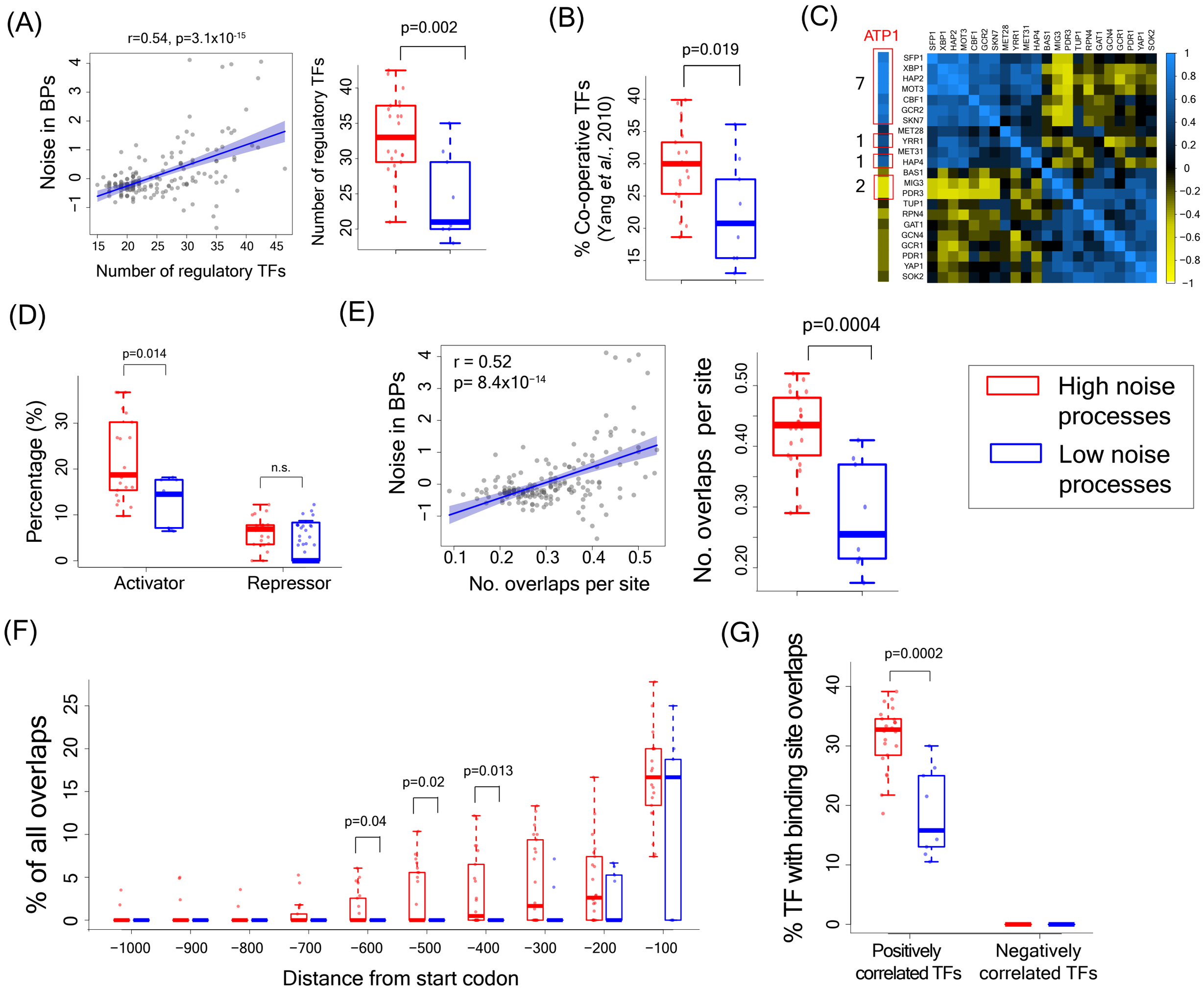
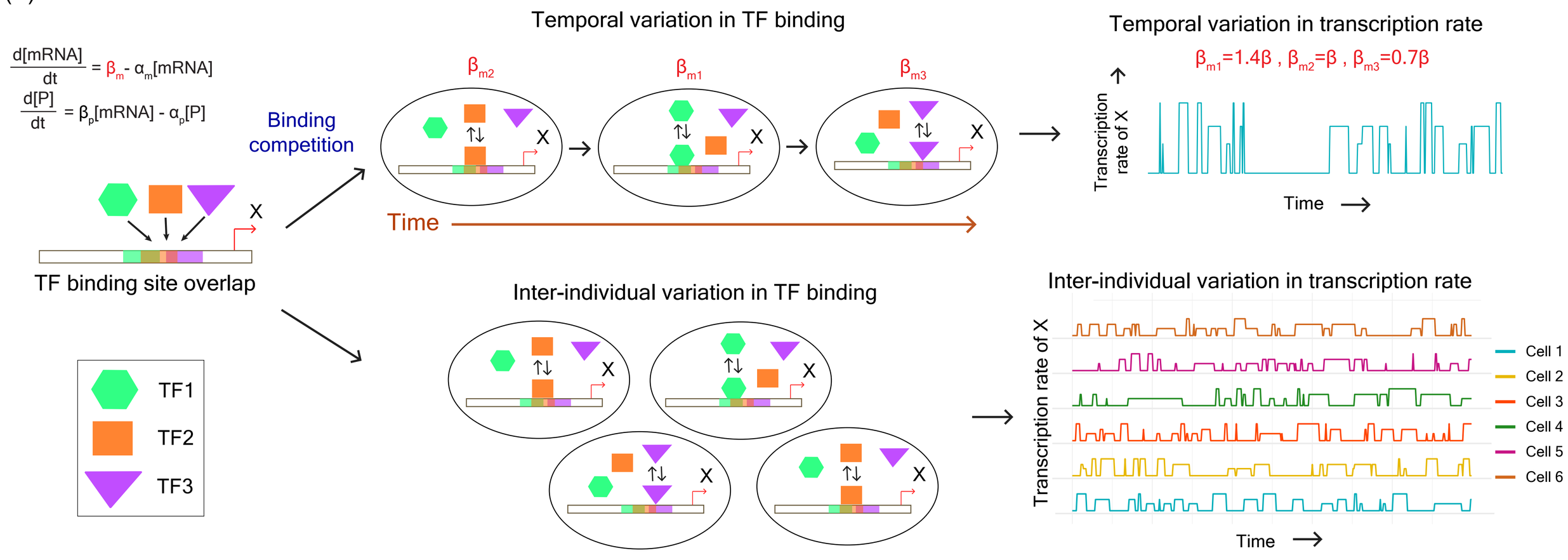



**Figure 2: Genes in high noise processes show higher percentage of co-operatively binding TFs, activators and overlaps in TF binding sites. (A)** Correlation between noise in biological processes and number of regulatory TFs of genes. **(B)** The percentage of co-operatively binding TFs of genes in high- and low- noise processes. **(C)** Expression correlation between a gene (ATP1) and all its regulatory TFs identifies activating and repressing TFs of the gene. The significant positive and negative correlations for the gene ATP1 are highlighted in red at the left of the matrix. The correlation matrix depicts the expression correlation within TFs of the ATP1 gene. **(D)** The percentage of activators and percentage of repressors of genes in high- and low- noise processes.**(E)** Left - Correlation of noise in biological processes with the number of overlaps in TF binding sites. Right - Number of overlaps per TF binding site for genes in high and low-noise processes **(F)** The overlaps in binding sites are up to 1000bp upstream region of the start codon. **(G)** The percentage of TFs that show binding site overlaps among the TFs that show positive expression correlation (co-expressing) and negative expression correlation.

Figure 3

(A)



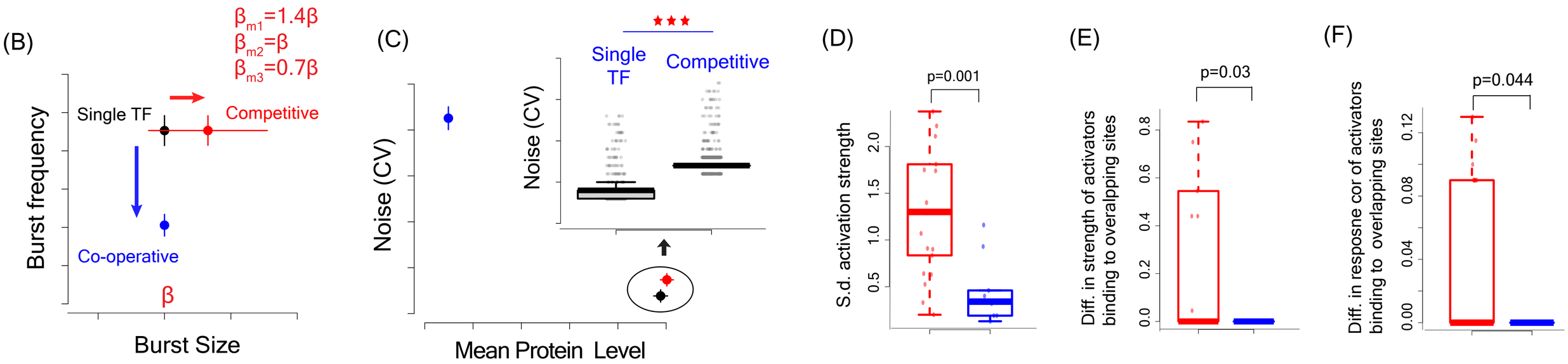




**Figure 3: Binding competition for overlapping TF binding sites drive noise. (A)** Overlap between TF binding sites lead to binding competition between TFs. Dynamics of binding and unbinding of TFs to promoter region in conjunction with variation in regulation strengths of TFs gives rise to temporal variation in TF binding in the same promoter region within a cell and leads to temporal variation in transcription rate. In addition, asynchrony in competitive TF binding among individual cells gives rise to inter-individual variation in TF binding and consequenctly, in transcription rate. The transcription rate variation profiles were obtianed from mathematical modeling and simulations. **(B)** Mathematical modeling and simulations show that co-operative TF binding changes the burst frequency but may not change the burst size, whereas competitive TF binding with TFs of varied regulation strengths alters burst size but may not affect burst frequency **(C)** Simulations show that both co-operative and competitve TF binding changes noise (CV). Inset - noise distribution in cases of gene regulation by single TF binding and competitve TF binding calculated across multiple time-points in 10,000 simulated cells. **(D)** Standard deviation in the strength of regulation of the activators for genes in high- and low- noise processes. **(E)** Difference in regulation strength of activators that bind to overlapping sites of genes in high- and low- noise processes. **(F)** Difference in response correlation of activators that bind to overlapping promoter sites.

Figure 4

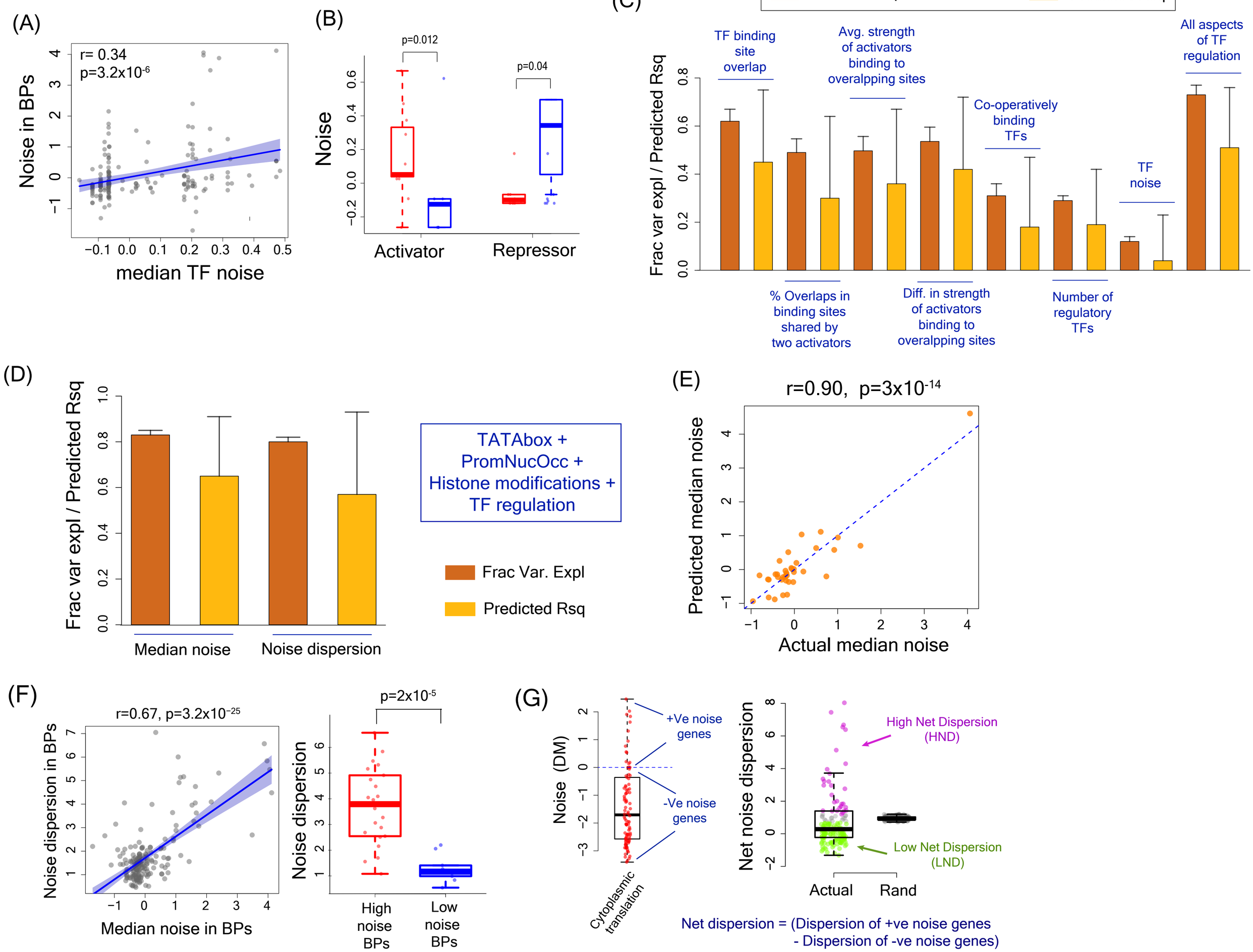



**Figure 4: Noise prediction using the integrated model and noise-balance in biological pathways. (A)** Left - Correlation between noise in biological processes with median TF noise. **(B)** Noise of activators and repressors of genes in high- and low- noise processes. **(C)** Comparison of impact of different aspects of regulation by TFs such as TF binding site overlap, Number of regulatory TFs, Co-operativity in TF binding, TF expression noise and all aspects of TF regulation on explaining and predicting noise in biological processes. **(D)** Using an ingerated model combining all factors including presence of TATA box, promoter nucleosome occupancy, histone modifications and TF binding for explaining and predicting median noise and noise dispersion. **(E)** A representative example of correlation between predicted median noise values and actual noise values. **(F)** Left - correlation between noise dispersion and median noise in biological processes. Right - Noise dispersion in processes with high and low median noise. **(G)** Net noise dispersion for a biological process is calculated as the difference between the dispersion of the positive noise genes and the dispersion of the negative noise genes. The processes in green show lower than expected net noise dispersion and were termed as low noise dispersion (LND) processes. The processes indicated in magenta showed higher than expected dispersion and were termed as high noise dispersion (HND) processes.

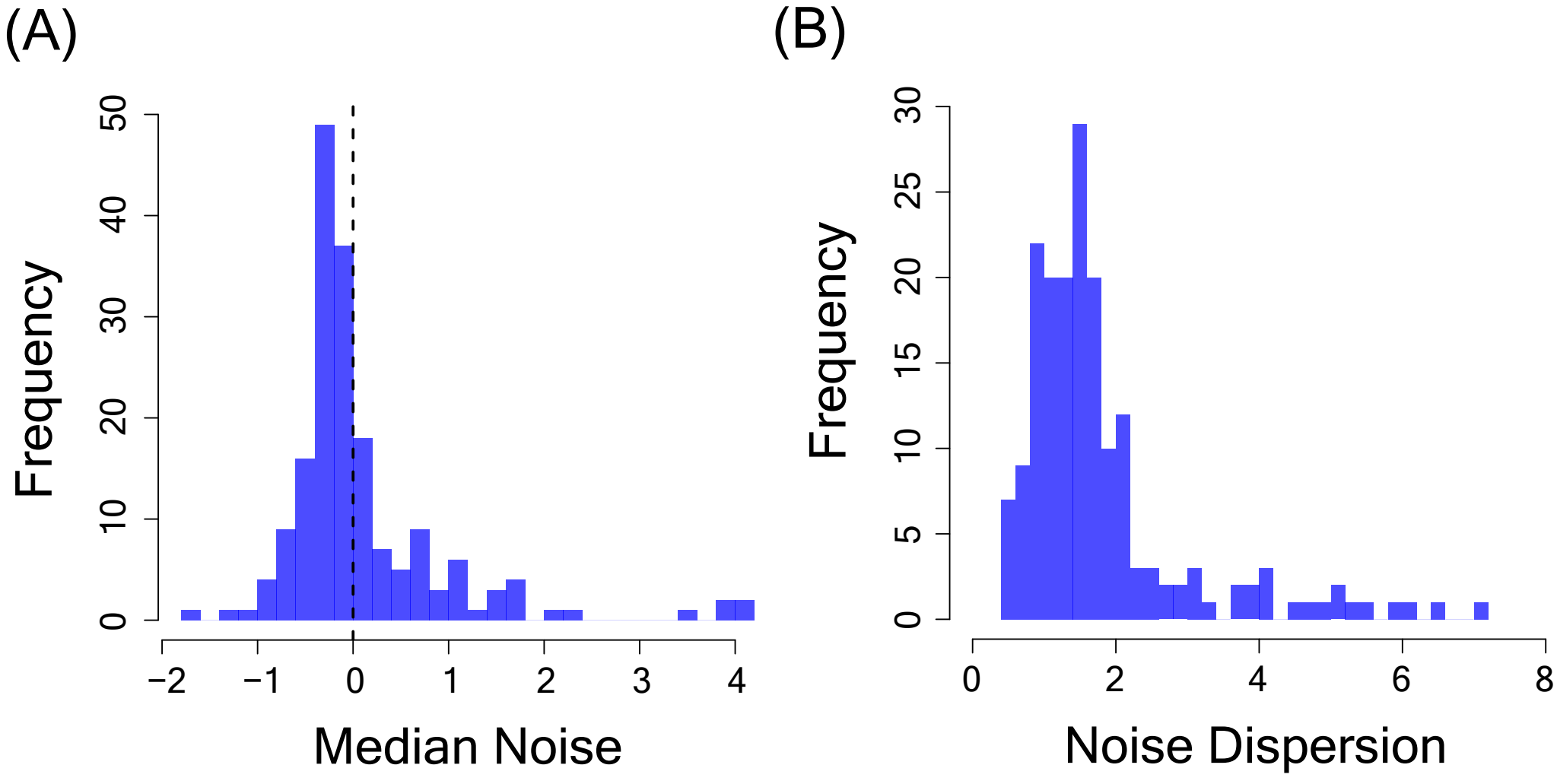



**Figure S1: Median noise and dispersion of noise across biological processes (A)** Distribution of median noise in 181 biological processes in the yeast Saccharomyces cerevisiae **(B)** Distribution of noise dispersion values, characterized by inter-quartile range, in these biological processes in yeast

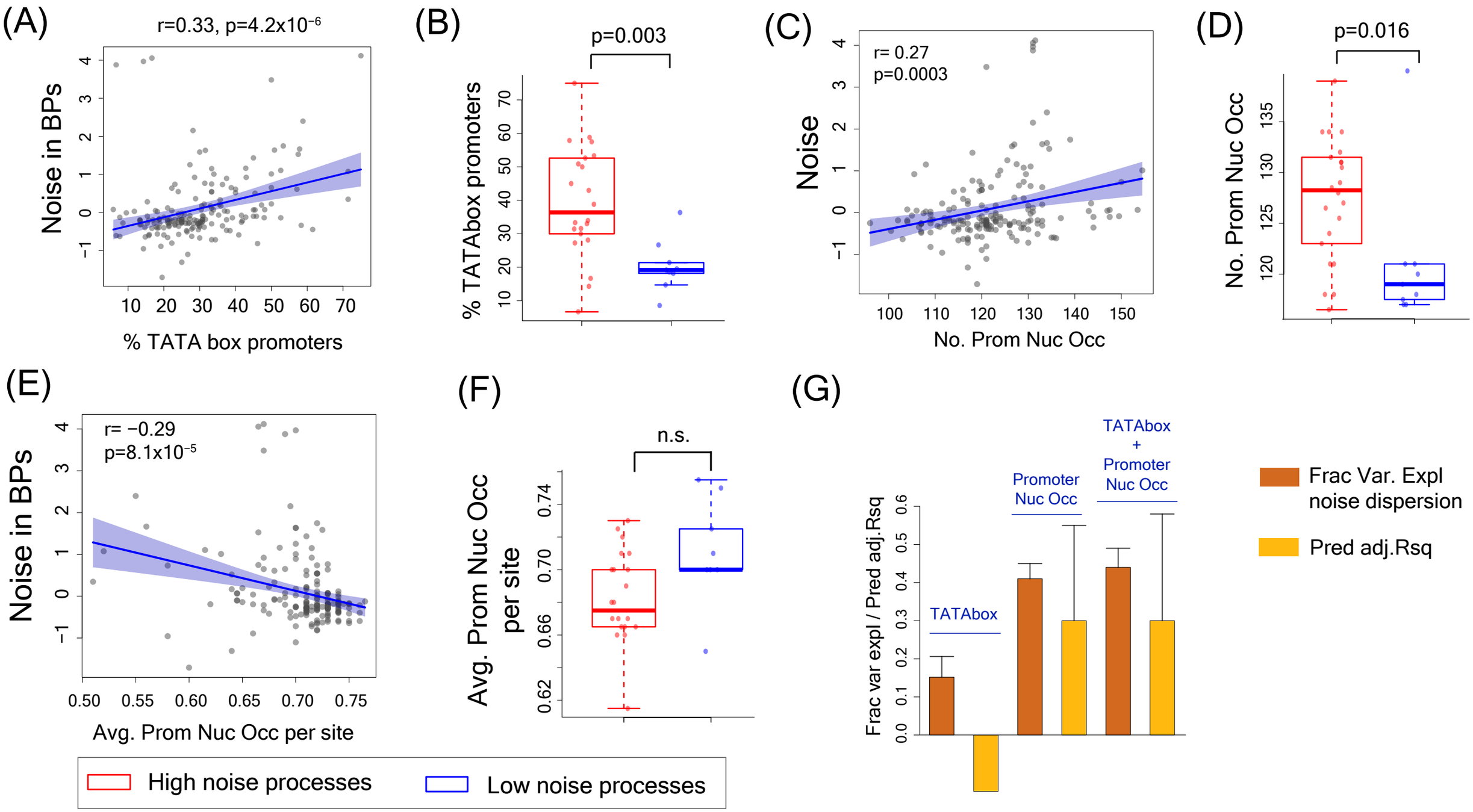



**Figure S2: Percentage of genes with TATA box in promoter sequence and promoter nucleosome occupancy correlate with noise in biological processes. (A)** Correlation between noise in biological processes and percentage of genes with TATA box sequence in their promoters.The blue region shows the region of 95% confidence interval. **(B)** Percentage of genes with TATA box sequence in the promoter in high- and low- noise processes. **(C)** Correlation between noise in biological processes and the number of nucleosome-occupied promoter sites. **(D)** Number of nucleosome occupied promoter sites of genes in high- and low- noise processes. **(E)** Correlation between noise in biological processes and average promoter nucleosome occupancy per occupied site **(F)** Average promoter nucleosome occupancy per occupied site of genes in high- and low- noise processes **(G)** Impact of presence of TATA box, Promoter nucleosome occupancy and promoter nucleosome occupancy in combination with presence of TATA box on explaining and prediction variation in noise dispersion.

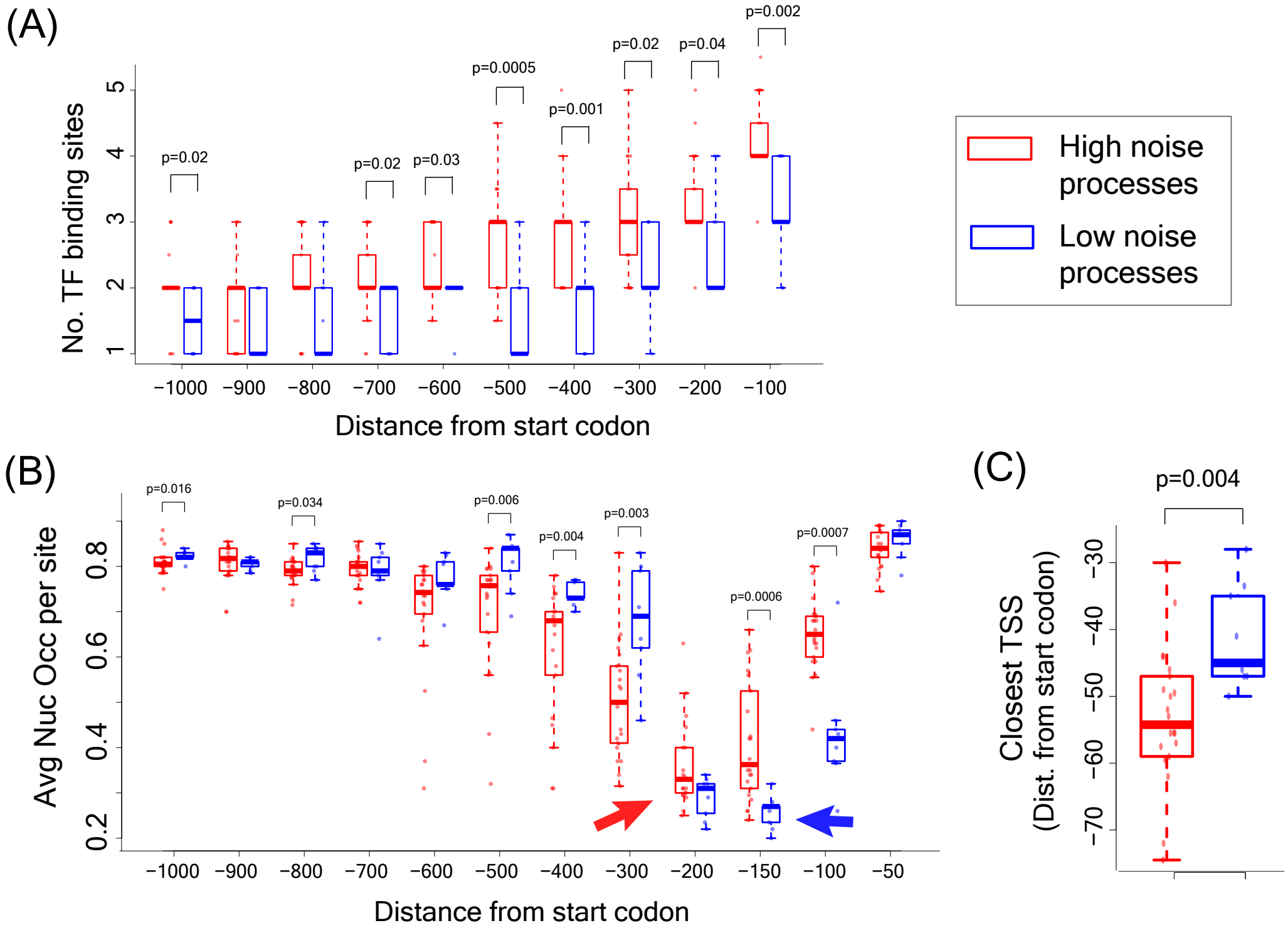



**Figure S3: TF binding sites, average promoter nucleosome occupancy per site and closest TSS position in the promoter region of genes in high- and low- noise processes. (A)** Distribution of TF binding sites up to 1000bp upstream region of genes in high- and low- noise processes. **(B)** Average promoter nucleosome occupancy per site in the promoter region for genes in high- and low- noise processes and the point of lowest occupancy (indicated by red and blue arrows) shows an upstream shift in high-noise processes. **(C)** The closest TSS shows a ~15bp upstream shift in the promoter region for genes in high-noise processes.

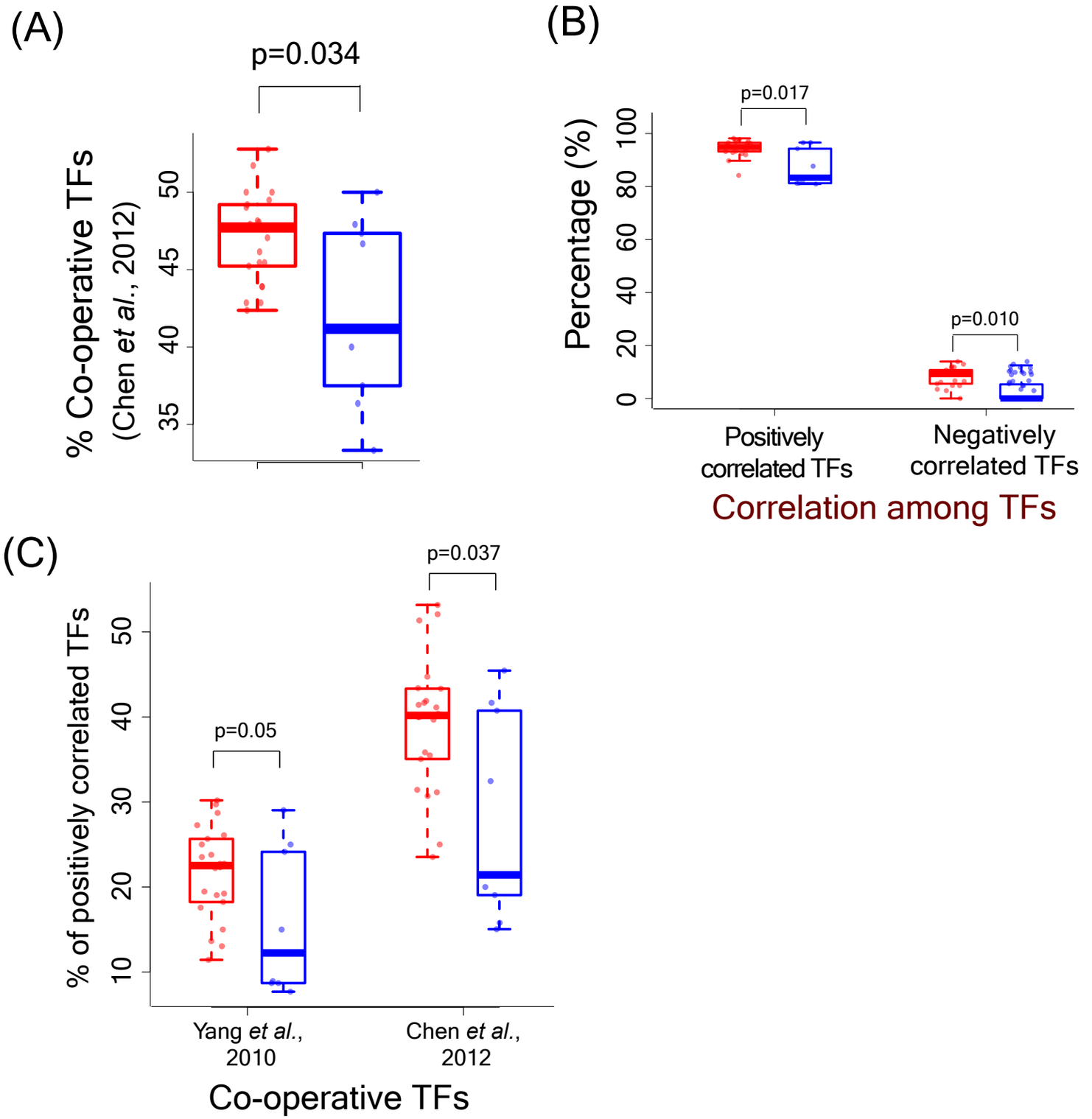



**Figure S4: Percentage of co-operative TFs and TFs with significant expression correlation among themselves. (A)** Percentage of co-operative TFs (Chen *et al.*, 2012) for genes in high- and low- noise processes. **(B)** Percentage of TFs showing positive correlation among themselves and percentage of TFs showing negative correlation among themselves **(C)** Percentage of co-operative TFs among positively correlated TFs for genes in high- and low- noise processes.

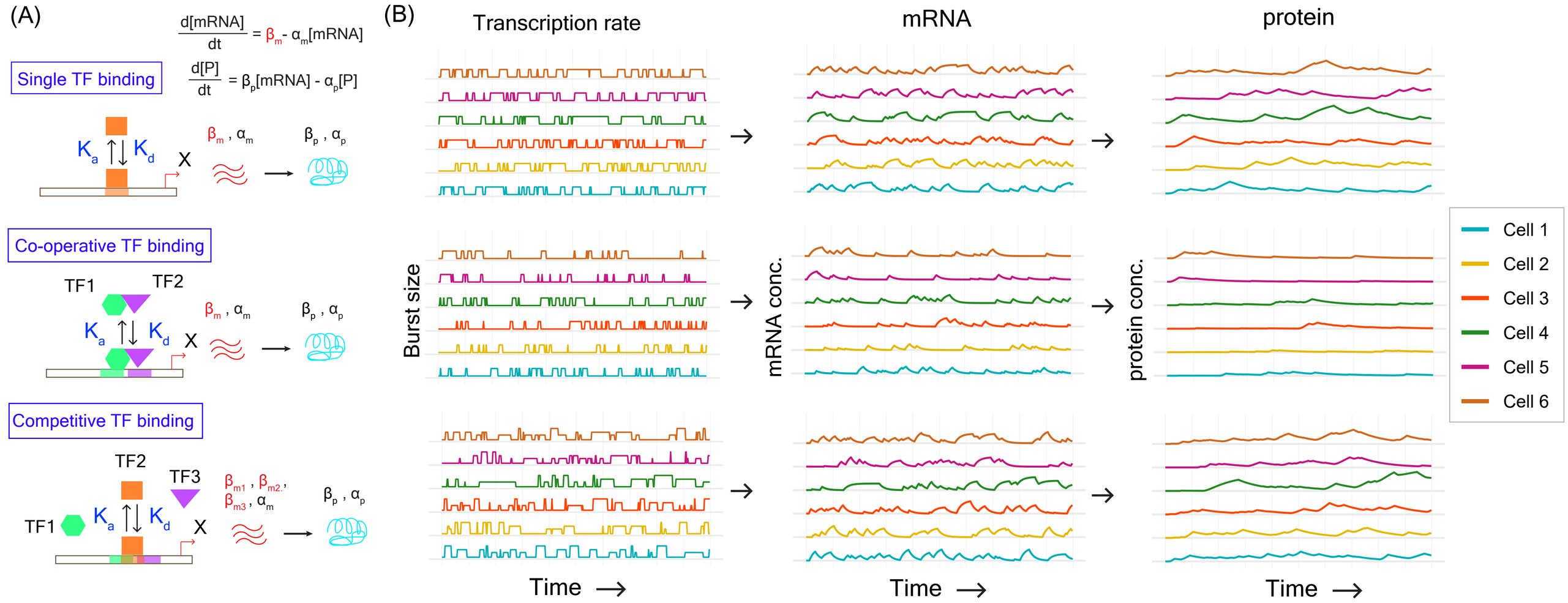



**Figure S5**: **Mathematical modeling and simulations of gene regulation by single TF, cooperative TF and competitive TF binding. (A)** Mathematical description of the model for deciphering the consequences of gene regulation by single TF binding, co-operative TF binding and competitive TF binding. **(B)** Examples of variation in transcription rate, mRNA concentration and protein concentration with time in cases of regulation by single TF biinding, co-operative TF binding and competitive TF binding in six representative cells.

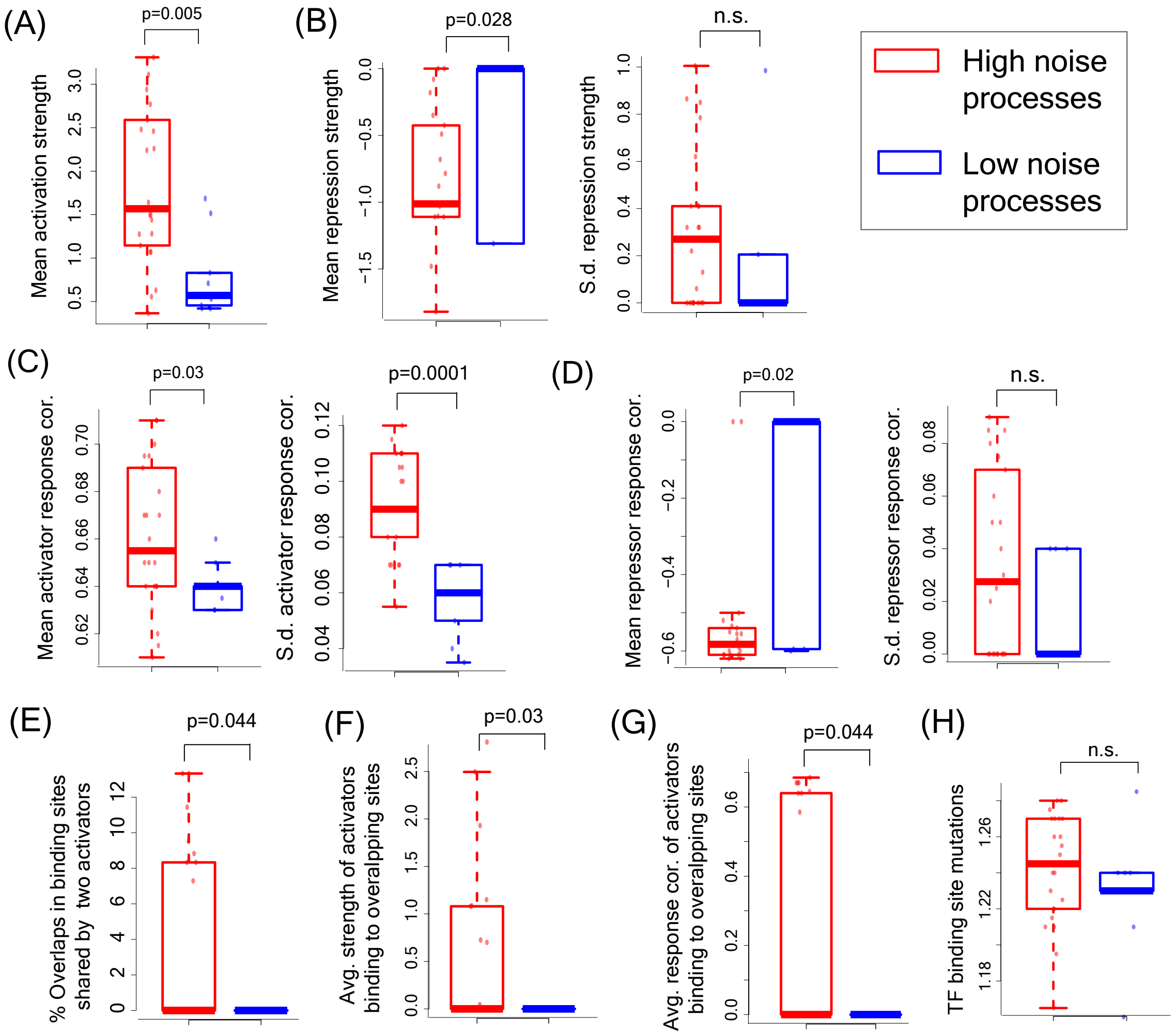



**Figure S6: Regulation strengths of TFs and response correlation between TFs and regulated genes. (A)** Mean regulation strength of activators in high and low-noise processes. **(B)** Mean and standard deviation of regulation strength of repressors **(C)** Mean and standard deviation of response correlation for activators **(D)** Mean and standard deviation of response correlation for repressors **(E)** Difference in percentage of overlapping sites bound by activators for genes in high- and low- noise processes. **(F)** Average regulation strength of activators that bind to overlapping sites. **(G)** Average response correlation of activators that bind to overlapping sites. **(H)** Average number of mutations per TF binding site in the promoter of genes in high- and low- noise processes.

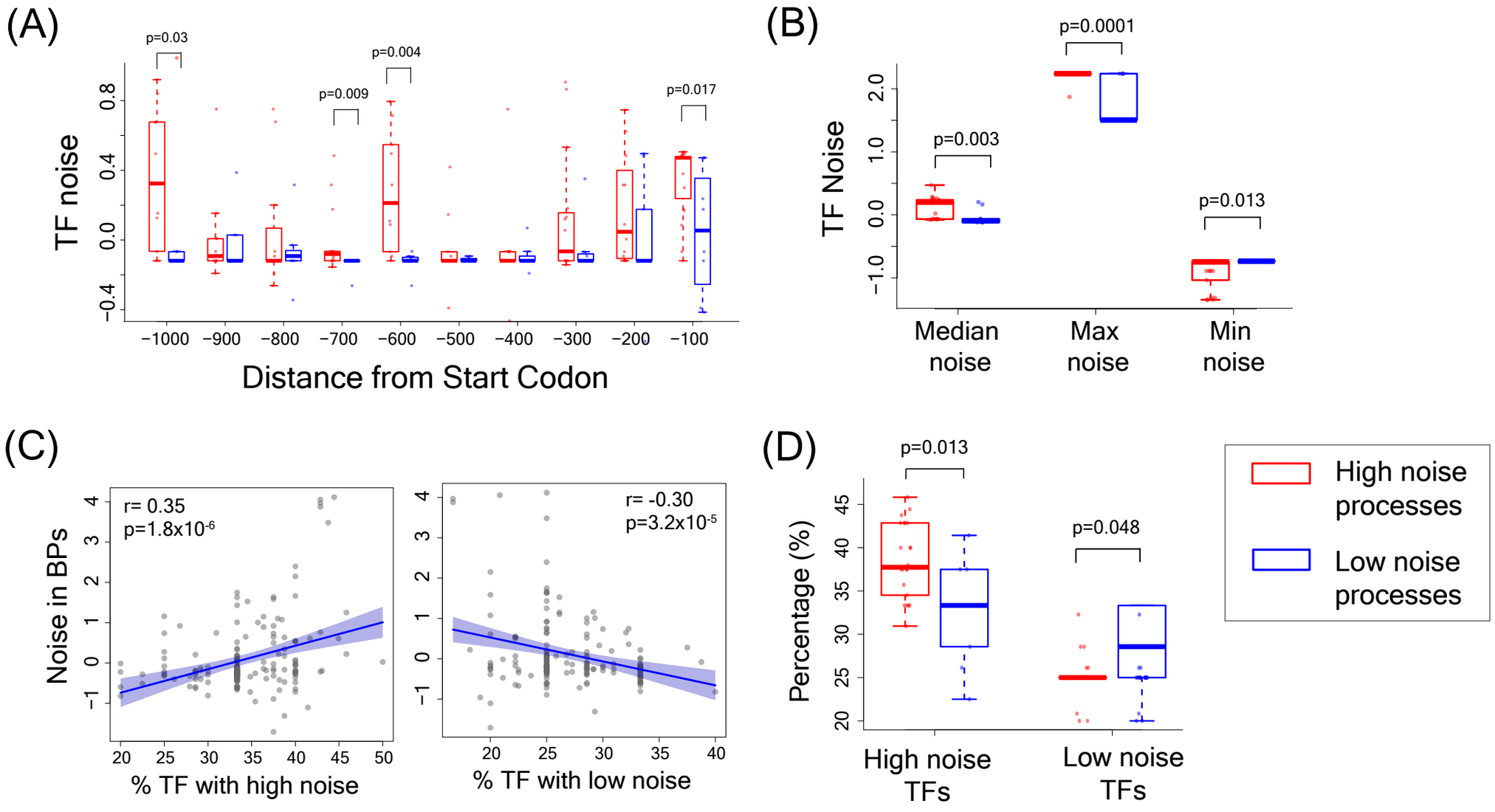



**Figure S7: Expression noise in regulatory TFs (A)** Distribution of noise of TFs that bound up to 1000 bp upstream region of genes for high-noise and low-noise processes. **(B)** Median, maximum and minimum TF noise in high- and low- noise processes. **(C)** Correlation between noise in biological processes with percentage of high- and low-noise TFs. **(D)** Percentage of high-noise and low-noise TFs of genes in high- and low- noise processes.

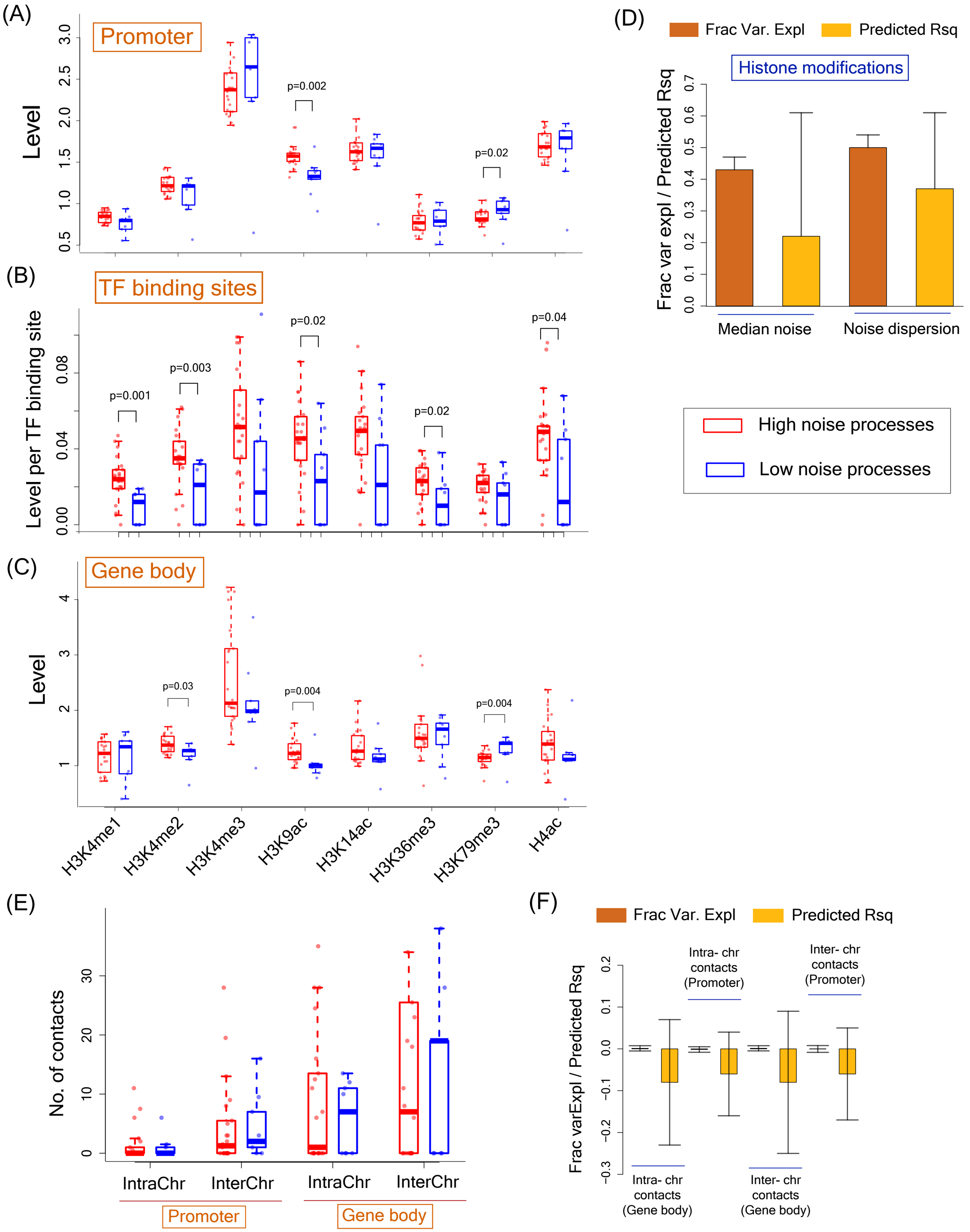



**Figure S8: Histone modification patterns and 3D genomic contacts have low power for noise prediction.** Distribution of histone modification patterns and associated proteins in the **(A)** promoter regions, **(B)** TF binding sites and **(C)** gene body for high and low-noise processes. **(D)** Fraction of variation in noise explained by histone modification pattern and the corresponding predictive power given by predicted $R^2$. **(E)** Distribution of number of intra-chromosomal and inter-chromosomal contacts in the promoter and gene-body in high- and low-noise processes. **(F)** Fraction of variation in noise explained by three dimensional genomic contacts and predictive power given by predicted $R^2$.

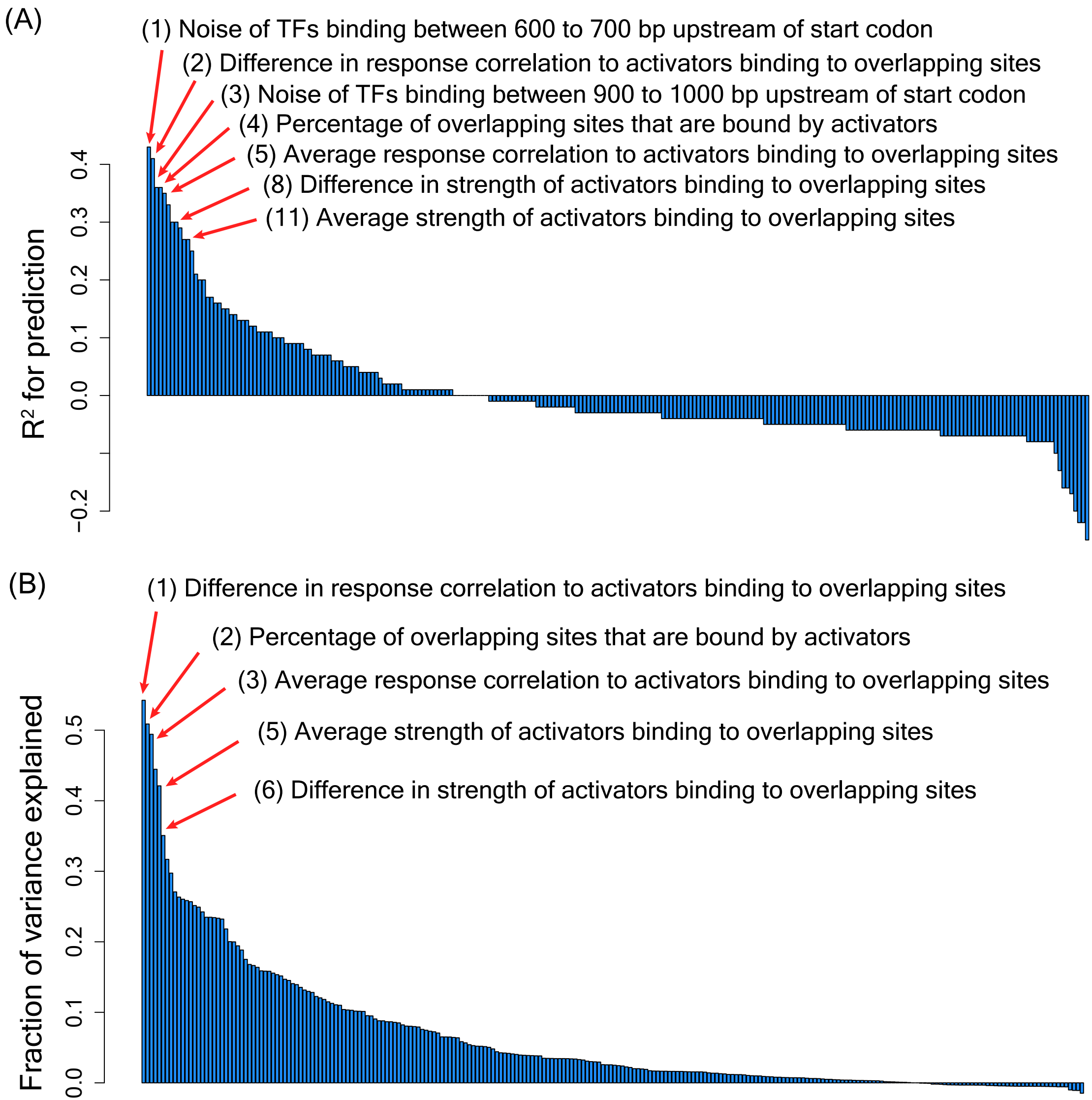



**Figure S9: Predictive power of and fraction of variance explained by individual molecular factors.** Impact of all factors individually on **(A)** predicting median noise given by predicted $R^2$ value and **(B)** fraction of variation explained

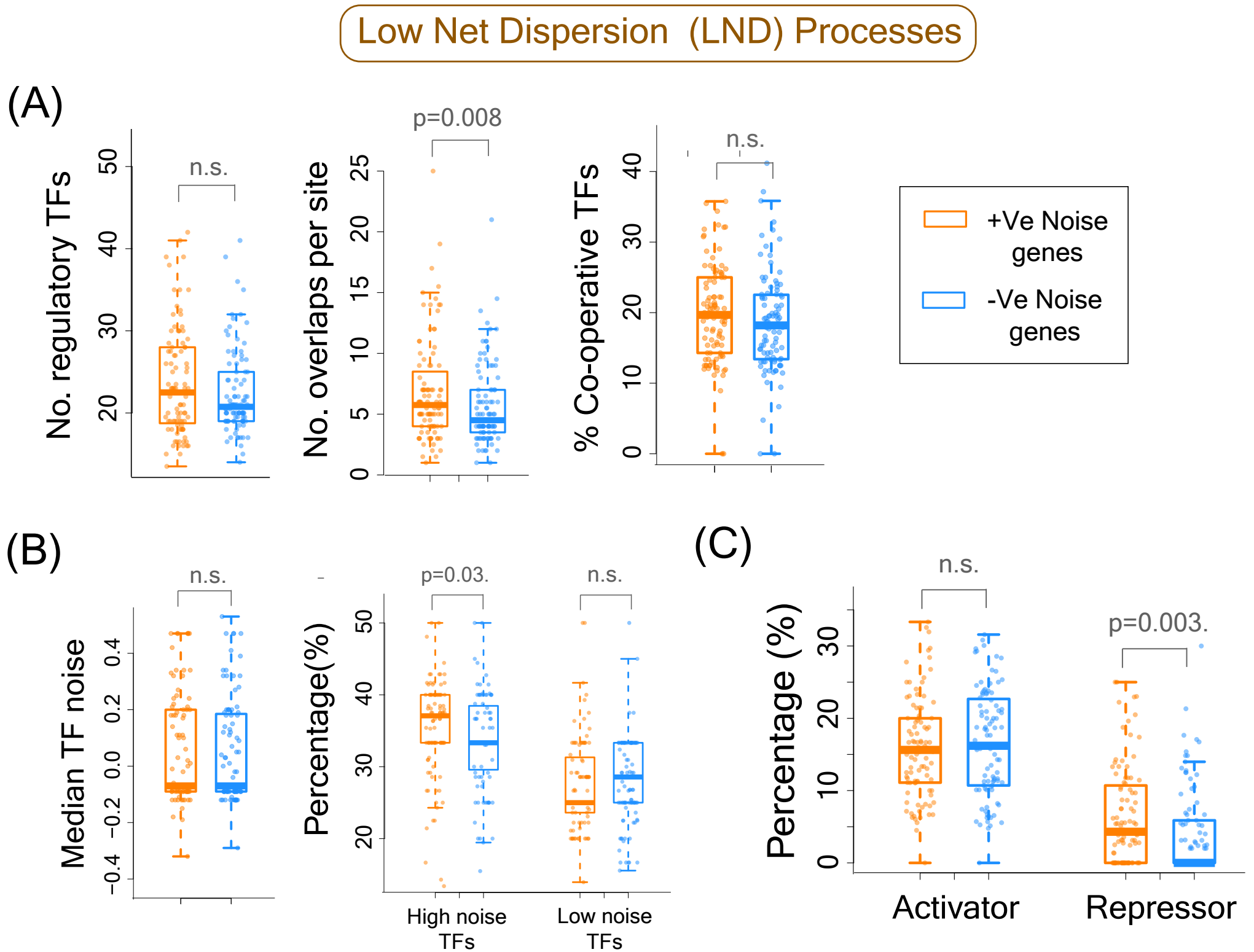



**Figure S10: Regulatory TF properties of genes underlie noise balance in biological processes. (A)** Comparison of number of regulatory TFs, number of overlaps per TF binding site and percentage of co-operative TFs between positive noise and negative noise genes in LND processes. **(B)** Median TF noise and percentage of high and low-noise TFs for positive noise and negative noise genes in LND processes. **(C)** Percentage of activators and repressors of positive noise and negative noise genes in LND processes..

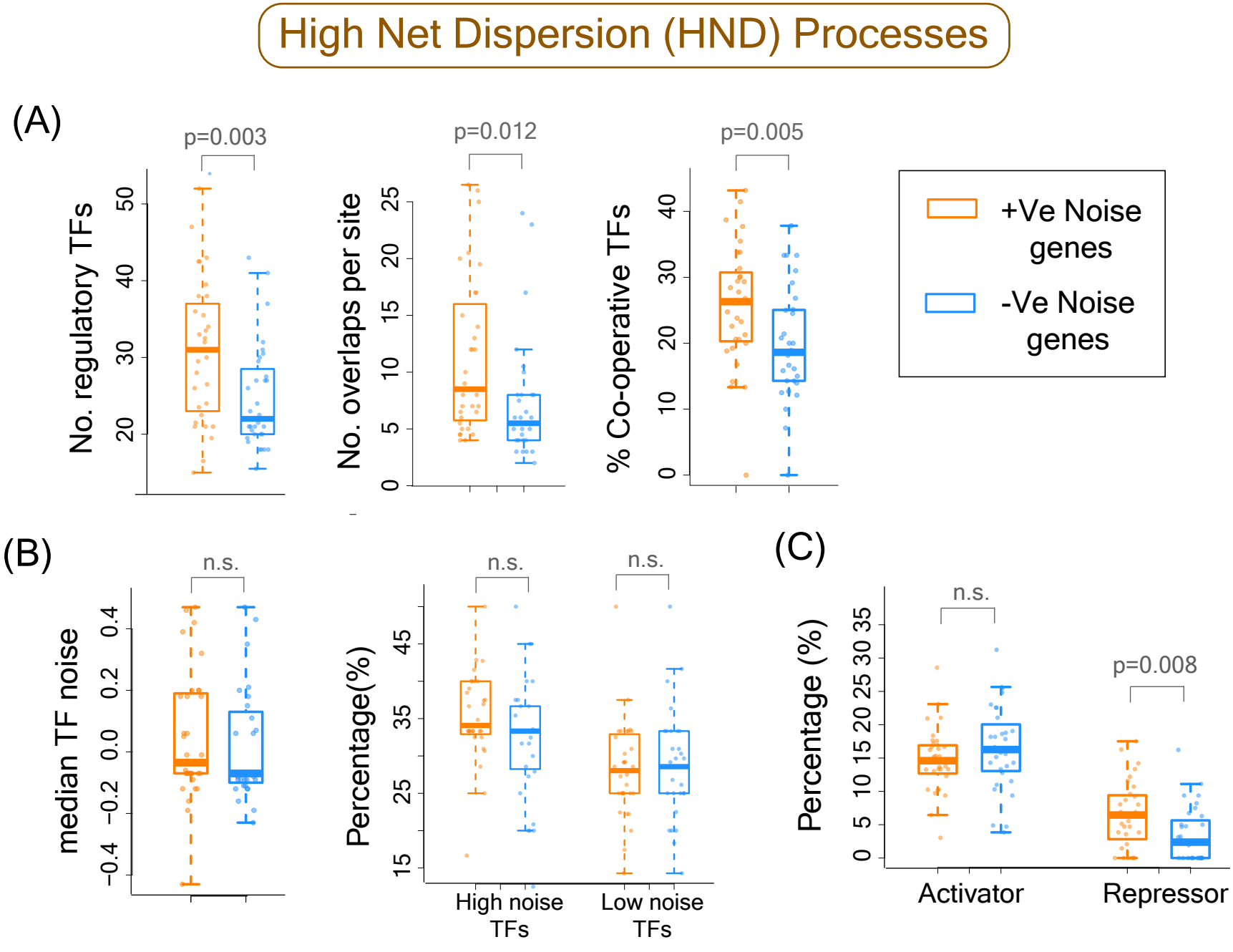



**Figure S11: Properties of regulatory TFs of positive and negative noise genes in HND processes.** Figures (A) to (C) show **(A)** Number of regulatory TFs and number of overlaps per TF site **(B)** median TF noise and percentage of high and low-noise TFs **(C)** percentages of activators and repressors of positive and negative noise genes for HND processes.

| GO_ID | Avg. Noise (DM) | Median Noise (DM) | adj. p-value | Function |
|---|---|---|---|---|
| GO:0006099 | 3.648 | 4.117 | 4.25E-03 | Tricarboxylic acid cycle |
| GO:0006754 | 5.583 | 4.056 | 9.30E-03 | ATP biosynthetic process |
| GO:0015986 | 5.499 | 3.967 | 5.89E-03 | ATP synthesis coupled proton transport |
| GO:0099132 | 4.977 | 3.878 | 8.29E-03 | ATP hydrolysis coupled cation transmembrane transport |
| GO:0006122 | 3.257 | 3.483 | 2.12E-02 | Mitochondrial electron transport, ubiquinol to cytochrome c |
| GO:0003333 | 5.034 | 2.398 | 9.30E-03 | Amino acid transmembrane transport |
| GO:0015992 | 3.935 | 2.152 | 5.91E-05 | Proton transport |
| GO:0098869 | 3.994 | 1.747 | 4.02E-03 | Cellular oxidant detoxification |
| GO:0006865 | 4.313 | 1.666 | 1.75E-02 | Amino acid transport |
| GO:0055114 | 3.029 | 1.638 | 1.22E-13 | oxidation-reduction process |
| GO:0006811 | 3.105 | 1.631 | 1.25E-05 | Ion transport |
| GO:1902600 | 2.252 | 1.586 | 1.73E-03 | Hydrogen ion transmembrane transport |
| GO:0055085 | 3.364 | 1.528 | 7.53E-06 | Transmembrane transport |
| GO:0045454 | 2.712 | 1.433 | 1.75E-02 | Cell redox homeostasis |
| GO:0009060 | 2.319 | 1.253 | 2.56E-02 | Aerobic respiration |
| GO:0005975 | 2.696 | 1.105 | 6.18E-02 | Carbohydrate metabolic process |
| GO:0034599 | 2.818 | 1.006 | 2.43E-04 | Cellular response to oxidative stress |
| GO:0008152 | 1.914 | 0.921 | 2.13E-06 | Metabolic process |
| GO:0032259 | 2.790 | 0.767 | 2.61E-02 | Methylation |
| GO:0019236 | 1.210 | 0.747 | 1.61E-02 | Response to pheromone |
| GO:0008652 | 1.369 | 0.651 | 8.50E-03 | Cellular amino acid biosynthetic process |
| GO:0006810 | 1.401 | 0.173 | 8.16E-04 | Transport |

**Table S1: List of high-noise biological processes in which genes show significantly positive noise (DM) values.**

| GO_ID | Avg. Noise (DM) | Median Noise (DM) | adj. p-value | Function |
|---|---|---|---|---|
| GO:0006368 | -0.291 | -0.279 | 9.85E-02 | Transcription elongation from RNA polymerase II promoter |
| GO:0006351 | -0.224 | -0.287 | 1.01E-02 | Transcription, DNA-templated |
| GO:0006486 | -0.306 | -0.356 | 9.93E-02 | Protein glycosylation |
| GO:0006355 | -0.252 | -0.397 | 5.41E-03 | Regulation of transcription, DNA-templated |
| GO:0032968 | -0.384 | -0.397 | 3.11E-02 | Positive regulation of transcription elongation from RNA polymerase II promoter |
| GO:0032543 | -0.299 | -0.566 | 4.05E-03 | Mitochondrial translation |
| GO:0006412 | -0.903 | -0.891 | 9.29E-12 | Translation |
| GO:0000027 | -0.972 | -1.103 | 1.21E-02 | Ribosomal large subunit assembly |
| GO:0002181 | -1.147 | -1.707 | 6.56E-08 | Cytoplasmic translation |

**Table S2: List of low-noise biological processes in which genes show significantly negative noise (DM) values.**