Multivariable regulation of gene expression plasticity in metazoans

Long Xiao[1,2,#], Zhiguang Zhao[1,2,#], Fei He[3,*], Zhuo Du[1,2,*]

[1] State Key Laboratory of Molecular Developmental Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, 100101, China

[2] University of Chinese Academy of Sciences, Beijing, 10049, China

[3] Biology Department, Brookhaven National Lab, Upton, NY, 11967, United States of America

[#] Equal contribution

* To whom correspondence should be addressed

Zhuo Du: Tel: 86-10-64801699; Fax: N/A; Email: zdu@genetics.ac.cn

Fei He: Tel: 01-347-265-4749; Fax: N/A; Email: feihe@ksu.edu

Present Address:

Fei He, Kansas State University, Manhattan, KS, 66506, United States of America

ABSTRACT

An important capacity of genes is the rapid change of expression levels to cope with environment, known as expression plasticity. Elucidating the genomic mechanisms determining expression plasticity is critical for understanding the molecular basis of phenotypic plasticity, fitness, and adaptation. In this study, we systematically quantified genome-wide gene expression plasticity in four metazoan species by integrating changes of expression levels under a large number of genetic and environmental conditions. From this, we demonstrated that expression plasticity measures a distinct feature of gene expression that is orthogonal to other well-studies features including gene expression potential and tissue specificity/broadness. Expression plasticity is conserved across species with important physiological implications. The magnitude of expression plasticity is highly correlated with gene function and genes with high plasticity are implicated in disease susceptibility. Genome-wide analysis identified many conserved promoter *cis*-elements, *trans*-acting factors (such as CFCF), and gene body histone modifications (H3K36me3, H3K79me2, and H4K20me1) that are significantly associated with expression plasticity. Analysis of expression changes in perturbation experiments further validated a causal role of specific transcription factors and histone modifications. Collectively, this work reveals general properties, physiological implications, and multivariable regulation of gene expression plasticity in metazoans, extending the mechanistic understanding of gene regulation.

INTRODUCTION

Gene expression connects genotypes to phenotypes. Gene expression plasticity, which concerns the capacity of genes to change their expression levels under diverse conditions, is critical for phenotypic plasticity, adaptation, and evolvability (1-4). It has been widely observed that the expression levels of certain genes (such as stress response genes) are intrinsically more flexible while those of other genes are more resistant (such as house-keeping genes). Gene expression plasticity has important implications for organismal fitness. Generally, low plasticity confers stability and allows the organism to maintain a steady state, while high plasticity allows an organism to rapidly remodel its gene expression programs

to cope with changing environments, enabling phenotypic adaption (2). Elucidating the genomic mechanisms underlying differential gene expression plasticity is an important but unresolved question in genome biology.

Pioneering studies in yeast have defined gene expression plasticity by measuring the magnitude of genome-wide gene expression changes across diverse genetic and environmental conditions (3,5,6). Functional analyses have revealed that two types of genetic and epigenetic signatures in promoters correlate with expression plasticity. The first signature is the TATA box, a conserved element present in many eukaryotic gene promoters. In yeast, TATA box-containing genes exhibit significantly higher levels of plasticity than TATA-less genes, and this phenomenon is conserved in other species (3). The second signature is nucleosome occupancy and organization near the transcription start site (TSS). The presence of well-positioned nucleosomes is associated with significantly higher expression plasticity in yeast (5,6).

In multicellular organisms, gene expression plasticity and its regulation are less well-understood. Gene expression changes have been profiled in response to a limited number of conditions in individual experiments in *C. elegans* (7-9)*, Drosophila* (10), and in human cells (11). The presence of specific transcription regulatory elements has been shown to allow dozens of genes to be co-regulated in response to heat shock stress (7). Systematic comparison of gene expression changes in five *C. elegans* strains cultured under five conditions identified that certain genes are more prone to response to *trans*-acting factors to mediate genotype–environment interactions, especially those with complex promoter architecture and mid-range expression levels (8). In addition, the genetical genomics approach has been employed to identify *cis*- and *trans*- loci that are strongly associated with gene expression changes in response to specific environmental conditions such as ambient temperature (9). These studies have characterized groups of genes that are prone to expression change under specific stress conditions, and revealed certain gene features that mediate plastic gene expression. However, many important questions remain to be addressed. For example, is gene expression plasticity an intrinsic gene property conserved in multicellular organisms? What are the biological implications of differential expression plasticity? Furthermore, given the significantly increased complexity of gene regulation in metazoan species, are

there additional *cis*-elements, *trans*-factors, and epigenetic regulators that underlie expression plasticity? To this end, a systematic analysis of the properties and genomic regulation of gene expression plasticity in metazoan species is highly desirable, but has yet to be performed.

In this study, we performed an integrative functional analysis of genome-wide gene expression programs, gene attributes, *cis*-regulatory motifs, *trans*-acting proteins, and histone modifications to decipher the properties, implications, and regulation of gene expression plasticity in four metazoan species. Our results revealed that gene expression plasticity is a conserved gene property implicated in cellular flexibility, disease susceptibility, and stress/environmental response. We provide genome-wide evidence that core promoter cis-elements, transcription factors (such as CTCF), and gene body histone modifications (H3K36me3, H3K79me2, and H4K20me1) play a causal role in determining expression plasticity. Our findings provide insights into the function and regulation of gene expression plasticity in metazoans.

RESULTS

Quantification of gene expression plasticity in metazoan species

Gene expression plasticity (GEP) is defined as the magnitude of gene expression change across diverse genetic and environmental conditions (**Figure 1A**). As done previously in yeast (3), we used gene expression datasets from a public database (12) or from the literature (13) to quantify GEP in four metazoan species (**Table S1**). For each condition (e.g. culture temperature, gene mutation, drug treatment), the magnitude of gene expression change after treatment was determined as a fold change, and expression changes across all conditions (range from >200 to >1,000 in different species) were averaged to represent GEP (**Figure S1A** and **Figure 1B**).

We performed a series of quality control checks to ensure reliable quantification of GEP. First, we verified that the total number of conditions used was sufficient to represent GEP. Simulation of the influence of condition number on GEP showed high stability of the score ($r = 0.9$) once the condition

number reached 100 (**Figure S1B**); as we used over 100 conditions for all species, our condition number was sufficient. Second, unlike yeast in which GEP is quantified in one cell across various conditions, GEP in multicellular organisms combines the results of different tissues and cell types. Thus, the quantifying organismal-level GEP may not be meaningful if cell-specific GEP is pervasive. To assess this possibility, we calculated GEP for specific human cell lines and found that cell-type-specific GEP values correlated significantly with total GEP and with each other (**Figure 1C**), suggesting that while GEP exhibits cell specificity, the total GEP recapitulates that of specific cell types. Third, we confirmed the biological relevance of quantified GEP using benchmark genes. As expected, human signaling-responsive genes (14) whose expression is expected to be conditional and dynamic showed significantly high GEP as compared to other genes (**Figure S1C**). In addition, consistent with the expectation that genes participating in stress response should have high GEP, our data showed that in both fly and worm (10,15), stress responsive genes and those required for stress response indeed exhibited significantly higher GEP (**Figure S1D**). Finally, our measurement of GEP is based on direct comparison of gene expression changes before and after treatments/perturbations (challenging conditions) which would in theory more recapitulate the potential of a gene to change its expression as compared to the natural variability of gene expression across tissue or cell types under normal condition. We applied the same method and calculate GEP using gene expression data across a large collection of normal samples (16). While GEP calculated using both data sources correlated moderately with each other (**Figure 1D**), the data source used here better represented GEP than that of gene expression across samples under normal conditions (**Figure 1E**). GEP of signaling-responsive genes calculated using challenging conditions was significantly higher than that using only normal conditions (Wilcoxon signed-rank test, $p < 0.001$).

Expression plasticity is an intrinsic gene property associated with gene function and disease susceptibility

With GEP quantified, we next determined whether GEP is an intrinsic gene property based on two criteria. First, a gene property should be evolutionarily conserved. Consistently, we found GEP to be

widely conserved between species. The strongest correlation coefficient was observed between human and mouse orthologs ($\rho = 0.46$, $p < 0.001$). Weaker but statistically significant correlations were also observed between distantly related species (**Figure S2A**). Second, a gene property should measure a distinct gene feature. Gene expression potential (average gene expression levels across samples) and broadness (also known as tissue specificity) are two important properties that quantify the abundancy and breadth of gene expression (16,17). While intuition suggests that GEP could be related to expression potential and broadness, our analysis showed that GEP to be poorly correlated with expression level (**Figure S2B**, $\rho = 0.079$ and 0.179 respectively for two different datasets) and expression broadness (**Figure S2C**, $\rho = 0.077$ and 0.080 respectively for two different measurements of expression broadness), suggesting GEP indicates an orthogonal feature of gene expression. Together, the above results confirm that expression plasticity is an intrinsic gene property indicating the changeability of gene expression.

We next sought to determine the potential implications of GEP. First, we established that GEP is significantly associated with specific gene functions. Consistent with intuition, analysis of gene functional annotations (18) revealed that genes with high GEP were significantly overrepresented in biological processes that are important for cellular flexibility, such as inflammatory response, immune response, and response to drug (**Figure 2A**). In addition to global functional classifications, we compiled lists of genes with well-defined physiological functions and confirmed again that GEP and biological function were nicely consistent. Specifically, homeobox genes, which are critical for specifying cell fate and body plan, exhibited significantly lower GEP (**Figure 2B**). So did hormones and their receptors, whose function is crucial for growth and development in a dosage-sensitive manner (**Figure 2C**). Conversely, the GEP of innate immune genes was significantly higher than that of other genes (**Figure 2D**), consistent with their important roles in immune response. Importantly, the association of gene function with characteristic GEP level was highly conserved between species; the average GEP values of genes in given ontology terms were significantly correlated for all pairwise species comparisons (**Figure 2E**).

Second, we revealed that genes with high GEP are implicated in disease susceptibility. Considering that genes with high GEP would confer flexibility, we tested whether their malfunction would be more

frequently implicated in disease. We found that the GEP values of human genes were positively correlated with their propensity for disease association. The average number of diseases associated with a gene increased with expression plasticity (**Figure 2F**). Similarly, the chances of a gene being cancer-related positively correlated with GEP (**Figure 2G**); the frequency of cancer-related genes increased from 1-2% in genes with low plasticity to 4-6% in those with high plasticity.

Altogether, we have systematically quantified gene expression plasticity in four metazoan species and demonstrated it to be an intrinsic gene property with important biological implications. Low-plasticity genes tend to function in cellular processes demanding high stability, whereas high-plasticity genes are enriched in cellular processes demanding high flexibility. In particular, genes with high expression plasticity tend to be crucial for maintaining organismal fitness, especially under challenges such as disease conditions. The broad conservation and important physiological implications of expression plasticity indicate its magnitude is a specific trait under tight regulation. In the following sections, we investigate the genomic regulation of expression plasticity.

Influence of core promoter *cis*-elements on gene expression plasticity

We began by investigating the influence of promoter elements on GEP. It has been reported that yeast genes with a TATA box motif exhibit significantly higher GEP compared to those without it (3,19). This trend was nicely recapitulated in our data. As shown in **Figure S3A**, the frequency of TATA box-containing promoters increased in genes with high GEP.

We then examined whether there exist additional promoter elements influencing plasticity. To facilitate cross-species comparison, we used a large collection of well-characterized DNA motifs sourced from multiple species (20). This approach identified 141 GEP-associated motifs (**Table S2**), including a TATA box-like motif (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$). An observed lack of GEP-associated motifs in *Drosophila* might have been due to underrepresentation of that species in the motif dataset. Because many motifs exhibited very similar sequence compositions, we further collapsed similar motifs into a single motif class. This yielded 40 unique classes, 33 of which

exhibited consistent effects on GEP (**Table S2**). Respectively, 19 and 14 types of motifs promoted or repressed GEP, and nine classes showed conserved roles in two or more species (**Figure 3A**). As the DNA elements in promoters might not be independent in their presence and function, we also determined the contribution of each of the 33 GEP-associated motif classes after controlling for the influence of all other classes. To do so, we performed stepwise model selections by calculating the Akaike information criterion (AIC). The stepwise AIC procedure iteratively adds and removes model predictors (motif classes) in order to identify the subset with the best performance (lowest AIC score). Only one motif class (C35) was removed from the model, indicating that most motif classes influence GEP after controlling for the effects of others (**Figure S3B**). Consistently, the effects of motif classes that influenced GEP in the same direction were accumulative; the magnitude of GEP change from a model increased progressively as the number of same-direction motif classes increased (**Figure 3B** and **Figure S3C**). These results suggest that many promoter *cis*-elements in addition to the TATA box influence GEP non-redundantly.

One promoter may contain both GEP-promoting and -repressing elements; is the presence of motif classes with opposite effects in the core promoter optimized to coordinate their influence on GEP? If so, we predict the co-occurrence of motif classes that influence GEP in the same direction would be enriched (higher than expected), whereas the co-occurrence of those working in opposite directions would be depleted (lower than expected). Indeed, for all pairwise motif classes, the ratio of enriched to depleted was significantly higher for those influencing GEP in the same direction as compared to those with opposite directions of effect (**Figure 3C** and **Figure S3D).** Hence, *cis*-element promoter architecture is optimized to maximize its effect on GEP.

To further characterize the interactions between DNA elements, we analyzed the combinatorial effects between pairwise motif classes to identify three types of interactions, namely additivity, enhancement, and dominance (**Figure S3E**). Most pairwise motif classes (88%, 87%, and 91% in human, mouse, and worm respectively) function additively to influence GEP. The observed GEP values of promoters containing both motif classes were similar to that expected from summing the effects of individual motif classes (**Figure 3D** and **Figure S3F**). Interestingly, a small number of motifs exhibited

non-additive interactions. In human (**Figure 3D**), we identified seven pairs of motif classes that demonstrated enhancement interactions in which the observed effects in promoters with both motif classes were significantly stronger than expected. In addition, seven pairs of motif classes exhibited dominance interactions in which the effect of one motif class was masked by that of the other.

The above analysis of the contributions of promoter *cis*-elements to expression plasticity reveals that many promoter *cis*-elements function individually or synthetically to influence expression plasticity. A DNA element often regulates gene expression through the binding of *trans*-acting proteins, such as transcription factors. However, our knowledge of the association between any given *cis*-element and its corresponding regulatory proteins is fairly limited. This uncertainty prevented us from performing a systematic analysis of how motifs regulate expression plasticity via regulatory proteins. To circumvent this complexity, we next adopted a regulatory protein-centric strategy in order to investigate whether certain *trans*-acting proteins regulate expression plasticity.

Promoter binding of specific *trans*-acting proteins is associated with gene expression plasticity

The availability of genome-wide binding patterns for many regulatory proteins provides rich opportunities to elucidate their functions in GEP. In particular, the Encyclopedia of DNA Elements (ENCODE) (21) has mapped *in vivo* binding regions for a large collection of human regulatory proteins in many samples, allowing us to systematically evaluate the potential functions of *trans*-acting proteins. If a regulatory protein influences GEP, its target genes would exhibit significantly higher or lower values compared to those it does not bind. Notably, where GEP describes a constant feature of gene expression, the occupancy of regulatory proteins at target genes is highly context-specific. We therefore required that the protein-GEP association be consistently detected in a majority of samples.

Through screening genome-wide binding data for 159 regulatory proteins in 505 samples (**Table S3**), we identified six (CEBPB, CTCF, RAD21, RELA, TBP, and TCF7L2) and four (NR2C2, GABPA, SIX5, and ZNF143) proteins that were positively and negatively associated with GEP, respectively

(**Figure 4**). Notably, the list of GEP-promoting regulators includes TBP, the TATA box binding protein, consistent with the observation that TATA box-containing promoters exhibit high GEP (3). We also found that co-bind regulatory proteins exhibited consistent effects on GEP. For example, CTCF, the CCCTC-binding factor and its interacting protein RAD21, the Scc1 component of the cohesin complex (22), were both associated with higher GEP (**Figure 4A**). Significantly, some of the identified proteins had effects consistent with the motif analysis results. For example, the effects of CEBPB and GABPA were captured by both motif-centric and regulatory protein-centric analyses (**Table S2** and **Figure 4**). Similarly, the motif MA0088.2 related to ZNF143 was associated with a lower GEP (adjusted $p = 0.011$), which, while not passing the cut-off for significance (adjusted $p < 0.01$), was consistent with the results of protein occupancy analysis.

Validating the Role of *trans*-acting Proteins

To test whether the above results were simple correlations or reflected causal regulatory relationships, we used gene expression datasets from knockout experiments to validate the causal roles of transcription factors. After careful curation, we identified five expression datasets for *CTCF*, *RAD21*, *RELA* (n=2), and *TCF7L2* (23-27) that met the following criteria: first, the regulatory protein was mutated (null mutants) or knocked out; and second, genome-wide gene expression was assayed for at least two different conditions for both control (*CTR*) and loss of function mutants (*LOF*).

Because all four proteins were predicted to be GEP-promoting (**Figure 4A**), we expected that their inactivation would induce a significant reduction of GEP. For each dataset (**Figure 5A**), we compared the magnitude of expression changes for very dynamically expressed genes (top 20% of genes with highest expression changes) as an approximation of GEP for a condition. We found a significant reduction (Mann–Whitney U test, $p < 0.001$ for all cases) in the magnitude of expression changes for all four regulatory proteins (4 out of 5 datasets) (**Figures 5B-E**). Of the two datasets available for *RELA*, one was not consistent with our prediction (data not shown).

Furthermore, we compared the magnitudes of expression change among genes that were dynamically expressed in both genotypes and showed the same directions of expression change (co-upregulation or co-downregulation). Again, we found that genes generally exhibited lower magnitude expression changes following the inactivation of regulatory proteins (Wilcoxon signed-rank test, $p < 0.001$ for all cases). The ratio of genes showing lower magnitude to those showing higher magnitude ranged from 1.4 to 3, which was significantly greater than expected (**Figures 5B-E**, Fisher's exact test, $p < 0.001$ for all cases). It is worth noting that while the predicted roles of these proteins was made using human data, the perturbation experiments performed using other organisms nicely validated their roles. Together, the above results reveal a previously unrecognized function for certain sequence-specific regulatory proteins in controlling expression plasticity.

Of these regulatory proteins, CTCF and RAD21 are of particular interest. Recent findings have revealed that CTCF and the associated cohesin complex could mediate higher-order chromosome folding and chromatin interactions (28). The binding of CTCF and cohesin at genomic sites defines the physical contact points (anchor) for the formation of a chromatin loop structure that is important in transcriptional regulation (29). It was reported that the CTCF/cohesin-bound anchor regions exhibit significantly higher transcriptional activity as compared to those located in the loop region (30). As related above, both genome-wide prediction and perturbation data support a positive role for CTCF and RAD21 in regulating expression plasticity (**Figures 4A**, **5B**, and **5C**). Another subunit of cohesion, SMC3, did not pass our stringent cut-off to identify regulators of GEP, but was also associated with higher expression plasticity in three out of four examined samples (**Table S3**). These results inspired us to test whether CTCF and the cohesin complex regulate GEP through a chromatin topology-based mechanism (**Figure S4A**). If it does, we could make two predictions: first, genes whose promoters are located in the anchor regions of a chromatin loop would exhibit significantly higher GEP than those located in loop regions; and second, CTCF binding sites outside the loop structure would not be associated with GEP. We tested this possibility using the GM12878 B-lymphocyte cell line, in which genome-wide binding data for CTCF, RAD21 and SMC3 and chromatin topology data are both available (21,30). Our results excluded the possibility of topology-based GEP regulation through the

following observations. First, genes exhibited similar GEP values regardless of whether their promoters were located in anchor or loop regions (Wilcoxon Signed-Rank test, $p > 0.01$) (**Figure S4B**). Second, binding of CTCF outside the loop structure remained significantly associated with higher GEP (**Figure S4C**, Mann–Whitney U test, $p < 0.001$). These findings suggest a different mechanism accounts for the GEP-promoting role of CTCF. Indeed, CTCF is known to be a multi-functional protein and is implicated in gene regulation through diverse mechanisms (31).

Gene body H3K36me3, H3K79me2, and H4K20me1 modifications are associated with restricted expression plasticity

In addition to DNA sequence and transcription factors, epigenetic modifications also play prominent roles in regulating gene expression. We collected diverse epigenomic datasets and examined their relationships with GEP. For the identification of GEP-associated epigenetic signatures, a similar strategy was adopted as for GEP-associated regulatory proteins. We did not find evidence that DNA methylation and histone modifications in the promoter regions (2 kb genomic regions centered on the TSS) correlated significantly with GEP. Importantly, we identified three histone modifications, mainly occurring in gene body regions (**Figure S5A**), that were strongly correlated with more restricted GEP: trimethylation of histone H3 at lysine 36 (H3K36me3), dimethylation of histone H3 at lysine 79 (H3K79me2) and monomethylation of histone H4 at lysine 20 (H4K20me1). Genes enriched for these modifications in the gene body exhibited significantly lower GEP than those depleted (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$) in a majority (≥80%) of the examined human samples (**Figure 6A**). Genes with high GEP values exhibited lower frequencies of these histone modifications in the gene body (**Figure S5B** for representative examples). Importantly, the repressive effects of H3K36me3, H3K79me2 and H4K20me1 on GEP were consistently detected in all four metazoan organisms (**Figures 6B-D**). Collectively, these results predict that gene body H3K36me3, H3K79me2, and H4K20me1 modifications function to restrict expression plasticity.

Validating the role of histone modifications

To determine a causal role for the above histone modifications, we analyzed the change of GEP in mutants in which histone modifications levels were perturbed. As for regulatory proteins, we mined the literature and collected four relevant datasets (**Figure 7A**) within which global gene expression was assayed for multiple conditions in both *wild type* and in mutants of the corresponding enzyme that adds or removes a modification of interest. We expected that the loss or gain of these histone modifications would induce a respective reduction or increase in GEP.

For H3K36me3, we used a *C. elegans* dataset in which the *met-1* gene that encodes the H3K36 methyltransferase was mutated (32). MET-1 is required for maintaining the global H3K36me3 level (33), and in *met-1* mutants the modification level is reduced by over 90% (32). We compared the magnitudes of global gene expression changes (day 2 versus day 12) between control and *met-1* mutants to approximate GEP. To specifically analyze the contribution of H3K36me3, we focused only on those genes enriched in H3K36me3 in the gene body regions of control samples at both Day 2 and Day 12. As predicted by our computational analysis (**Figure 6**), the magnitude of gene expression change was significantly increased following the loss of H3K36me3 (**Figure 7B**). We also compared the magnitudes of expression change for genes that showed identical directions of change (co-upregulation or co-downregulation). The ratio of genes showing higher magnitude to those showing lower magnitude was significantly higher than expected (3.12 versus 1, Fisher's exact test, $p < 0.001$). These findings confirm a GEP-repressive role for H3K36me3. Interestingly, the authors of this dataset reported a similar function for H3K36me3 in restricting age-dependent gene expression changes, which impacts the *C. elegans* lifespan (32). In addition, they identified similar results using *Drosophila* data, suggesting this phenomenon is conserved across metazoans. H3K36me3 is known to be enriched at actively transcribed genes (34), the increased magnitudes of expression change induced by inactivation of H3K36me3 thus cannot be simply explained by its role in transcriptional activation.

For H3K79me2, we identified two mouse datasets in which the *DOT1L* gene that encodes the histone H3K79 methyltransferase was either knocked out (35) or selectively inhibited (36). In both

mutants, the level of H3K79me2 was dramatically reduced. As above, we compared the magnitudes of expression changes in multiple conditions and found them to be significantly increased (**Figures 7C** and **7D**). In genes showing identical directions of expression change, the ratio of higher magnitude to lower magnitude changes was significantly higher than expected (**Figures 7C** and **7D**). To specifically measure the effect of H3K79me2, we focused on genes enriched for the modification in the gene body region in knockout control samples. For the inhibition dataset, genome-wide H3K79me2 modification data was not available, therefore our analysis was applied to all genes.

For H4K20me1, we identified one mouse dataset in which a neuron-specific isoform of *LSD1* (*LSD1n*), a H4K20 demethylase, was conditionally knocked out (37). This resulted in a significant and specific elevation of H4K20me1 levels (37). We expected this gain of function for H4K20me1 would induce a reduction in GEP. Consistently, loss of *LSD1n* resulted in significantly decreased magnitudes of gene expression change (**Figure 7E**). Among genes showing identical directions of expression change, the ratio of those with higher magnitude changes to lower magnitude changes was significantly lower than the expectation (**Figure 7E**).

Collectively, the above data reveal that certain gene body histone modifications determine gene expression plasticity, in addition to their the well-established functions in regulating gene transcription (38), splicing (34,39), DNA replication (40) and DNA repair (41).

DISCUSSION

Expression plasticity is a conserved gene property

Although expression plasticity has been quantified under a variety of genetic and environmental conditions in different species, there is surprisingly widespread correlation of expression plasticity between orthologs. Importantly, expression plasticity is widely associated with particular gene functions, and this association is also evolutionarily conserved. These findings suggest that the expression plasticity of different genes is optimized to exert biological functions and that plasticity is a

specific target of regulation through evolutionarily conserved mechanisms. An important further question is to what extent expression plasticity varies across cell types. Due to the relative scarcity of cell-specific expression change datasets, we performed a meta-analysis on combined data from cell lines, tissues, and whole organisms. As per the preliminary results shown in **Figure 1C**, different cell types seem to exhibit differential expression plasticity. However, the relatively small number of conditions included (less than 100 for most cases) prevents us from drawing a strong conclusion. It will be interesting to assay and compare the magnitude of gene expression changes in different cell types in response to diverse conditions. Such an assay will provide insights on the implications of expression plasticity for cellular function and on cell-specific mechanisms of plasticity regulation.

Independent regulation of gene expression level, plasticity, and noise

Expression plasticity correlates poorly with gene expression potential (**Figure S2B**), indicating that different mechanisms are used to independently regulate various properties of gene expression. As expected, we found gene expression levels and plasticity to be regulated by distinct genetic and epigenetic mechanisms. As for expression plasticity, we identified DNA motifs that are associated with expression levels in multiple cell lines. However, none of the motifs that influenced expression plasticity were associated with expression level (data not shown). In addition, a genome-wide analysis of the contribution of 38 types of histone modifications to gene expression levels in human cells revealed that H2BK5ac, H3K27ac, H3K79me1, and H4K20me1 are the most important histone modifications for predicting expression levels, all of which positively correlate with expression (42). Again, none of these modifications play similar roles in regulating expression plasticity. It appears that cells use distinct regulatory programs to separately modulate gene expression levels and plasticity.

Genome-wide studies in yeast have revealed gene expression plasticity and expression noise to be highly correlated (1,3,6,19,43). Genes with high expression variability among isogenic cells (high expression noise) tend to exhibit high magnitudes of expression change in response to stimuli (high expression plasticity). This thus suggests the existence of a common mechanism underlying expression

noise and plasticity. Consistent with this, the TATA box promoter element is positively correlated with both gene expression plasticity and cell-to-cell expression noise (4,44). Recent studies have also identified certain types of histone modification as significantly associated with expression noise (44,45), some of which were independently identified here to regulate expression plasticity. However, closer examination of the relationship between noise and plasticity demonstrates that the coupling between the two is highly conditional and evolvable in both yeast (4) and *E. coli* (46). For example, the correlation between noise and plasticity is disfavored for essential genes and haploinsufficient genes, whereby certain genes could have both high plasticity and low noise. It is unclear whether the noise-plasticity coupling also exists in higher organisms. We evaluated the noise-plasticity relationship using a recently published single-cell transcriptome dataset from mouse embryonic stem cells under three culture conditions (47). Where expression noise had a relatively high correlation coefficient between conditions (**Figure S6A**), the correlation between noise and plasticity was significantly lower in all conditions (**Figure S6B**). This suggests the coupling between noise and plasticity is not a general rule in higher organisms. Furthermore, while H3K27me3, H3K4me1, and H3K9ac are associated with high expression noise in mouse embryonic stem cells (44), we found that these modifications do not affect expression plasticity. Interestingly, H3K36me3, H3K79me2, and H4K20me1 have been independently identified as significantly associated with restricted expression noise (45) and plasticity (**Figure 6**). However, the expression noise data (48) and plasticity data used in each study are not correlated (**Figure S6C**, $\rho = 0.023$, $p = 0.059$). This suggests these histone modifications regulate expression noise and plasticity through distinct mechanisms.

Effect of histone modifications on gene expression plasticity

While many histone modifications regulate gene expression level, fewer regulate expression plasticity. Out of all 30 types of histone modifications examined, three modifications (H3K36me3, H3K79me2, and H4K20me1) were found to restrict plasticity, and none promoted plasticity. Previous works have shown that H3K36me3 functions to maintain gene expression stability and fidelity (49,50), and to

maintain epigenetic memory of gene transcription in germ cells (51-53). The negative role of H3K36me3 in regulating expression plasticity is consistent with the reported roles. Whether H3K79me2 and H4K20me1 play a similar function in above processes is worth testing in future studies. All identified histone modifications are plasticity-restricting, raising the possibility that histone modifications may mainly function to down-regulate plasticity to ensure stable and robust expression. All plasticity-restricting modifications were found to be enriched in gene body regions (**Figure S5A**), suggesting transcriptional elongation might be a regulatory target for controlling expression plasticity. A previous study showed that the *C. elegans* ZFP-1/DOT-1.1 complex, a H3K79 methyltransferase, promotes Pol II pausing (54). In addition, inactivation of mouse *LSD1n*, a methylase that removes histone H4K20 methylation, increases Pol II pausing (37). These findings raise the possibility that these gene body histone modifications restrict expression plasticity by inducing Pol II pausing, an important mechanism for regulating gene expression (55). Using human Global Run-on Sequencing (GRO-seq) data (56-58), we found that genes with or without Pol II pausing exhibited similar expression plasticity (**Figure S7**). Pol II pausing index values (59,60), a parameter for quantifying Pol II pausing, were indistinguishable for genes with high and low GEP. Whether gene body histone modifications function through affecting other aspects of transcriptional elongation would be an important question to test in the future.

CONCLUSIONS

In summary, in this study we systematically quantified gene expression plasticity in four metazoan species and performed a comprehensive functional analysis of its properties, implications, and multivariable regulation with two major findings (**Figure 7F**). First, we revealed that expression plasticity is a conserved gene property related to gene function and implicated in disease susceptibility and cellular stability. This finding establishes that the changeability of gene expression is an intrinsic gene property with broad biological implications. Second, we identified genomic and epigenomic signatures that determine expression plasticity genome-wide. These findings significantly expand the

functional repertoire of *cis*-elements, transcription factors, and histone modifications in gene regulation. Together, our work provides insights into the genomic regulation of gene expression flexibility in multicellular organisms.

MATERIALS AND METHODS

Quantification of GEP

The method for GEP quantification used here is based on a previous study (3) which represented GEP as the magnitudes of gene expression change under diverse conditions. We collected expression datasets in which genome-wide gene expression was assayed under at least two different conditions. The magnitude of gene expression change for a condition was determined as a fold change, and the average fold change across a large number of genetic or environmental conditions was used to quantify GEP. The Expression Atlas database (12) collects a large number of manually curated and uniformly processed gene expression datasets spanning many species and biological samples. In particular, its differential experiments section provides processed fold changes in gene expression across many conditions. Expression fold change datasets for human, mouse, and fly genes were directly extracted from the Expression Atlas. Due to the paucity of *C. elegans* expression datasets in the Expression Atlas, worm data was obtained from a previous study in which fold changes of genes were determined across more than 400 conditions (13). This resource has been widely used in the field to study gene co-expression regulatory networks. A considerable number of *C. elegans* genes (n = 2,395 according to WormBase) are organized in operons, and genes within an operon are co-expressed as a polycistron from the same promoter. We excluded such genes from analysis as comparisons of plasticity between them would not be not meaningful.

We calculated GEP (**Table S1**) using a similar method as done previously for yeast data, using the following equation:

$$\mathrm{GEP} = \log_2^{\frac{1}{k}\sum_{i=0}^{k}(\log_2 r_i)^2}$$

where r_i is the fold change of mRNA abundance under a given condition as compared to the control. We required expression datasets to have been assayed using the same platform when a sufficient number of conditions (more than 100) were available for that platform. Otherwise, expression datasets from multiple platforms were combined to achieve the necessary condition number.

Gene attributes

Gene orthologs and functional classification: Lists of orthologous genes were downloaded from the Ensembl genome browser (https://www.ensembl.org) (61). We only considered one-to-one orthologs between species pairs. Enrichment analysis of biological processes was performed for genes with high and low expression plasticity using the Database for Annotation, Visualization and Integrated Discovery (DAVID) functional classification tool (https://david.ncifcrf.gov) with default parameters (18). Conservation of GEP-gene function associations between species was determined using Gene Ontology terms (http://geneontology.org) (62). We first populated each GO term recursively to its parent terms. Only GO terms associated with 20 to 500 genes were considered in our analysis. Then, we calculated the average expression plasticity for genes associated with each GO term. We also collected lists of genes with well-defined biological functions from corresponding databases, including homeobox genes (63), hormone and receptor genes (64), and genes that function in the innate immune system (65).

Gene expression potential and broadness: Gene expression potential is defined as the average gene expression level across a large number of samples. We used expression datasets from two previous studies to measure expression potential: first, the Lukk dataset in which human gene expression level was measured based on 5,372 curated human microarray datasets (17), and second, the functional annotation of the mammalian genome 5 (FANTOM5) project in which genome-wide gene expression was quantified across a large number of tissues and cell types (16). For FANTOM5 dataset, log2 transformed TPM (Transcripts Per Kilobase Million reads) expression data in 230 normal human tissue or cell lines was used. The most abundant transcript was used to represent the gene if multiple transcripts

were associated with a same gene. Broadness of gene expression was calculated by two different methods using the FANTOM5 dataset. In the first method, broadness was quantified as the frequency of samples a gene is expressed. A gene was defined as being expressed in a sample if log (TPM+1) > 0.1. In the second method, broadness was quantified as the frequency of samples a gene is expressed specifically. Specificity of gene expression in a sample was quantified as the expression level in that sample divided by the average expression level across all samples. We defined a gene to be specifically expressed in a sample if the specificity score was over 2.

Gene-disease association, cancer related genes, and stress/environmental response genes: Disease genes were extracted from DisNet (http://www.disgenet.org) (66) and GeneCards (https://www.genecards.org) (67,68). Cancer related genes were defined by IntGOen (https://www.intogen.org) (69) and the Cancer Gene Consensus (http://cancer.sanger.ac.uk/census) (70). Worm genes that are critical for stress/environmental response were curated based on phenotype data from mutants or RNAi-mediated gene knockdown. Gene-associated phenotypes were retrieved from WormBase (15) using simplemine (https://wormbase.org/tools/mine/simplemine.cgi). From all phenotypic terms, we manually curated a collection of 69 phenotypes that were associated with stress/environmental response. A gene was considered required for stress/environmental response if any of those 69 phenotypes was detected in its mutants or RNAi experiments.

DNA motif-centric analysis

TATA box-containing genes were identified using the FindM tool (http://ccg.vital-it.ch/ssa/findm.php) of the Signal search analysis server (71), which scanned for the presence of the TATA box motif in the genomic region from −99 to 0 relative to the transcription start site (TSS) using default parameters. To identify GEP-associated motifs, we used the position weight matrix file for all characterized DNA motifs (CORE 2016 dataset, n = 1,014) obtained from the JASPAR database (http://jaspar2016.genereg.net) (20). We scanned for the presence of each motif in the core promoter regions of genes from four metazoan species (from −200 to +100 relative to the TSS) using the Find

Individual Motif Occurrences (FIMO) tool of the MEME suite with default parameters (72). A DNA motif was considered associated with expression plasticity if a significant difference (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$) in plasticity was detected between genes with and without the motif in their core promoters (**Table S2**). Because the position weight matrices of many DNA motifs are very similar, we further collapsed individual motifs into distinct classes based on matrix similarity. First, cluster information for each of the five taxonomic groups was extracted from JASPAR 2018 (http://jaspar.genereg.net/matrix-clusters), and GEP-associated motifs belonging to a given cluster were combined. Then, motif clusters were compared between taxonomic groups to further collapse similar motifs into one class. Through this process, the 141 individual motifs associated with expression plasticity were collapsed into 40 distinct motif classes.

Evaluation of pairwise combinations of DNA motifs revealed three types of relationships: additivity, enhancement, and dominance. For this analysis, we predicted the expected GEP of promoters containing both motifs (GEP_{exp}) by summing the effects of the individual motifs and then comparing that prediction with the observed GEP values from promoters with both motifs (GEP_{obs}). A relationship was defined to be additive if GEP_{obs} was indistinguishable from GEP_{exp} ($GEP_{obs} = GEP_{exp}$). Otherwise, GEP_{obs} was further compared to the GEP values of promoters containing only one of the two motifs (GEP_{m1} or GEP_{m2}) to determine whether enhancement or dominance occurred using the rules listed in **Figure S3E**. Enhancement was defined as the combined effect of two motifs exerting significantly stronger influence on GEP than the sum of their separate effects, and occurs only between motifs that influence GEP in the same direction. A dominance interaction occurred when the effect of one motif was masked by that of another, and applied only in cases of motifs influencing GEP in opposite directions. Statistical significance of GEP values were determined using the Mann-Whitney U test with a p value cutoff of 0.05.

As GEP was estimated from changes in gene expression, DNA motifs were considered to influence GEP if they influenced gene expression levels. We examined the contribution of 1,014 motifs to expression levels using three baseline gene expression datasets sourced from the Expression Atlas including GTEx (53 normal tissues) (73), the Illumina Body Map (16 normal tissue types) (74), and the

Roadmap Epigenomics Project (57 tissue and cell lines) (75). For each sample, we compared expression levels between genes with and without a specific motif in their core promoter regions (from −200 to +100 relative to the TSS). A motif was considered to be associated with gene expression level if the expression levels between motif-containing and motif-less genes differed significantly ((Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$) and consistently in more than 80% of the total 126 samples analyzed.

Regulatory protein-centric analysis

Genome-wide *in vivo* binding data for human regulatory proteins was produced by the ENCODE project (21) and extracted from the UCSC Genome Browser (https://www.genome.ucsc.edu/ENCODE) (76). In total, these data consisted of 505 datasets containing binding data for 159 regulatory proteins in 91 cell lines (**Table S3**). Binding patterns were generated for a large collection of regulatory proteins in many distinct samples by ChIP-seq experiments, and uniformly processed to identify protein binding peaks throughout the genome. To identify regulatory protein occupancy in the promoter regions of target genes, protein-bound peaks were intersected with the promoter regions of all human genes (from −500 bp to +500 bp relative to the annotated TSS defined by Ensembl Genome Browser). To determine whether the binding of a certain regulatory protein was associated with expression plasticity, we compared expression plasticity between genes with and without binding of that protein in their promoter regions. We defined a regulatory protein as being associated with expression plasticity if it met the following criteria: (i) the difference in GEP was statistically significant (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$), (ii) the directions of difference were consistent in at least 80% of all examined samples, and (iii) genome-wide binding was assayed in more than three distinct biological samples.

Histone modification analysis

Histone modification datasets were collected from the ENCODE project, Roadmap Epigenomics Project, and the GEO database. Processing for peak regions of histone modifications was performed in the original studies. To identify evolutionarily conserved histone modifications that regulate expression plasticity, we first identified candidate modifications from the human data and then examined whether they could be validated using data from other species. We used human histone modification data from the Roadmap Epigenomics Project (http://egg2.wustl.edu/roadmap), which contained 978 datasets illustrating the genome-wide distribution of 30 types of histone modifications in 127 samples (75). For each dataset, we downloaded the uniformly processed broadpeaks and narrowpeaks files to examine their enrichment in gene body regions. Only peaks with p values $< 10^{-4}$ were used. If a peak overlapped with the gene body, the gene was considered to contain the modification. A histone modification was considered to regulate expression plasticity in the human genome if: (1) expression plasticity between genes with or without the modification was significantly different (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$), (2) the effects of the histone modification on expression plasticity (promoting or repressive) were consistent in over 80% of examined samples, and (3) data was available for more than three samples.

Magnitude of gene expression change in perturbation experiments

We searched exhaustively for gene expression datasets in which expression levels were assayed for at least two different conditions in both wild type and mutant samples. These datasets were necessary to examine whether inactivation of regulatory proteins or histone modifications induces differential magnitudes of gene expression change between conditions. To exclude the possibility that a gene perturbation induced dramatic changes in the gene expression program, which would make comparisons of magnitude meaningless, we only considered datasets where a majority of genes (>60%) were consistently up-regulated or down-regulated between conditions in the *wild type* and mutants. The processed expression values for all microarray probes were downloaded from the NCBI GEO database for each series. To determine gene expression levels, we mapped the probes and associated values onto

all protein-coding genes. Probes that matched to multiple genes were discarded. If multiple probes matched the same gene, their values were averaged. Average gene expression and fold changes were analyzed using the *limma* algorithm (77).

Statistics

All statistical methods and corresponding *p* values were described in the main text, figures or figure legends. Statistical analyses including Pearson correlation, partial correlation, Spearman rank correlation, Mann-Whitney U Test, Wilcoxon signed-rank test, t-test, and Fisher's exact test were performed using Python.

AUTHORS' CONTRIBUTIONS

Z.D. and F.H. conceived and designed the project, F.H.; Z.D., L.X., and Z.Z. performed the data analysis; Z.D. wrote the manuscript with the input of other authors.

COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENTS

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences [XDB19000000 to Z.D.] and National Natural Science Foundation of China [31571535, 31722035 to Z.D.].

REFERENCES

1. Tirosh, I., Barkai, N. and Verstrepen, K.J. (2009) Promoter architecture and the evolvability of gene expression. *Journal of biology*, **8**, 95.
2. Lopez-Maury, L., Marguerat, S. and Bahler, J. (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature reviews. Genetics*, **9**, 583-593.
3. Tirosh, I., Weinberger, A., Carmi, M. and Barkai, N. (2006) A genetic signature of interspecies variations in gene expression. *Nature genetics*, **38**, 830-834.
4. Lehner, B. (2010) Conflict between noise and plasticity in yeast. *PLoS genetics*, **6**, e1001185.
5. Tirosh, I. and Barkai, N. (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res*, **18**, 1084-1091.
6. Choi, J.K. and Kim, Y.J. (2009) Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nature genetics*, **41**, 498-503.
7. GuhaThakurta, D., Palomar, L., Stormo, G.D., Tedesco, P., Johnson, T.E., Walker, D.W., Lithgow, G., Kim, S. and Link, C.D. (2002) Identification of a novel cis-regulatory element involved in the heat shock response in Caenorhabditis elegans using microarray gene expression and computational methods. *Genome Res*, **12**, 701-712.
8. Grishkevich, V., Ben-Elazar, S., Hashimshony, T., Schott, D.H., Hunter, C.P. and Yanai, I. (2012) A genomic bias for genotype-environment interactions in C. elegans. *Mol Syst Biol*, **8**, 587.
9. Li, Y., Alvarez, O.A., Gutteling, E.W., Tijsterman, M., Fu, J., Riksen, J.A., Hazendonk, E., Prins, P., Plasterk, R.H., Jansen, R.C. *et al.* (2006) Mapping determinants of gene expression plasticity by genetical genomics in C. elegans. *PLoS genetics*, **2**, e222.
10. Girardot, F., Monnier, V. and Tricoire, H. (2004) Genome wide analysis of common and specific stress responses in adult drosophila melanogaster. *BMC Genomics*, **5**, 74.

11. Murray, J.I., Whitfield, M.L., Trinklein, N.D., Myers, R.M., Brown, P.O. and Botstein, D. (2004) Diverse and specific gene expression responses to stresses in cultured human cells. *Molecular biology of the cell*, **15**, 2361-2374.

12. Papatheodorou, I., Fonseca, N.A., Keays, M., Tang, Y.A., Barrera, E., Bazant, W., Burke, M., Fullgrabe, A., Fuentes, A.M., George, N. *et al.* (2018) Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res*, **46**, D246-D251.

13. Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N. and Davidson, G.S. (2001) A gene expression map for Caenorhabditis elegans. *Science*, **293**, 2087-2092.

14. Kandasamy, K., Mohan, S.S., Raju, R., Keerthikumar, S., Kumar, G.S., Venugopal, A.K., Telikicherla, D., Navarro, J.D., Mathivanan, S., Pecquet, C. *et al.* (2010) NetPath: a public resource of curated signal transduction pathways. *Genome Biol*, **11**, R3.

15. Howe, K.L., Bolt, B.J., Cain, S., Chan, J., Chen, W.J., Davis, P., Done, J., Down, T., Gao, S., Grove, C. *et al.* (2016) WormBase 2016: expanding to enable helminth genomic research. *Nucleic Acids Res*, **44**, D774-780.

16. Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T. *et al.* (2014) A promoter-level mammalian expression atlas. *Nature*, **507**, 462-470.

17. Lukk, M., Kapushesky, M., Nikkila, J., Parkinson, H., Goncalves, A., Huber, W., Ukkonen, E. and Brazma, A. (2010) A global map of human gene expression. *Nature biotechnology*, **28**, 322-324.

18. Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C. and Lempicki, R.A. (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*, **8**, R183.

19. Landry, C.R., Lemos, B., Rifkin, S.A., Dickinson, W.J. and Hartl, D.L. (2007) Genetic properties influencing the evolvability of gene expression. *Science*, **317**, 118-121.

20. Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R. *et al.* (2016) JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*, **44**, D110-115.

21. Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57-74.

22. Xiao, T., Wallace, J. and Felsenfeld, G. (2011) Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Molecular and cellular biology*, **31**, 2174-2183.

23. Rhodes, J.M., Bentley, F.K., Print, C.G., Dorsett, D., Misulovin, Z., Dickinson, E.J., Crosier, K.E., Crosier, P.S. and Horsfield, J.A. (2010) Positive regulation of c-Myc by cohesin is direct, and evolutionarily conserved. *Developmental biology*, **344**, 637-649.

24. Hirayama, T., Tarusawa, E., Yoshimura, Y., Galjart, N. and Yagi, T. (2012) CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell reports*, **2**, 345-357.

25. Boj, S.F., van Es, J.H., Huch, M., Li, V.S., Jose, A., Hatzis, P., Mokry, M., Haegebarth, A., van den Born, M., Chambon, P. *et al.* (2012) Diabetes risk gene and Wnt effector Tcf7l2/TCF4 controls hepatic response to perinatal and adult metabolic demand. *Cell*, **151**, 1595-1607.

26. Lazarova, D.L., Chiaro, C. and Bordonaro, M. (2014) Butyrate induced changes in Wnt-signaling specific gene expression in colorectal cancer cells. *BMC research notes*, **7**, 226.

27. Lovas, A., Radke, D., Albrecht, D., Yilmaz, Z.B., Moller, U., Habenicht, A.J. and Weih, F. (2008) Differential RelA- and RelB-dependent gene transcription in LTbetaR-stimulated mouse embryonic fibroblasts. *BMC Genomics*, **9**, 606.

28. Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T. *et al.* (2008) Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell*, **132**, 422-433.

29. Seitan, V.C., Faure, A.J., Zhan, Y., McCord, R.P., Lajoie, B.R., Ing-Simmons, E., Lenhard, B., Giorgetti, L., Heard, E., Fisher, A.G. *et al.* (2013) Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res*, **23**, 2066-2077.

30. Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B. *et al.* (2015) CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*, **163**, 1611-1627.

31. Lu, Y., Shan, G., Xue, J., Chen, C. and Zhang, C. (2016) Defining the multivalent functions of CTCF from chromatin state and three-dimensional chromatin interactions. *Nucleic Acids Res*, **44**, 6200-6212.

32. Pu, M., Ni, Z., Wang, M., Wang, X., Wood, J.G., Helfand, S.L., Yu, H. and Lee, S.S. (2015) Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. *Genes & development*, **29**, 718-731.

33. Andersen, E.C. and Horvitz, H.R. (2007) Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. *Development*, **134**, 2991-2999.

34. Kolasinska-Zwierz, P., Down, T., Latorre, I., Liu, T., Liu, X.S. and Ahringer, J. (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature genetics*, **41**, 376-381.

35. Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A.V., Feng, Z., Punt, N., Daigle, A., Bullinger, L. *et al.* (2011) MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer cell*, **20**, 66-78.

36. Daigle, S.R., Olhava, E.J., Therkelsen, C.A., Majer, C.R., Sneeringer, C.J., Song, J., Johnston, L.D., Scott, M.P., Smith, J.J., Xiao, Y. *et al.* (2011) Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer cell*, **20**, 53-65.

37. Wang, J., Telese, F., Tan, Y., Li, W., Jin, C., He, X., Basnet, H., Ma, Q., Merkurjev, D., Zhu, X. *et al.* (2015) LSD1n is an H4K20 demethylase regulating memory formation via transcriptional elongation control. *Nature neuroscience*, **18**, 1256-1264.

38. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823-837.

39. Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M. and Misteli, T. (2010) Regulation of alternative splicing by histone modifications. *Science*, **327**, 996-1000.

40. Han, J., Zhou, H., Horazdovsky, B., Zhang, K., Xu, R.M. and Zhang, Z. (2007) Rtt109 acetylates histone H3 lysine 56 and functions in DNA replication. *Science*, **315**, 653-655.

41. Sanders, S.L., Portoso, M., Mata, J., Bahler, J., Allshire, R.C. and Kouzarides, T. (2004) Methylation of histone H4 lysine 20 controls recruitment of Crb2 to sites of DNA damage. *Cell*, **119**, 603-614.

42. Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K. and Vingron, M. (2010) Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci U S A*, **107**, 2926-2931.

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

44. Faure, A.J., Schmiedel, J.M. and Lehner, B. (2017) Systematic Analysis of the Determinants of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst*, **5**, 471-484 e474.

45. Wu, S., Li, K., Li, Y., Zhao, T., Li, T., Yang, Y.F. and Qian, W. (2017) Independent regulation of gene expression level and noise by histone modifications. *PLoS Comput Biol*, **13**, e1005585.

46. Singh, G.P. (2013) Coupling between noise and plasticity in E. coli. *G3*, **3**, 2115-2120.

47. Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C., Ilicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Buhler, M., Liu, P. *et al.* (2015) Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. *Cell Stem Cell*, **17**, 471-485.

48. Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., DaleyKeyser, A., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J. *et al.* (2014) Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, **516**, 56-61.

49. Sen, P., Dang, W., Donahue, G., Dai, J., Dorsey, J., Cao, X., Liu, W., Cao, K., Perry, R., Lee, J.Y. *et al.* (2015) H3K36 methylation promotes longevity by enhancing transcriptional fidelity. *Genes & development*, **29**, 1362-1376.

50. Meers, M.P., Henriques, T., Lavender, C.A., McKay, D.J., Strahl, B.D., Duronio, R.J., Adelman, K. and Matera, A.G. (2017) Histone gene replacement reveals a post-transcriptional role for H3K36 in maintaining metazoan transcriptome fidelity. *eLife*, **6**.

51. Furuhashi, H., Takasaki, T., Rechtsteiner, A., Li, T., Kimura, H., Checchi, P.M., Strome, S. and Kelly, W.G. (2010) Trans-generational epigenetic regulation of C. elegans primordial germ cells. *Epigenetics Chromatin*, **3**, 15.

52. Rechtsteiner, A., Ercan, S., Takasaki, T., Phippen, T.M., Egelhofer, T.A., Wang, W., Kimura, H., Lieb, J.D. and Strome, S. (2010) The histone H3K36 methyltransferase MES-4 acts epigenetically to transmit the memory of germline gene expression to progeny. *PLoS genetics*, **6**, e1001091.

53. Kreher, J., Takasaki, T., Cockrum, C., Sidoli, S., Garcia, B.A., Jensen, O.N. and Strome, S. (2018) Distinct Roles of Two Histone Methyltransferases in Transmitting H3K36me3-Based Epigenetic Memory Across Generations in Caenorhabditis elegans. *Genetics*, **210**, 969-982.

54. Cecere, G., Hoersch, S., Jensen, M.B., Dixit, S. and Grishok, A. (2013) The ZFP-1(AF10)/DOT-1 complex opposes H2B ubiquitination to reduce Pol II transcription. *Mol Cell*, **50**, 894-907.

55. Adelman, K. and Lis, J.T. (2012) Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nature reviews. Genetics*, **13**, 720-731.

56. Gao, W.W., Xiao, R.Q., Zhang, W.J., Hu, Y.R., Peng, B.L., Li, W.J., He, Y.H., Shen, H.F., Ding, J.C., Huang, Q.X. *et al.* (2018) JMJD6 Licenses ERalpha-Dependent Enhancer and Coding Gene Activation by Modulating the Recruitment of the CARM1/MED12 Co-activator Complex. *Mol Cell*, **70**, 340-357 e348.

57. Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T. and Sandelin, A. (2014) Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat Commun*, **5**, 5336.

58. Chen, L., Chen, J.Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H. *et al.* (2017) R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol Cell*, **68**, 745-757 e745.

59. Chae, M., Danko, C.G. and Kraus, W.L. (2015) groHMM: a computational tool for identifying unannotated and cell type-specific transcription units from global run-on sequencing data. *BMC Bioinformatics*, **16**, 222.

60. Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K. and Lis, J.T. (2012) Defining the status of RNA polymerase at promoters. *Cell reports*, **2**, 1025-1035.

61. Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M., Amode, R., Brent, S. *et al.* (2016) Ensembl comparative genomics resources. *Database : the journal of biological databases and curation*, **2016**.

62. Gene Ontology, C. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res*, **43**, D1049-1056.

63. Zhong, Y.F. and Holland, P.W. (2011) HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol Dev*, **13**, 567-568.

64. Rashid, M., Singla, D., Sharma, A., Kumar, M. and Raghava, G.P. (2009) Hmrbase: a database of hormones and their receptors. *BMC Genomics*, **10**, 307.

65. Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L., Hancock, R.E., Brinkman, F.S. and Lynn, D.J. (2013) InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Res*, **41**, D1228-1233.

66. Pinero, J., Queralt-Rosinach, N., Bravo, A., Deu-Pons, J., Bauer-Mehren, A., Baron, M., Sanz, F. and Furlong, L.I. (2015) DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database : the journal of biological databases and curation*, **2015**, bav028.

67. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y. *et al.* (2016) The GeneCards Suite: From Gene Data Mining

to Disease Genome Sequence Analyses. *Current protocols in bioinformatics*, **54**, 1 30 31-31 30 33.

68. Rebhan, M., Chalifa-Caspi, V., Prilusky, J. and Lancet, D. (1997) GeneCards: integrating information about genes, proteins and diseases. *Trends in genetics : TIG*, **13**, 163.

69. Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M.P., Jene-Sanz, A., Santos, A. and Lopez-Bigas, N. (2013) IntOGen-mutations identifies cancer drivers across tumor types. *Nature methods*, **10**, 1081-1082.

70. Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N. and Stratton, M.R. (2004) A census of human cancer genes. *Nature reviews. Cancer*, **4**, 177-183.

71. Ambrosini, G., Praz, V., Jagannathan, V. and Bucher, P. (2003) Signal search analysis server. *Nucleic Acids Res*, **31**, 3618-3620.

72. Grant, C.E., Bailey, T.L. and Noble, W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017-1018.

73. Consortium, G.T. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648-660.

74. Asmann, Y.W., Necela, B.M., Kalari, K.R., Hossain, A., Baker, T.R., Carr, J.M., Davis, C., Getz, J.E., Hostetter, G., Li, X. *et al.* (2012) Detection of redundant fusion transcripts as biomarkers or disease-specific therapeutic targets in breast cancer. *Cancer Res*, **72**, 1921-1928.

75. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. *et al.* (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317-330.

76. Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G. *et al.* (2013) ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res*, **41**, D56-63.

77. Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*, **3**, Article3.

78. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T.W., Greven, M.C., Pierce, B.G., Dong, X., Kundaje, A., Cheng, Y. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res*, **22**, 1798-1812.

FIGURES LEGENDS

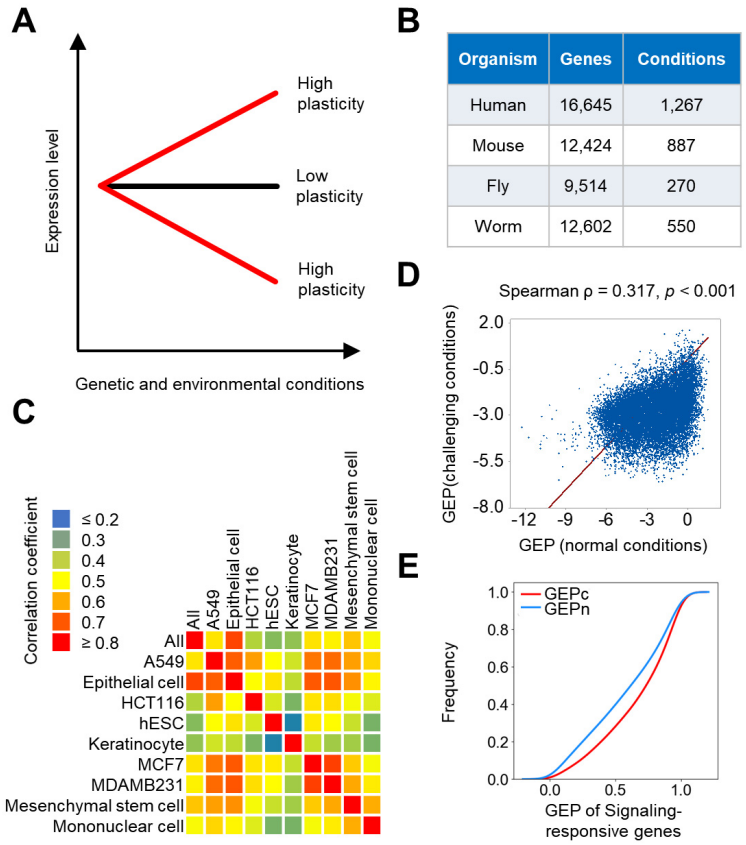


Figure 1. Quantification of gene expression plasticity in four metazoan species

(**A**) Schematic diagram shows the definition of GEP based on changes of expression levels across genetic and environmental conditions. Genes with a high GEP exhibit dynamic expression across conditions, while genes with a low GEP exhibit stable expression. (**B**) Summary of data sources. Figure shows the species, numbers of genes, and numbers of conditions used for quantifying GEP. (**C**) Heatmap showing correlation coefficient between global and cell-specific GEP. (**D**) Correlation between GEP measured using gene expression under challenging and normal conditions. For normal conditions, fold changes of gene expression was quantified between expression level in a sample and the average level across all samples. (**E**) Signaling-responsive genes exhibited higher GEP measured using gene expression changes under challenging conditions (GEPc) than that under normal conditions (GEPn). To compare GEP level of a same gene calculated using different expression datasets, GEP

levels were normalized as percentile rank across all genes and figure plots the cumulative distribution of normalized GEP levels.

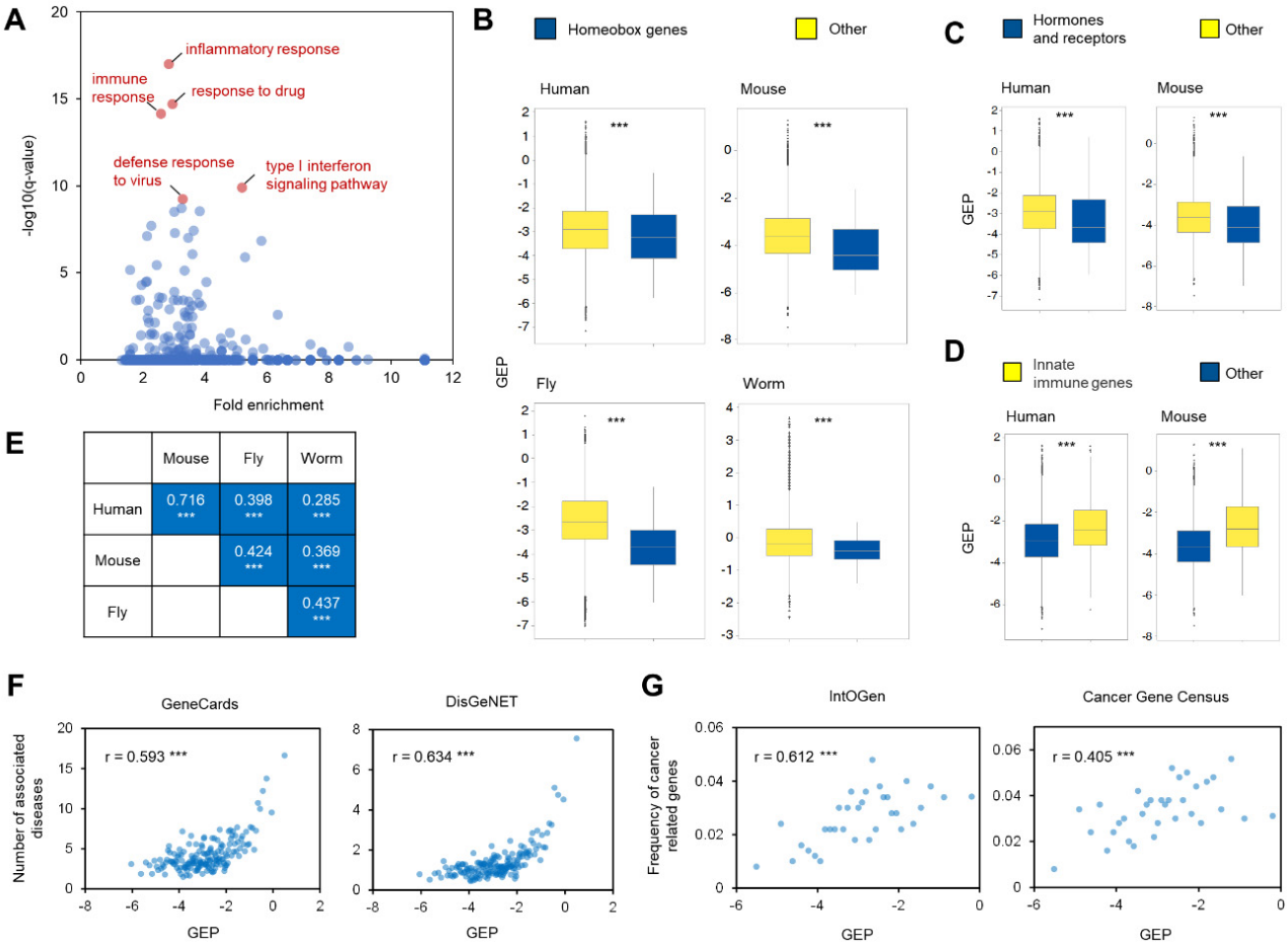


Figure 2. Expression plasticity is associated with gene function and disease susceptibility

(**A**) Enrichment biological processes for genes with high GEP. Scatter plot shows the fold enrichment and q values of all biological processes for genes exhibiting high GEP (top 10%, n = 1,643). The top five most significantly enriched processes are marked in red. (**B**) Homeobox genes exhibited significantly lower GEP. Homeobox genes were defined by the HomeoDB2 classification (63) and GEP values were compared between homeobox genes (n = 198 for human, n = 152 for mouse, n = 77 for fly, and n = 62 for worm) and all other genes in each species. *** denotes Mann–Whitney U test, $p < 0.001$. (**C**) Hormones and receptor genes exhibited significantly lower GEP in human and mouse. Genes were taken from Hmrbase (n = 181 and 107 for human and mouse respectively) (64). (**D**) Innate immune genes exhibited significantly higher GEP in human and mouse. Innate immune genes were taken from InnateDB (n = 898 and 431 for human and mouse respectively) (65). (**E**) Ontology term GEP is highly

correlated between species. GEP values for each ontology term were calculated by averaging the GEP of all genes belonging to that term. Figure shows the correlation coefficients between ontology term GEP values for all cross-species comparisons. *** denotes $p < 0.001$. (**F**) Genes with high GEP are implicated in disease susceptibility. Genes were ranked by GEP and the average number of associated diseases calculated for each bin (100 genes). Gene-disease associations were extracted from GeneCards (68) and DisGeNet (66). (**G**) Genes with high GEP are implicated in cancer. Genes were ranked by GEP and the frequency of cancer related genes was calculated for each bin (500 genes). Cancer-related genes were obtained from IntOGen (69) and the Cancer Gene Consensus (70).

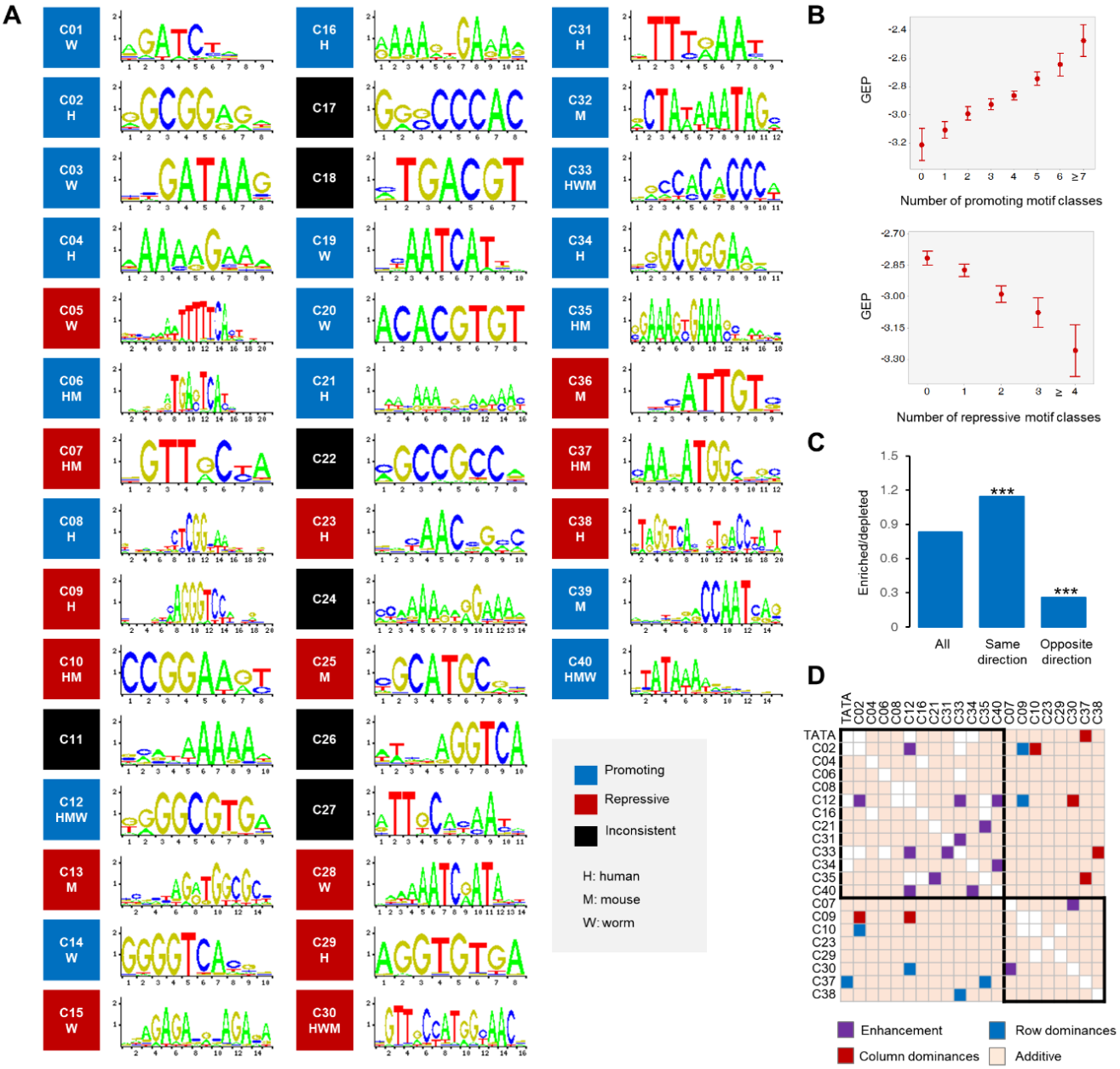


Figure 3. Influence of promoter elements on gene expression plasticity.

(**A**) Figure shows the sequence logo and effects of motif classes associated with GEP. Black indicates motif classes with inconsistent effects; blue and red indicates the 33 motif classes that have GEP-promoting or -repressive roles in corresponding species. Because individual motifs within motif classes exhibit very similar DNA sequence composition, representative motif sequence logos are used. (**B**) Cumulative effects of motif classes on human GEP. Figure summarizes the changes in GEP as more GEP-promoting (left) and -repressing motif classes occur in gene promoters. See Figure S2C for results

from mouse and worm. (**C**) Optimization of motif co-occurrence in human promoters. Co-occurrence was evaluated for all motif pairs that affected GEP in the same or opposite directions. A motif pair was considered enriched or depleted if their observed co-occurrence was significantly higher or lower than expected (Fisher's exact test, $p < 0.05$), respectively. *** denotes a ratio significantly different from that of all motif pairs (Fisher's exact test, $p < 0.001$). See S2D Figure for results from mouse and worm. (**D**) Combinatorial effects of DNA motifs on human GEP. Matrix shows additive (light orange), enhancement (purple), and dominance (blue and red) relationships between all pairwise motif classes. Black boxes highlight motif classes that affect GEP in the same direction. See S2F Figure for results from mouse and worm.

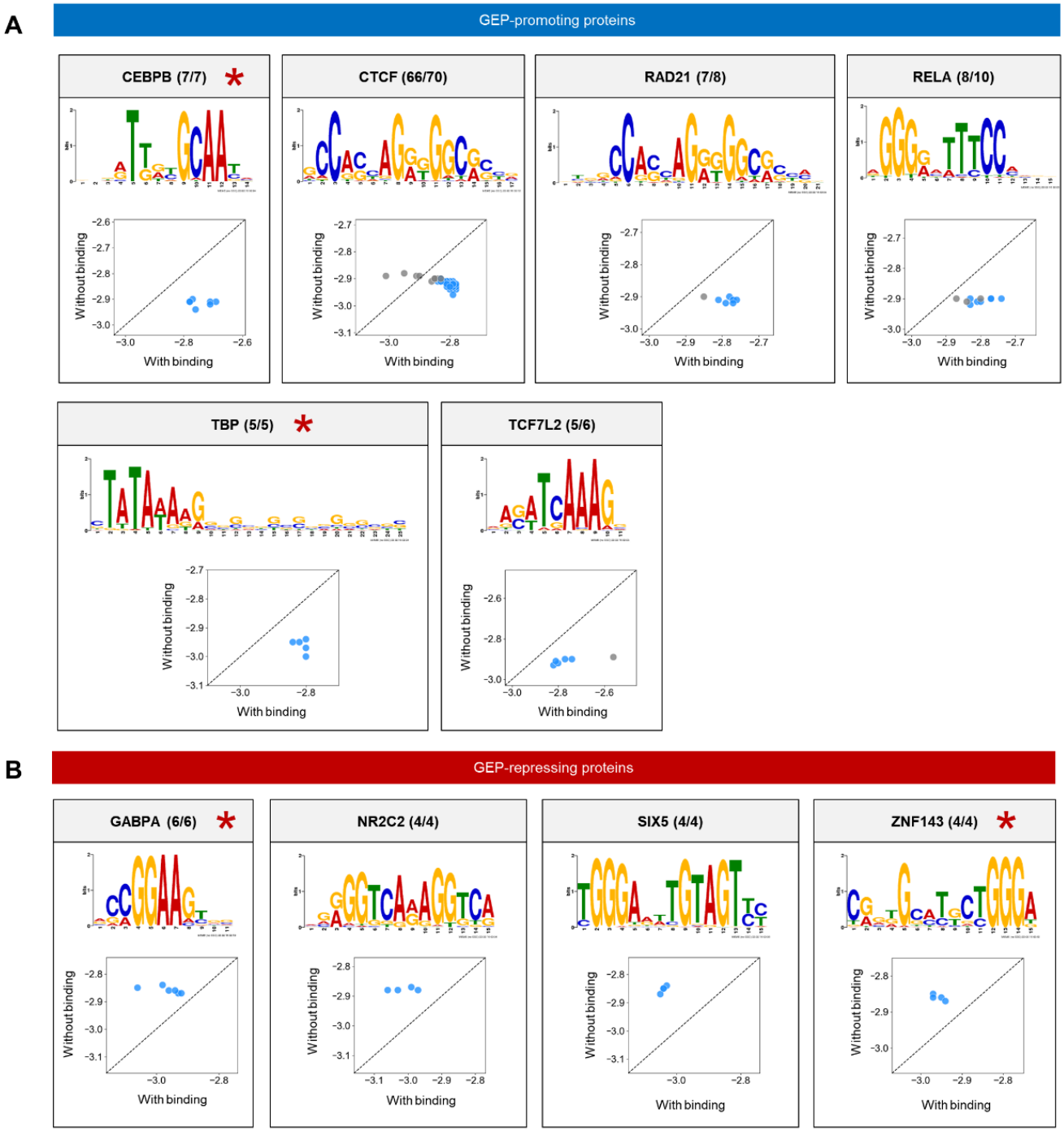


Figure 4. Influence of Regulatory Proteins on Gene Expression Plasticity

GEP-associated regulatory proteins. GEP-promoting (**A**) and -repressing proteins (**B**) were defined as those whose target genes exhibited significantly (Mann–Whitney U Test, Benjamini–Hochberg corrected $p < 0.01$) higher or lower GEP as compared to other genes. Figure shows regulatory protein name, the number of samples with significant GEP association, the total samples examined, and the

preferential binding motif as defined by Factorbook (78). Star indicates the protein binding result is consistent with that from motif analysis. Scatterplot compares GEP between genes with (X-axis) and without (Y-axis) binding in the promoter region (from −500 to +500 relative to the TSS) of regulatory proteins having a promoting (A) or repressive (B) role. Blue indicates $p < 0.01$ (Mann–Whitney U test, Benjamini–Hochberg corrected).

A

Regulatory Proteins	Organism	GEO Accession	Perturbation	Conditions
CTCF	Mouse	GSE38673	Nex-Cre mediated conditional knockout	Cerebral cortex at P7 and hippocampus at P7
RAD21	Zebrafish	GSE18795	Null mutant *rad21*^{nz171}	24hpf and 48hpf
RELA	Mouse	GSE79233	Tissue-specific *RELA* knockout	With and without TNF-α treatment
RELA	Mouse	GSE11963	*RELA*$^{-/-}$ mouse embryonic 3T3 fibroblasts	LTβR stimulation, 0 h and 10 h
TCF7L2/TCF4	Mouse	GSE41284	Disrupted DNA-binding HMG box, null mutant	Liver sample, fasted and fed

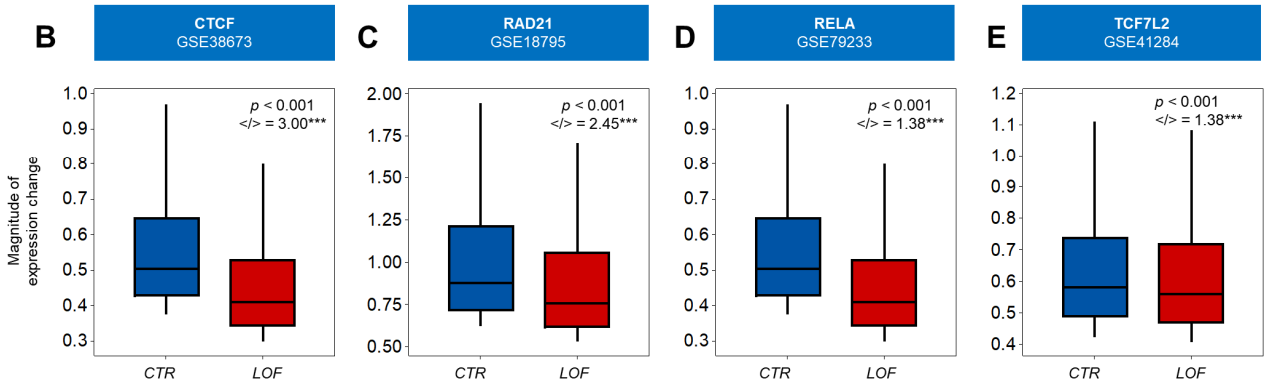


Figure 5. Perturbing regulatory proteins affects magnitude of gene expression changes.

(**A**) Gene expression datasets used to validate the influence of regulatory proteins on GEP. **(B-E)** Loss of function of four regulatory proteins predicted to promote expression plasticity reduced magnitude of gene expression changes. Absolute log2 fold changes of gene expression between two conditions were used to approximate changes in GEP. Box plots show magnitudes for the mostly dynamically expressed genes (top 20%) in control (*CTR*) and loss of function (*LOF*) mutants. Gene numbers for each dataset are: n = 1817 for CTCF (B), n = 1,130 for RAD21 (C), n = 675 for RELA (D), and n = 2,114 for TCF7L2 (E). </> indicates the ratio of genes showing lower change magnitude to those showing higher magnitude. Only genes that exhibited expression changes in the same direction between two conditions (co-upregulation or co-downregulation) were considered. *** denotes a ratio significantly different from expected (Fisher's exact test, $p < 0.001$).

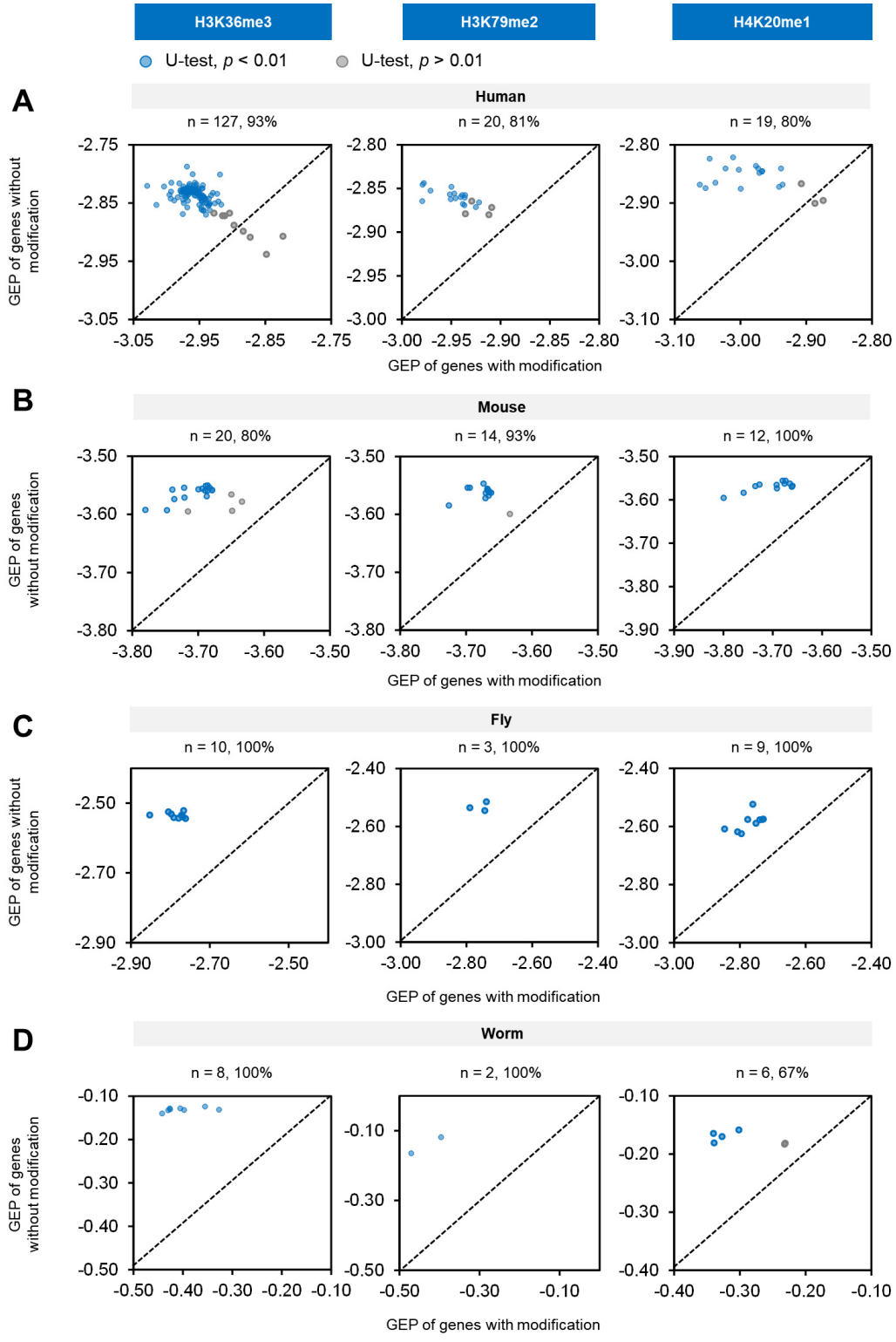


Figure 6. Genes enriched for gene body H3K36me3, H3K79me2, and H4K20me1 modifications exhibit restricted expression plasticity.

(**A-D**) Gene body H3K36me3, H3K79me2, and H4K20me1 modifications restrict GEP in four species. Each scatterplot shows median GEP for genes with (X axis) or without (Y axis) a histone modification in the gene body region. Samples showing significant difference (Mann–Whitney U test, Benjamini–Hochberg corrected $p < 0.01$) are indicated in blue. Diagonal line indicates equal GEP. The number of samples analyzed and the percentage of samples with consistent differences are shown above each scatterplot.

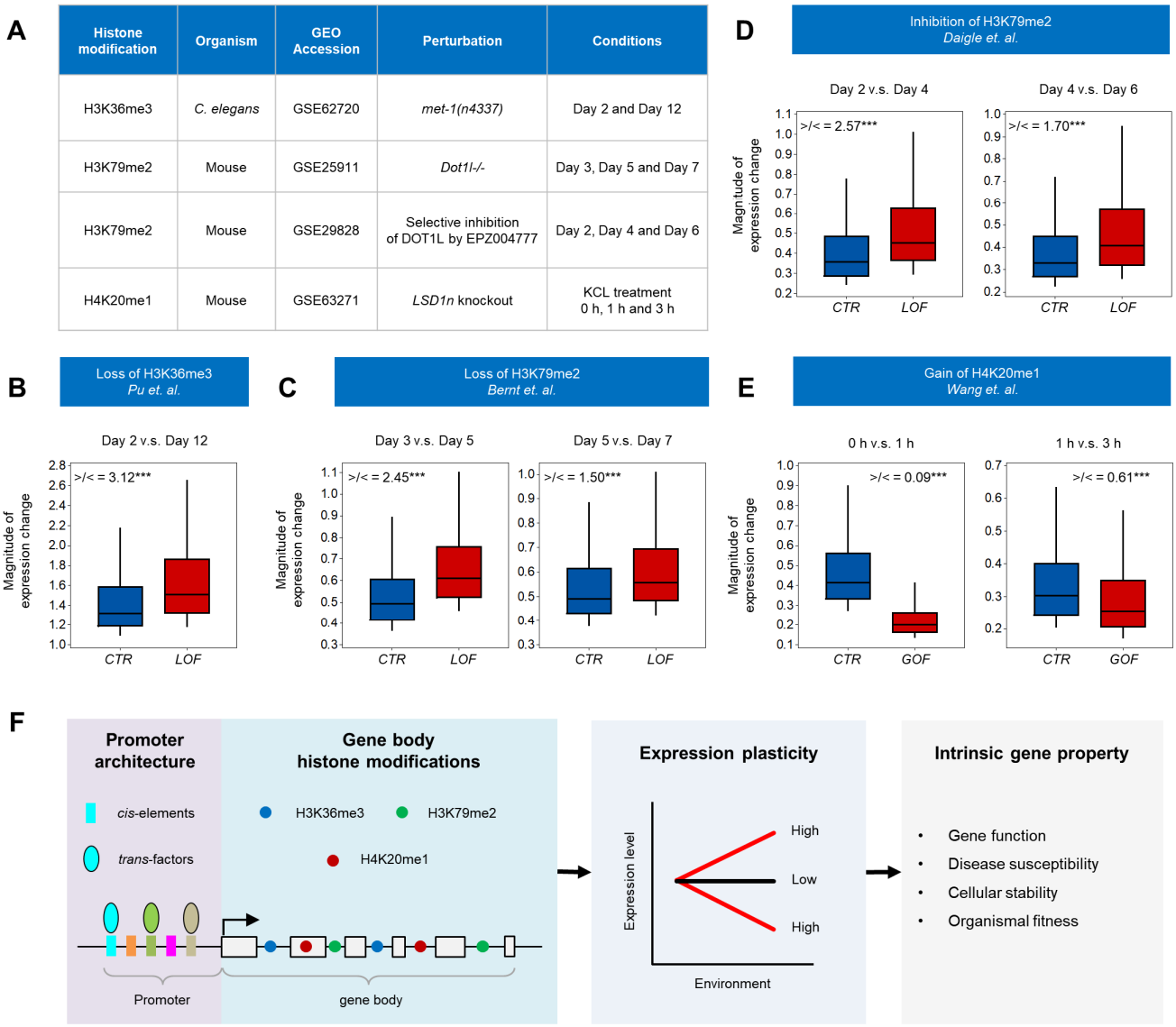


Figure 7. Perturbing histone modifications affects magnitude of gene expression changes.

(**A**) Gene expression datasets used to validate the effect of histone modifications on GEP. (**B-E**) Perturbation of H3K36me3 (B), H3K79me2 (C, D), and H4K20me1 (E) induced expected changes in the magnitudes of gene expression changes between two conditions. Absolute log2 fold changes of gene expression between two conditions were used to proximate GEP changes. Box plots show magnitudes for the most dynamically expressed genes (top 20%) in control (*CTR*) and loss (LOF) or gain (GOF) of function mutations. Gene numbers for all datasets are: n = 535 from Pu *et. al.*; n = 141 (Day 3 v.s. Day

5) and n = 190 (Day 5 v.s. Day 7) from Bernt *et. al.*; n = 1,483 (Day 2 v.s. Day 4) and n = 460 (Day 4 v.s. Day 6) from Daigle *et. al.*; and n = 1,062 (0 h v.s. 1 h) and n = 1,106 (1 h v.s. 3 h) from Wang *et. al.*. >/< indicates the ratio of the number of genes exhibiting larger magnitudes of change to those exhibiting smaller magnitudes for genes having the same direction of expression change between two conditions. *** denotes a ratio significantly different from expected (Fisher's exact test, $p < 0.001$). (**F**) Summary of findings. Gene structure is illustrated by linked gray boxes with an arrow indicating transcription start site. Rectangles indicate *cis*-elements and colored ovals depict regulatory proteins binding in the promoter region. Circles indicate gene body histone modifications.

SUPPLEMENTARY DATA

SUPPLEMENTARY FIGURES

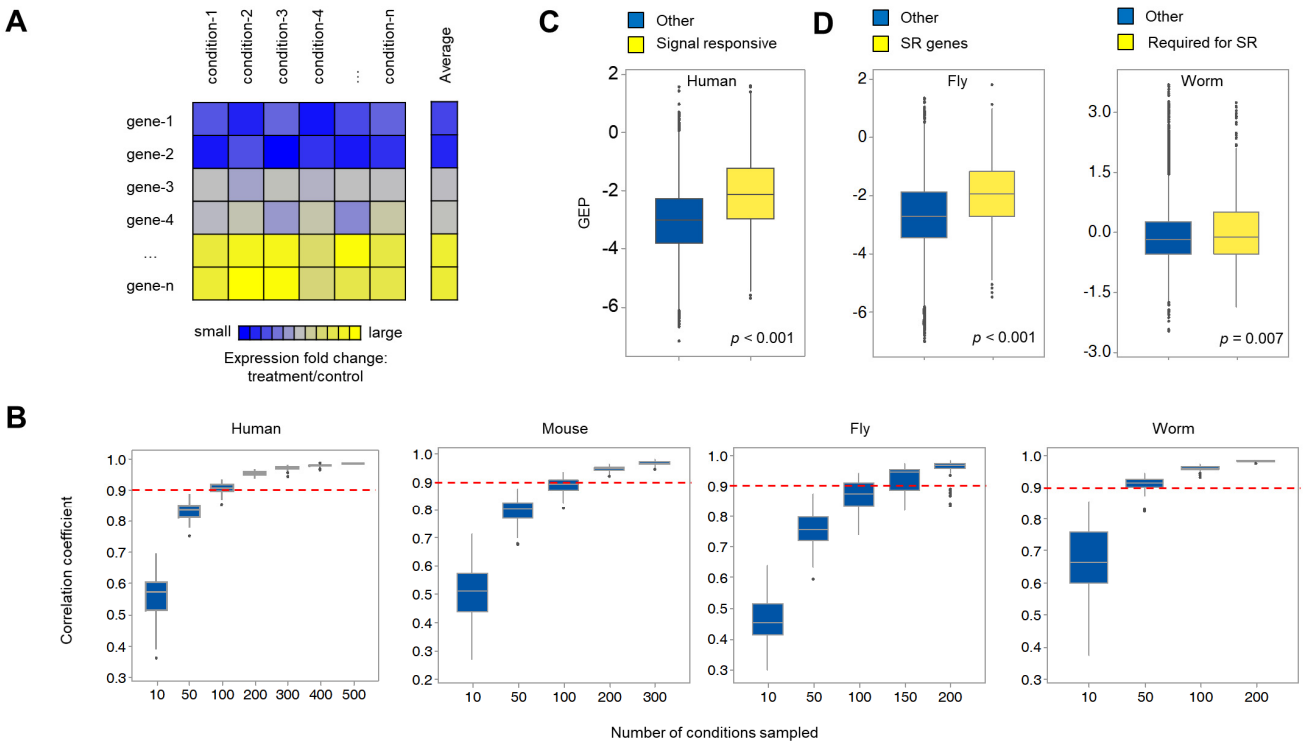


Figure S1. Quality control of gene expression plasticity.

(**A**) Graphic showing the strategy for quantifying GEP from changes of gene expression measured across diverse conditions. (**B**) Sufficiency of condition number for GEP quantification. The influence of condition numbers on GEP quantification was simulated by comparing the Pearson correlation coefficients of GEP values calculated using all conditions and those using only a subset of all conditions in all species. Each sampling was repeated 100 times. Red dashed line indicates a correlation coefficient of 0.9. (**C**) Human signaling responsive genes (n = 2,221) (14) exhibit significantly higher GEP than other genes. (**D**) Genes implicated in stress response show significantly higher GEP. Fly stress response (SR) genes (n = 891) were identified previously (10); worm genes (n = 505) required for stress response were those whose inactivation (by RNAi or mutation) caused hypersensitivity to stress conditions (69 associated phenotypic terms), obtained from WormBase (15).

A

Organism	Gene number	Spearman's Rank Correlation
Human-Mouse	10,260	ρ = 0.46 ***
Human-Fly	2,102	ρ = 0.15 ***
Human-Worm	904	ρ = 0.12 ***
Mouse-Fly	1,569	ρ = 0.18 ***
Mouse-Worm	687	ρ = 0.17 ***
Fly-Worm	1,075	ρ = 0.14 ***

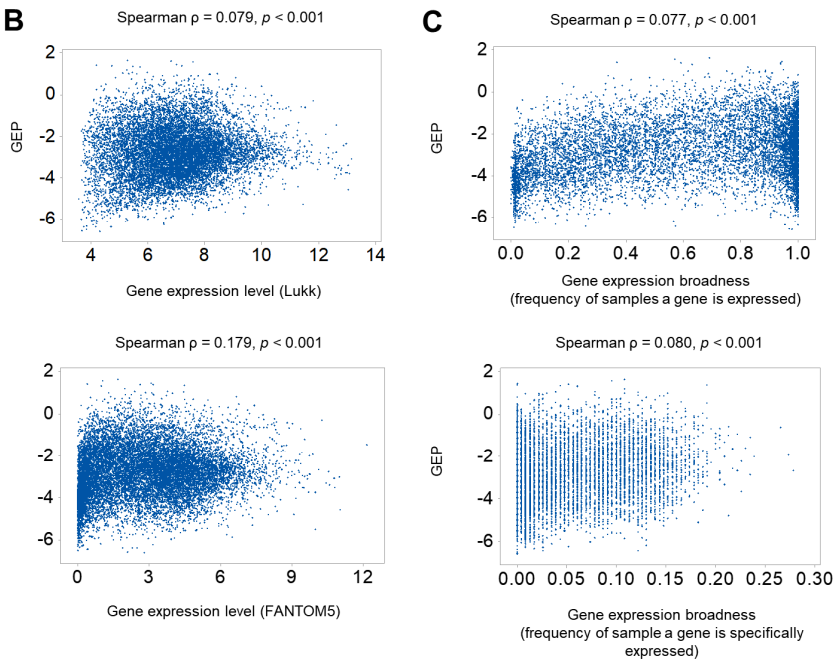
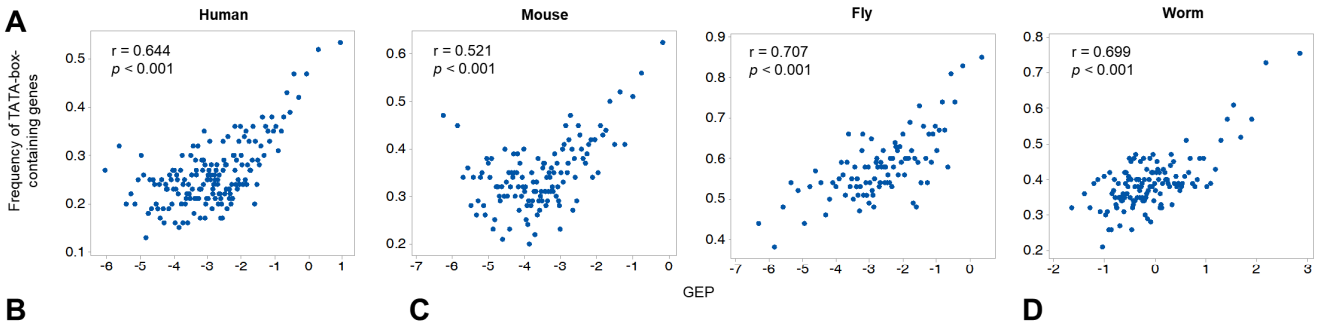
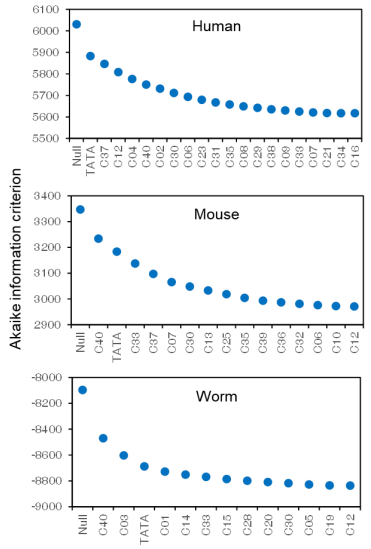
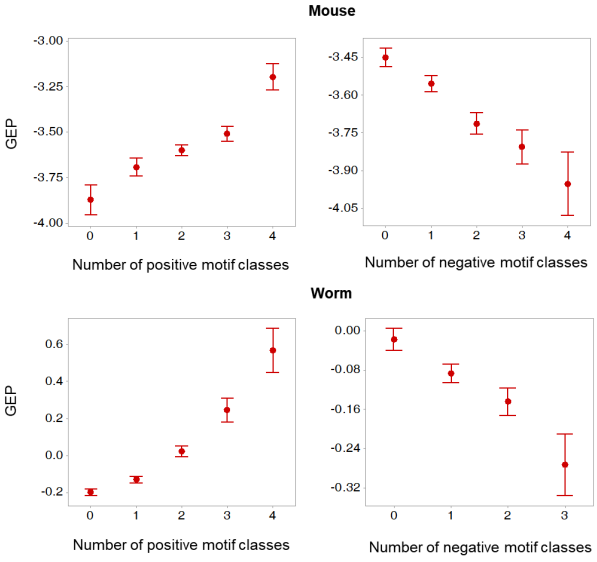


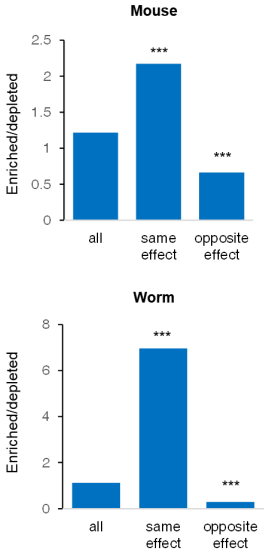
Figure S2. Gene expression plasticity is conserved across species and measure a distinct feature of gene expression.

(**A**) GEP of orthologous genes is conserved in four metazoan species. Pairwise correlation coefficients were calculated for GEP of one-to-one orthologous genes in four metazoan species (61). Figure shows the overall Spearman's rank correlation coefficient between species. *** denotes $p < 0.001$. (**B**) GEP is poorly correlated with gene expression potential. Scatterplot compares GEP and expression potential measured using two different datasets. (**C**) GEP is poorly correlated with gene expression broadness measured using two different methods.









E

Motif 1	Motif 2	Enhancement	Dominance (motif 1 masks 2)	Dominance (motif 2 masks 1)	Additive
Promoting	Promoting	$GEP_{obs}>GEP_{exp}$ $GEP_{obs}>GEP_{M1}$ $GEP_{obs}>GEP_{M2}$	NA	NA	$GEP_{obs}=GEP_{exp}$
Repressive	Repressive	$GEP_{obs}<GEP_{exp}$ $GEP_{obs}<GEP_{M1}$ $GEP_{obs}<GEP_{M2}$	NA	NA	$GEP_{obs}=GEP_{exp}$
Promoting	Repressive	NA	$GEP_{obs}>GEP_{exp}$ $GEP_{obs}=GEP_{M1}$ $GEP_{obs}>GEP_{M2}$	$GEP_{obs}<GEP_{exp}$ $GEP_{obs}<GEP_{M1}$ $GEP_{obs}=GEP_{M2}$	$GEP_{obs}=GEP_{exp}$
Repressive	Promoting	NA	$GEP_{obs}<GEP_{exp}$ $GEP_{obs}=GEP_{M1}$ $GEP_{obs}<GEP_{M2}$	$GEP_{obs}>GEP_{exp}$ $GEP_{obs}>GEP_{M1}$ $GEP_{obs}=GEP_{M2}$	$GEP_{obs}=GEP_{exp}$

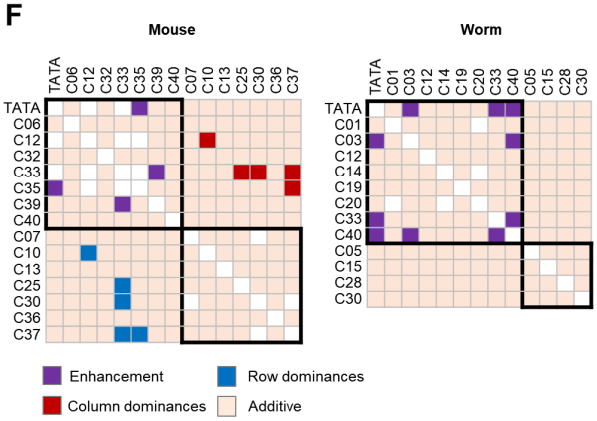


Figure S3 *Cis*-regulation of gene expression plasticity.

(**A**) Promoter TATA box is associated with high GEP in four metazoan species. Genes were ordered by GEP and the frequency of TATA box-containing genes calculated for each bin of 100 genes. (**B**) Contributions of DNA motif classes to GEP. Figure shows the reduction of Akaike information criterion (AIC) score following inclusion of all selected motif classes in the predictive model for each species. Motif selection was performed using a stepwise AIC procedure to identify the contribution of each motif class after considering the influence of other motifs. (**C**) Cumulative effect of motifs. Changes in GEP as a function of more GEP-promoting (left) or GEP-repressing (right) motif classes in promoters of mouse and worm genes. (**D**) Optimization of motif co-occurrence. Bar graph gives the ratios of enriched to depleted co-occurrence of two DNA motif classes for all pairwise motif pairs (All) and motif pairs that influence GEP in the same or opposite directions. Enriched and depleted motif pairs are defined as those with observed co-occurrence significantly higher or lower than expected (Fisher's exact test, $p < 0.05$), respectively. *** denotes ratio significantly different from that of all motif pairs (Fisher's exact test, $p < 0.001$). (**E**) Rules used to identify interactions between pairwise motifs. Details described in Methods section. (**F**) Combinatorial effects of motif classes on GEP in mouse and worm. Black boxes highlight motif classes that affect GEP in the same direction.

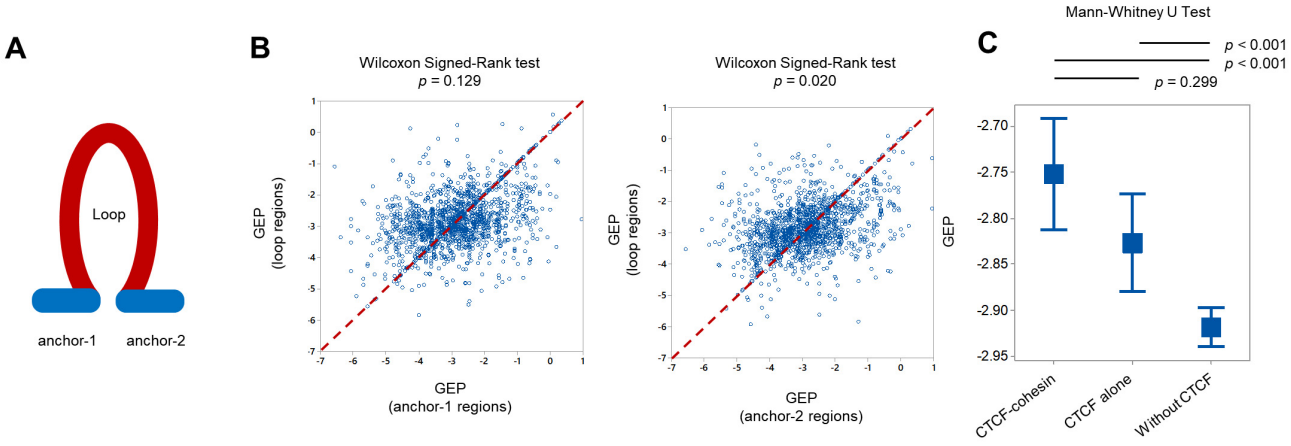


Figure S4. CTCF-mediated chromatin looping is not associated with expression plasticity.

(**A**) Schematic of chromatin loop structure. (**B**) Scatterplot compares GEP between genes with promoters located within anchors (X-axis, n = 2139 for anchor-1 and n = 2,084 for anchor-2) and those with promoters located in loop regions (Y-axis). If promoters of multiple genes are present in an anchor or loop region, their GEP was averaged. Diagonal line indicates equal GEP levels. (**C**) Comparison of GEP between genes whose promoter is bound by CTCF-cohesin (associated with chromatin loop, n = 1,469), bound by CTCF alone (n = 1,685), and not bound by CTCF (n = 12,545). Data is represented as mean ± 95% confidence interval.

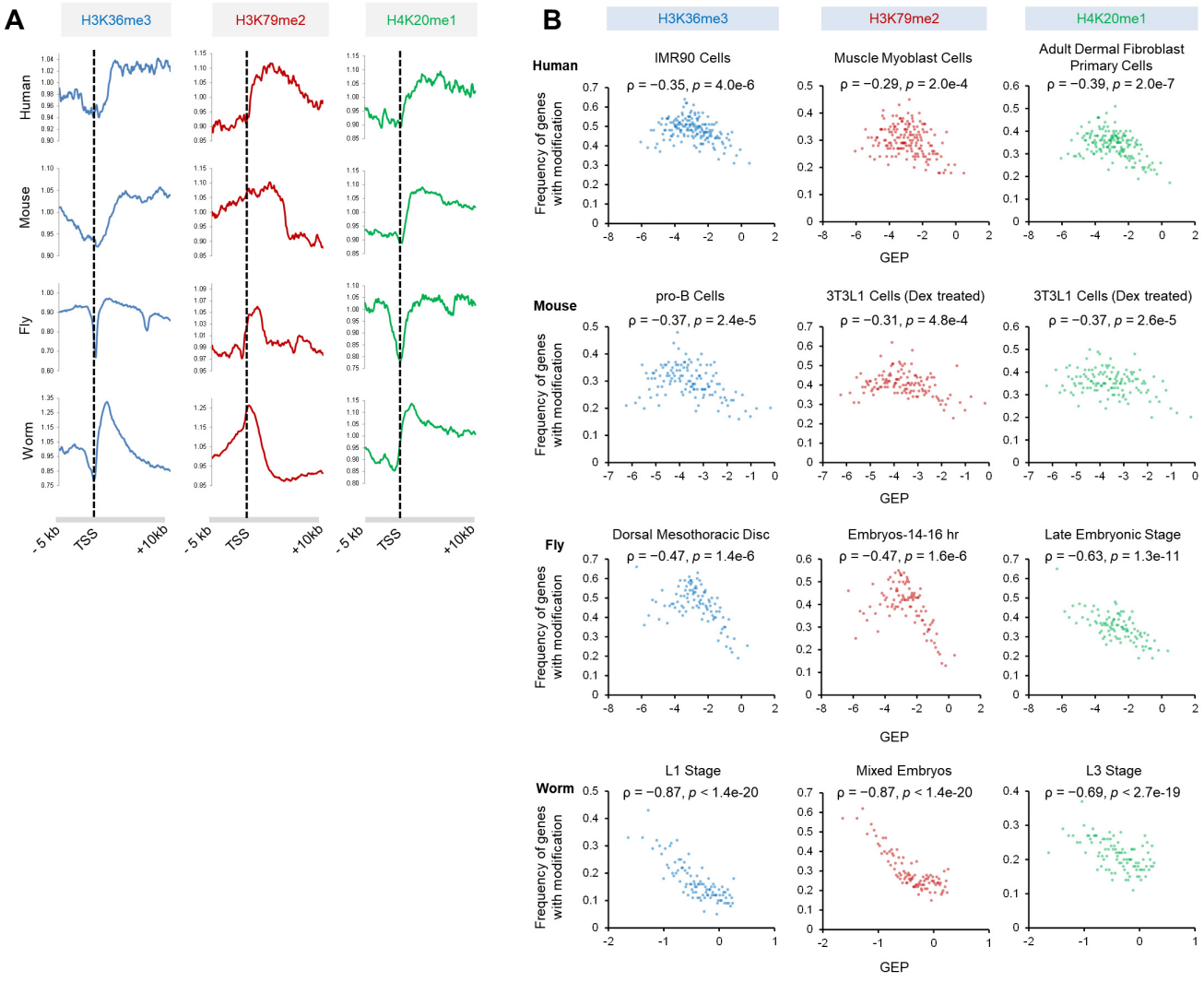


Figure S5. Influence of gene body histone modifications on gene expression plasticity.

(**A**) Enrichment of H3K36me3, H3K79me2, and H4K20me1 in gene body. Figure shows the relative distribution of H3K36me3 (blue), H3K79me2 (red), and H4K20me1 (green) modifications in genomic regions from −5,000 to +10,000 relative to the annotated TSS for all genes. Plots show one representative sample for each modification in each species. In fly and worm, the sharp drop in histone modification signal immediately downstream of the TSS may have been due to their small gene sizes. (**B**) Gene body H3K36me3, H3K79me2, and H4K20me1 are associated with restricted GEP. Genes were ranked by expression plasticity and the average frequency of genes enriched for H3K36me3 (blue),

H3K79me2 (red), and H4K20me1 (green) modifications in the gene body was calculated for each bin (100 genes). One representative sample for each modification are shown for each species.

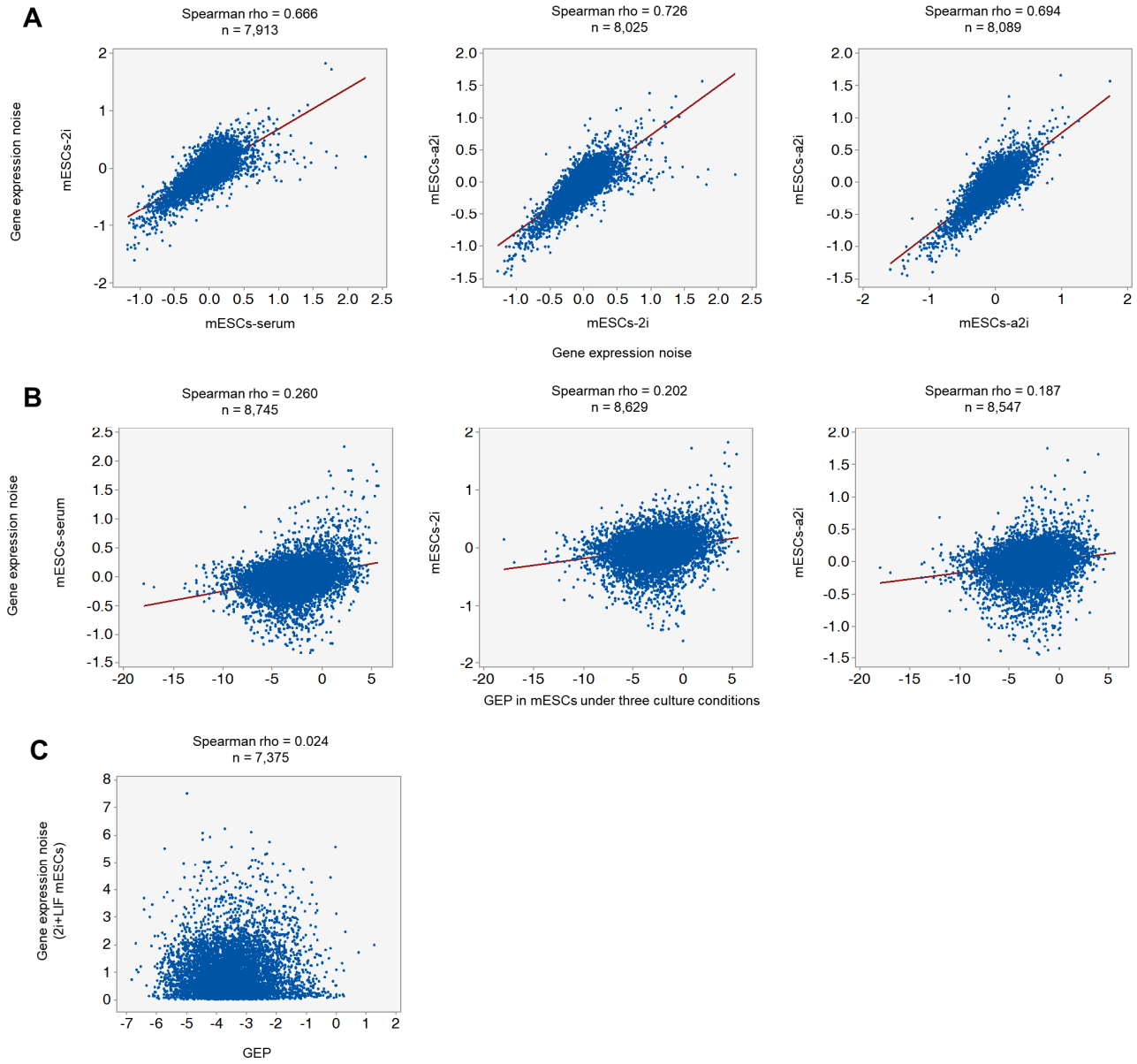


Figure S6. Decoupling of gene expression plasticity and noise.

(**A**) Correlation of gene expression noise under different culture conditions. Scatterplots show pairwise correlations of gene expression noise in mouse embryonic stem cells (mESCs) cultured under three different conditions (serum, 2i, and a2i). Gene expression noise was determined by the original study (47). (**B**) Correlation between GEP and gene expression noise in mESCs. Scatterplots show the correlation between GEP in mESCs and expression noise in three culture conditions. GEP in mESCs was measured using the same method as in Figure 1 by determining the magnitudes of gene

expression fold changes between the culture conditions. (**C**) Scatter plot shows correlation between GEP and gene expression noise in which the same types of histone modifications are identified to restrict GEP and expression noise respectively. As in the previous study (45), coefficients of variation of gene expression levels in 94 single mESCs cultured under 2i+LIF conditions (48) were used to quantify expression noise.

A

GRO-seq Dataset	GEP of genes without Pol II pausing	GEP of genes with Pol II pausing	Mann-Whitney U test P value
GSM2706105_MCF7	−2.8387 n = 9,375	−2.8532 n = 4,689	0.0436
GSM2706107_MCF7	−2.8355 n = 9,353	−2.8481 n = 4,608	0.1595
GSE62046_HeLa	−2.8318 n = 5,658	−2.8145 n = 7,381	0.4723
GSM2551016_HEK293T	−2.8503 n = 8,703	−2.8144 n = 4,601	0.0479

B

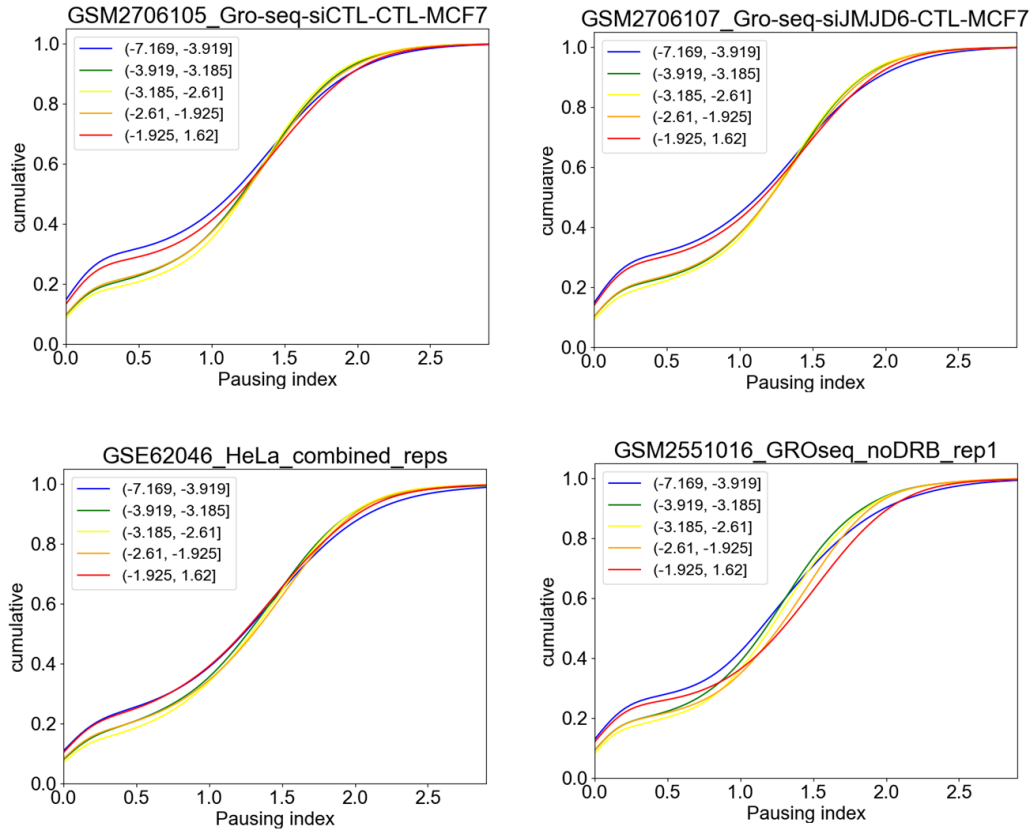


Fig. S7 RNA Pol II pausing is not associated with GEP.

(**A**) Genes with or without Pol II pausing exhibit similar GEP values. RNA Pol II pausing was determined from four Global Run-on Sequencing (GRO-seq) datasets (56-58) using the groHMM algorithm (59). Nascent mRNA read densities were compared between promoter-proximal regions (−50 to +300 relative to the TSS) and in gene body regions (from +300 relative to the TSS to +3000 from the gene end) and genes were identified as having Pol II pausing if disproportionately high (Fisher's exact test, $p < 0.001$) read density was present in the promoter-proximal region. Only actively

transcribed genes were analyzed, as determined by significantly enriched (p <0.01) read density in the gene body compared to background (bottom 1% read density of all genes), as done in a previous study (60). (**B**) Comparison of Pol II pausing index between genes with different GEP. Genes were divided into five bins according to GEP levels and the cumulative distribution of Pol II pausing index (log2 transformed) plotted. The pausing index was calculated as the ratio of read density in promoter-proximal regions to that of gene body regions for all actively transcribed genes.

SUPPLEMENTARY TABLES

Table S1. GEP of four metazoan species.

Sheets 1 to 4. Gene expression plasticity of human, mouse, fly, and worm genes.

Table S2. Core promoter motifs associated with GEP.

Sheet 1. Identified DNA motifs, effects on GEP, motif classes, and associated factors in four species.

Sheets 2-4. Presence of motif classes in human, mouse, and worm genes.

Table S3. Binding of regulatory proteins at the promoter is associated with GEP.

Sheet 1. Effects of all analyzed regulatory proteins on expression plasticity in all samples.

Sheet 2. Summary of results.

Table S4. Data sources for histone modifications.

Table shows data sources for H3K36me3, H3K79me2, and H4K20m1 modifications for all species.