



Chromatin enables precise and scalable gene regulation with factors of limited specificity

Mindy Liu Perkins^a , Justin Crocker^a, and Gašper Tkačik^{b,1}

Edited by Joanna Wysocka, Stanford University School of Medicine, Stanford, CA; received June 13, 2024; accepted November 22, 2024

Biophysical constraints limit the specificity with which transcription factors (TFs) can target regulatory DNA. While individual nontarget binding events may be low affinity, the sheer number of such interactions could present a challenge for gene regulation by degrading its precision or possibly leading to an erroneous induction state. Chromatin can prevent nontarget binding by rendering DNA physically inaccessible to TFs, at the cost of energy-consuming remodeling orchestrated by pioneer factors (PFs). Under what conditions and by how much can chromatin reduce regulatory errors on a global scale? We use a theoretical approach to compare two scenarios for gene regulation: one that relies on TF binding to free DNA alone and one that uses a combination of TFs and chromatin-regulating PFs to achieve desired gene expression patterns. We find, first, that chromatin effectively silences groups of genes that should be simultaneously OFF, thereby allowing more accurate graded control of expression for the remaining ON genes. Second, chromatin buffers the deleterious consequences of nontarget binding as the number of OFF genes grows, permitting a substantial expansion in regulatory complexity. Third, chromatin-based regulation productively co-opts nontarget TF binding for ON genes in order to establish a “leaky” baseline expression level, which targeted activator or repressor binding subsequently up- or down-modulates. Thus, on a global scale, using chromatin simultaneously alleviates pressure for high specificity of regulatory interactions and enables an increase in genome size with minimal impact on global expression error.

chromatin | gene regulation | crosstalk | optimization | regulatory networks

Many biological functions require certain subsets of genes to be active independently of others. The targeted activation or repression of gene expression is achieved by regulatory factors including transcription factors (TFs), which must specifically locate and bind the regulatory regions for their target genes. However, some level of nontarget binding is unavoidable due to the chemical nature of protein–DNA interactions (1–5). Organisms have therefore evolved several different mechanisms for increasing the ratio of target to nontarget binding. In prokaryotes, TFs bind almost uniquely to their cognate binding sites, which are long enough to appear only very rarely by chance in the genome. Paradoxically, though the eukaryotic genome is two to three orders of magnitude larger than the prokaryotic genome, binding sites in eukaryotes tend to be shorter and more degenerate than in prokaryotes and, consequently, cannot specify unique genomic locations (6). This biophysical observation has motivated researchers to propose alternative means for eukaryotes to maintain targeted gene regulation, including cooperativity (7–9), clustering of low-affinity sites (6, 10–13), kinetic proofreading (14–18), and combinatorial control of transcription (9, 19–22).

One notable possibility for reducing unwanted gene expression involves the use of chromatin, the physical packaging of DNA by nucleosomes, and associated proteins largely unique to eukaryotes (23). Nucleosomes may physically obstruct TF binding to DNA wrapped around the core (24–26) or directly inhibit transcription initiation by blocking the assembly of the preinitiation complex (27–31). Longer-term silencing of gene expression additionally requires specific patterns of DNA methylation (32), histone modifications (33), and/or higher-order chromatin structures to maintain a transcriptionally repressive state (34–36). Although condensed or compacted chromatin may hinder large regulatory factors or complexes from accessing binding sites (26, 37–40) and most TFs, if they manage to bind nucleosomal DNA, will dissociate quickly (26), some factors are capable of reversing chromatin-based repression by recruiting other factors to decompactify chromatin (41–43) or evict, displace, or disassemble nucleosomes from DNA (38, 44). Such TFs with “pioneering activity” (PFs) induce the dynamic changes to genetic silencing that are associated with proper cell fate specification during embryonic development (32, 34, 45, 46). Moreover, introducing binding sites for PFs

Significance

Reliably keeping a gene off is as important as controlling its expression level when the gene is on. Yet both tasks become challenging in the packed nuclear environment of a eukaryotic cell, where the numerous and diverse regulatory proteins that are present cannot bind enhancers for target genes with perfect specificity. While regulatory schemes based on prokaryotic models would be overwhelmed by errors in such conditions, we show that chromatin-based regulation, an evolutionary innovation of eukaryotic cells, successfully rescues precise gene expression control by reliably keeping desired genes off. Our systems-level computational analysis demonstrates that this result is nontrivial, because chromatin opening must itself be correctly regulated. We furthermore identify when and how chromatin-based regulation outperforms alternative schemes.

Author affiliations: ^aDevelopmental Biology Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany; and ^bInstitute of Science and Technology Austria, AT-3400 Klosterneuburg, Austria

Author contributions: M.L.P., J.C., and G.T. designed research; M.L.P. performed research; M.L.P. and G.T. analyzed data; and M.L.P., J.C., and G.T. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2024 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹To whom correspondence may be addressed. Email: gatkacik@ist.ac.at.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2411887121/-DCSupplemental>.

Published December 30, 2024.

into regulatory sequences (47, 48)—or even into random sequences (49)—is sufficient to induce or boost expression in many contexts. Whether “pioneer factors” are an entirely distinct category from “ordinary” TFs (50), or whether TFs simply vary in their capacity to induce “pioneering activity” based on context, remains a subject of open discussion (51).

Introducing chromatin alleviates some nontarget TF binding by rendering DNA inaccessible—but not without a cost. In order to selectively (de)silence genes, chromatin remodelers and PFs must themselves be able to target certain binding sites in chromatin while ignoring others. Indeed, it has been proposed that PFs may enhance the recruitment of chromatin remodelers to appropriate nucleosomes (16) as part of an energy-consuming kinetic proofreading scheme (52, 53), variants of which may improve specificity by a factor of 300 to 400 over equilibrium schemes (54). The relationship between target specificity and chromatin appears even more intricate in light of recent work indicating that chromatin can improve the functional specificity of proteins with otherwise degenerate DNA binding domains, for instance, the Hox class of homeodomain proteins (7, 55). Thus, chromatin-based transcriptional regulation may contribute in nontrivial ways to the resolution of the binding site specificity paradox in eukaryotes (12).

While many works have considered chromatin-based transcriptional regulation at individual loci (24, 29, 42, 43, 56–62), we are unaware of any that consider the global implications of using a combination of chromatin and free DNA-binding TFs for gene regulatory programs. The omission is notable for two reasons. First, using both chromatin and TFs enables new modes of genetic regulation spanning multiple genes or regulatory regions; for example, the same TF could be reused in multiple networks simply by changing the accessibility of its target regulatory regions. Second, limited binding site specificity constitutes a challenge only inasmuch as multiple factors and multiple nontarget sites are present simultaneously; in other words, it emerges as a nontrivial constraint only when considered at the global level (9).

Here, we address this knowledge gap by employing a theoretical approach. Our focus is on global gene expression patterning: the task of accurately modulating individual expression levels for multiple active (ON) genes simultaneously, while keeping subsets of silent (OFF) genes from expressing altogether. We ask whether gene expression patterning benefits from chromatin-based mechanisms relative to regulatory schemes relying on TFs alone and find biophysically realistic regimes of operation where chromatin lowers gene expression patterning errors by one to two orders of magnitude. The systems-level benefits we identify in this work synergize with the recently identified evolutionary benefits of chromatin-based regulation, suggesting how such complex schemes could have evolved and become selectively maintained (63).

Model Setup

Our model of chromatin landscaping and transcriptional regulation is deliberately simple, retaining only the key features of gene expression patterning that can be treated independently of the exact molecular implementation. Where possible, we have chosen parameter values to match measurements or estimates available in the literature.

The ultimate goal of our study is to comparatively evaluate two scenarios for gene regulation. In the first scenario, the *free DNA* scenario, gene expression levels depend on the concentrations of TFs only; there is no chromatin and no PFs. In the *chromatin*

scenario, gene expression depends on the concentrations of both TFs and PFs. Both scenarios are set up so that their “wiring diagrams” are equivalent, and in the hypothetical limit where all regulating factors (PFs and TFs) have perfect specificity, they would perform equally well. In this paper, we compute what happens in both scenarios outside of this limit, at finite factor specificity: How precisely gene expression levels could be driven toward different target expression patterns, assuming that the concentrations of their regulating factors could be optimized without constraint toward each target (Fig. 1). Because no practical control scenario could ever outperform the mathematical optimum, the residual (minimal) *global expression error* (GEE) of this “best-case” outcome is a well-defined and theoretically justified performance measure—a loss function—for each scenario, introduced in previous work as detrimental “regulatory crosstalk” (9).

In order to perform the mathematical optimization, we must first specify which TFs or PFs regulate which genes. Experimental evidence indicates that genes can be grouped into clusters that share temporal patterns of chromatin accessibility (64). We therefore group genes into clusters of fixed size, with each cluster being regulated by one distinct *multitarget* factor. Experimental evidence also suggests that TFs outnumber factors with strong pioneering activity (65, 66). Therefore, in addition to being regulated by a multitarget factor, we allow each gene to be regulated by one *single-target* factor that does not regulate any other gene (i.e., there is a one-to-one mapping from single-target factors to genes). Thus, each gene is targeted by exactly one single-target and one multitarget factor (Fig. 1A). We initially assume all regulatory factors behave as activators and lift this assumption later.

The two regulatory scenarios to be compared employ the regulatory factors as follows: 1) in the *free DNA scenario*, both the single-target and the multitarget factors are TFs; and 2) in the *chromatin scenario*, the single-target factors are TFs and the multitarget factors are PFs. In the free DNA case, both TFs need to be bound simultaneously to activate a gene, but there is no sense of temporal ordering—that is, it does not matter whether the multitarget or single-target factor binds first. The expression level of a gene is therefore simply the probability that both the single-target and multitarget factors are bound. In contrast, in the chromatin scenario, PFs must bind first to open enhancer chromatin (35, 64, 67) before TFs can bind to modulate transcription levels (Fig. 1B); gene expression is thus the product of the steady-state probabilities that chromatin is permissive and the single-target TF is bound (35, 64, 67). This scenario does not require simultaneous binding by PFs and TFs at a regulatory locus.

As a consequence, once we fix the number of genes and clusters, we can assess a regulatory system comparatively in both the free DNA and chromatin scenarios, simply by changing the model for multitarget factor binding from TFs to PFs, respectively (*SI Appendix, section S1*).

In principle, transcription or chromatin opening can be initiated by any binding event—be it “target” or “nontarget”—albeit with different probability. To account for this observation, we assume that nontarget binding is weaker than target binding and occurs only among sites of the same type; that is, TFs cannot bind PF sites and vice versa. In contrast to the chromatin scenario, the free DNA scenario therefore experiences nontarget binding between single- and multitarget factors. The two scenarios also differ in their ability to distinguish between target and nontarget binding. While TFs follow a typical thermodynamic (equilibrium) binding scheme, we adopt a kinetic proofreading

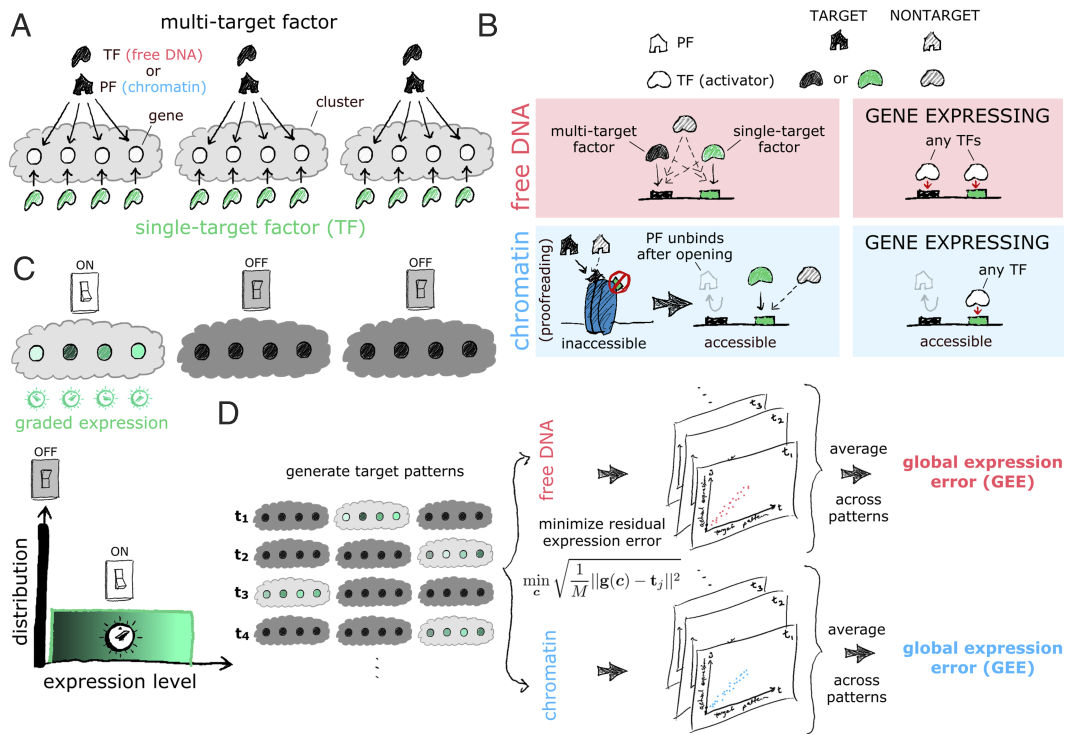


Fig. 1. Model schematic. (A) Each gene (circle) is targeted for regulation by one multitarget factor (black; house-shaped PF in chromatin, bean-shaped TF in free DNA) and one single-target factor (green; TF). Genes targeted by the same multitarget factor are said to belong to the same cluster (gray cloud). (B) PFs (house shapes) and TFs (bean shapes) bind target and nontarget sites on inaccessible chromatin and free DNA/accessible chromatin, respectively, leading to chromatin opening or gene expression. (C) A target pattern is randomly generated by assigning all genes in the same cluster to the same ON or OFF state. OFF genes have zero target expression (black), while ON genes have randomly drawn graded expression, here a uniform distribution (green). (D) Global gene expression error (GEE) measures the patterning accuracy of a regulatory scenario (*free DNA* in red vs. *chromatin* in blue) given a fixed number of genes and clusters (as in panel A). GEE is calculated by optimizing regulatory factor concentrations to achieve target expression patterns as closely as possible for all genes simultaneously. Expression levels are calculated based on mathematical models for TF and PF binding (*SI Appendix, section S1*).

scheme for chromatin opening that allows for stronger rejection of nontarget binding by PFs than is possible at equilibrium (*SI Appendix, Fig. S1*). Kinetic proofreading expends energy to pass through at least one intermediate transition step between initial PF binding and final transcriptionally permissive state, e.g., by recruiting remodelers to stabilize chromatin in an open conformation (5, 68).

To enable fair comparison between the two scenarios, we assign to all factors, TFs and PFs alike, the same intrinsic dissociation constants: K_T for their target and K_{NT} for all other (nontarget) sites. The ratio $S = K_{NT}/K_T$ is the *intrinsic specificity* of the factors; we vary it between simulations by changing K_{NT} . Our model is agnostic to whether nontarget binding arises from low-affinity binding, from truly nonspecific TF–DNA interactions, or by some other molecular mechanism. Further details of the models may be found in *SI Appendix, section S1*.

We can now mathematically define the GEE, our main metric for comparing the “expressive capacity” of different regulatory scenarios. To compute the GEE, we begin by generating N_T random target gene expression patterns $\mathbf{t} = \{t_i\}$ for genes $i = 1, \dots, M$ according to two rules: 1) all genes in the same cluster (i.e., sharing a multitarget factor) are randomly chosen to either not express (OFF, $t_i = 0$) or express (ON, $t_i > 0$), such that the number of ON clusters is held fixed; 2) each gene within an ON cluster is independently assigned a random nonnegative expression level from a uniform distribution. Importantly, this assignment of expression levels implies that genes belonging to the same ON cluster do not necessarily share the same target expression level (while all genes belonging to an OFF cluster will share a target expression level of zero). Thus, target

expression levels across all genes follow a “spike and slab” distribution (Fig. 1C), qualitatively recapitulating observations in early embryos (69). For each scenario (free DNA or chromatin) and each target pattern, the concentrations \mathbf{c} of regulatory factors (PFs and TFs) are optimized to reduce the root mean square error between target patterns \mathbf{t} and expression levels \mathbf{g} realized by our model. The residual RMS error averaged over many target patterns finally yields the GEE (Fig. 1D):

$$\text{GEE} = \frac{1}{N_T} \sum_{j=1}^{N_T} \left(\min_{\mathbf{c}} \sqrt{\frac{1}{M} \|\mathbf{g}(\mathbf{c}) - \mathbf{t}_j\|^2} \right). \quad [1]$$

Simulation details that define the regulatory network (number of clusters, number of genes per cluster, etc.) as well as the baseline parameter values are reported in *Materials and Methods*, with further details given in *SI Appendix*.

Results

Chromatin Outperforms Free DNA in Expressive Capacity. We optimized the concentrations of all regulatory factors for each target gene expression pattern to minimize GEE (Eq. 1), both in the chromatin as well as the free DNA scenario. At high intrinsic specificity of the factors ($S \gg 10^3$), target patterns are achieved with high accuracy in both scenarios. Yet as S decreases, the free DNA scenario suffers visibly larger deviations from target expression levels compared to chromatin (Fig. 2A). This is reflected in the respective GEEs of the two scenarios (Fig. 2B), where chromatin outperforms free DNA for every S , with highest

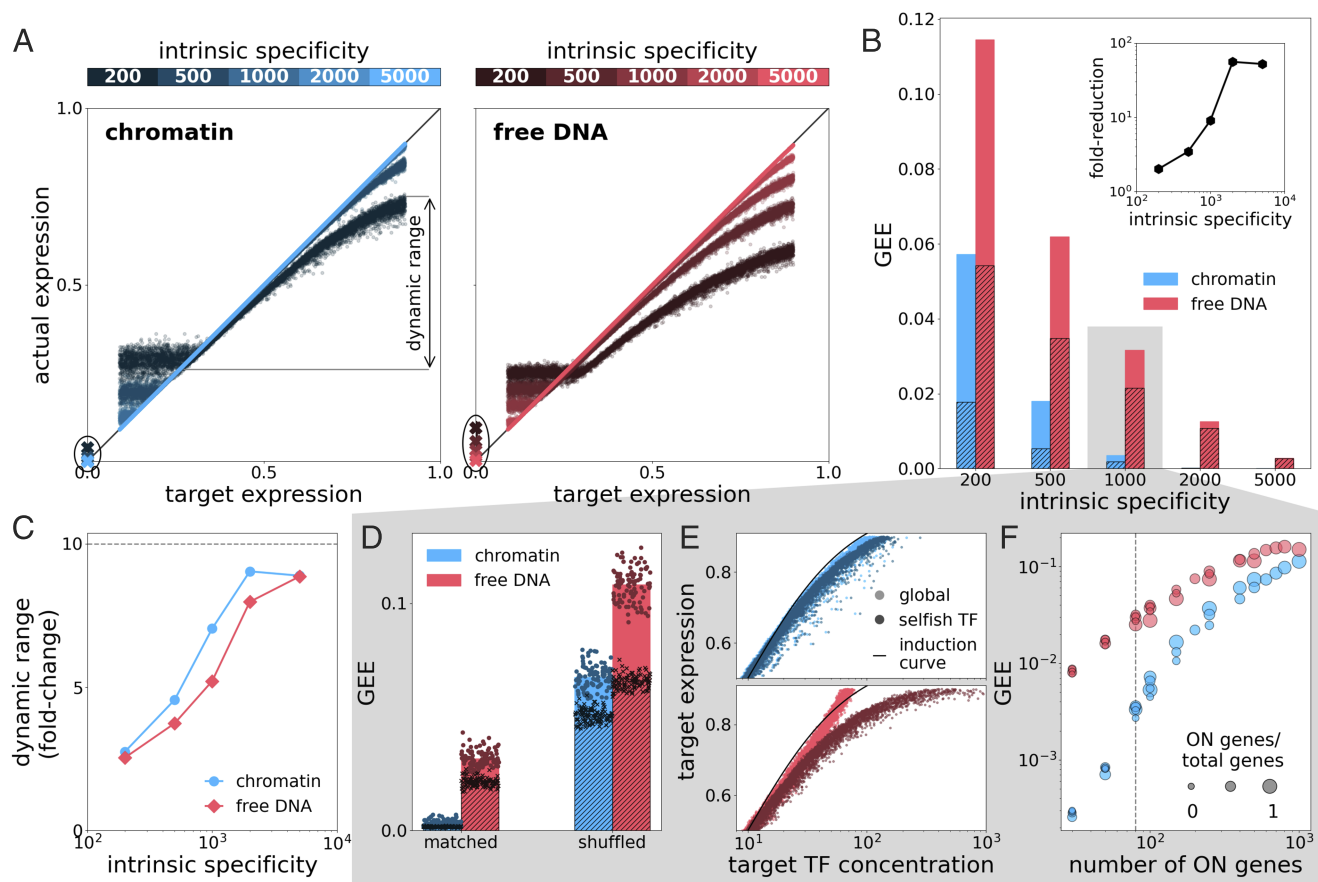


Fig. 2. Chromatin achieves smaller GEE compared to free DNA. (A) Actual vs. target expression levels after optimization of regulatory factor concentrations for chromatin (blue, *Left*) and free DNA (red, *Right*) for different intrinsic specificities S (color bar, *Top*). The displayed range for S corresponds to binding energy differences between target and nontarget binding of $5.3k_B T$ – $8.5k_B T$ (see *SI Appendix, Fig. S3* for results over a larger range). Points are ON genes in individual target patterns; crosses are OFF genes. Compared to free DNA, chromatin achieves expression levels for OFF genes much closer to zero (circled). (B) GEE is lower for chromatin (blue) than free DNA (red). Hatching indicates the GEE fraction due to erroneous expression of OFF genes. *Inset*: GEE fold-reduction for chromatin vs. free DNA. (C) Dynamic range (denoted in A) is greater for chromatin than free DNA. The dashed line indicates the maximal target dynamic range for this simulation. (D) GEE is lower when genes in the same cluster are either ON or OFF together (“matched”; *Left*) compared to when OFF genes are randomly assigned to clusters (“shuffled”; *Right*). Points are individual target patterns; bars are as in B. (E) Chromatin decouples regulatory control over ON genes. Light dots are individual gene expression levels achieved via global optimization over all PF and TF concentrations. Dark dots are optimization over individual single-target TF concentrations (see text). Ideal single-gene induction curve shown in black. (F) GEE is primarily determined by the number of ON genes rather than the absolute genome size or fraction of active genes (circle sizes) (*SI Appendix, Fig. S4*). Dashed line indicates number of ON genes used for analyses in the other subpanels.

(10- to 50-fold) reductions in error seen at realistic values of $S \gtrsim 10^3$. Similar results hold for larger simulated networks and across a wider range of intrinsic specificity S (*SI Appendix, Fig. S3*). To put this into perspective, the intrinsic specificity of regulatory factors in the free DNA scenario would have to be ~ 5 to $10\times$ as high compared to the chromatin scenario to become competitive, corresponding to an increase in binding energy difference between a cognate and noncognate site of $\sim 2k_B T$. While this appears small, such a systematic increase across all factors would be a significant fraction of the entire energy difference between binding a cognate and a noncognate site for metazoan TFs ($\sim 8k_B T$, corresponding to $S \sim 3 \cdot 10^3$, or roughly 12 bits of binding motif information) (3, 6).

Detailed analysis reveals that two effects predominantly contribute to the substantial improvements in expressive capacity of chromatin over free DNA. First, chromatin much more effectively silences the genes that are supposed to be OFF, i.e., their expression can be regulated closer toward the 0 target (Fig. 2A, circled crosses); this is reflected in a lower contribution to GEE from the OFF genes (Fig. 2B, hatched) compared to free DNA. Second, while regulators in both free

DNA and chromatin have difficulty inducing genes toward high expression levels, the failure at limited intrinsic specificity is much more pronounced in the free DNA scenario. This is reflected in its significantly smaller dynamic range of regulation—the mean difference between the highest and lowest achievable expression for the ON genes—compared to chromatin (Fig. 2C).

Intuitively, the benefits of chromatin silencing should be maximized when genes are grouped into clusters that turn ON or OFF together (i.e., ON/OFF state is “matched” to the cluster), such that each cluster is regulated by a dedicated PF and TFs are employed solely to fine-tune the precise expression levels of ON genes. Our analysis in Fig. 2D confirms this intuition: when clusters are matched to ON/OFF state, the GEE is threefold lower already for the free DNA scenario compared to its shuffled control, with the difference growing to more than 10-fold for chromatin. In the shuffled control, OFF genes can no longer be protected from unwanted expression: the need to express ON genes forces PFs to open chromatin, thereby exposing OFF genes in mixed clusters to regulatory influences from nontarget binding, which increases GEE. Therefore, for optimal expressive

capacity, genes belonging to the same cluster should share the same ON/OFF state.

Chromatin-based regulation has another important and biologically relevant benefit: it decouples and modularizes the regulation of ON genes. Consider the following. At global optimum, no gene will have perfectly achieved its target expression level. Suppose we choose a “selfish” gene that, starting from this global optimum, must reach its target expression level at any cost to other genes. The simplest way to regulate the selfish gene is to modulate the concentration of its corresponding single-target TF. Fig. 2E shows that, in the chromatin case, we need not change the single-target TF concentration very much to appease the selfish gene: the required change is in line with the effective optimal induction curve of individual genes. In contrast, in the free DNA scenario, the concentration of single-target TF must be significantly increased if the selfish gene is to reach a high target expression level, at a cost of increasing error due to nontarget binding elsewhere. Chromatin-based regulation mitigates this effect by leveraging a division of labor between PFs and TFs, where PFs keep OFF genes OFF nearly without fail, while each TF modularly changes the expression level of its target ON gene without much impact on other expression levels.

Owing to this division of labor between PFs and TFs, chromatin retains an absolute advantage in GEE across a broad range of regulatory network sizes. The error tends to track the number of ON genes and varies little with the number of OFF genes (Fig. 2F and SI Appendix, Fig. S4). For a fixed error, chromatin admits a much larger number of ON genes than free DNA, across two orders of magnitude in absolute genome size.

Thus, regulating gene expression through chromatin accessibility could conceivably support the evolution of larger genomes with more genes active simultaneously, even if the intrinsic specificity of binding factors remains fixed.

Optimal Chromatin-Based Regulation Co-opts Nontarget Binding. In all globally optimal solutions, genes with low target expression levels tend to be overexpressed relative to their targets, except at the highest specificities S that we consider (Fig. 2A). This excess expression is almost entirely due to the high prevalence of nontarget binding (Fig. 3A): already for ON genes with a middling desired expression level of 0.5, nontarget binding of single- and/or multitarget factors drives about 14% of expression in chromatin and 20% in the free-DNA scenario (Fig. 3B). Simply ignoring nontarget binding after optimization also yields significant increases in the GEE (SI Appendix, Fig. S2). These observations led us to wonder whether globally optimal solutions that minimize GEE also implicitly minimize nontarget binding and vice versa, or whether GEE minimization perhaps productively co-opts nontarget binding to achieve target expression levels.

To test the co-option hypothesis, we again turned to optimization. However, instead of optimizing factor concentrations toward matching target expression levels by any kind of binding, target or nontarget, as before, the new analysis only counts target binding as correct and all nontarget binding as erroneous regulation (SI Appendix, section S2.1). We find that this alternative strategy based on target binding alone does allow graded expression of ON genes, but systematically over- (under-) expresses genes with low (high) target levels (Fig. 3C). As a

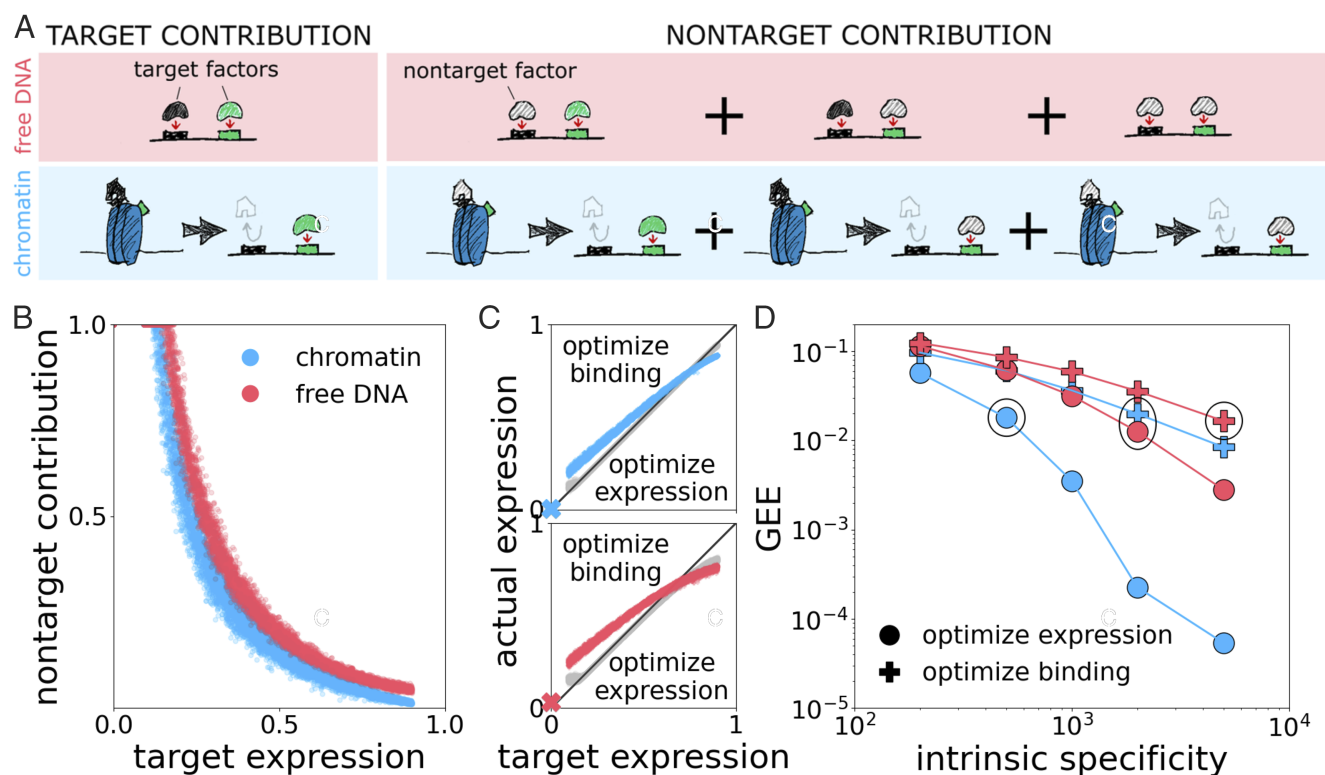


Fig. 3. Co-opting nontarget binding contributes substantially to gene expressive capacity. (A) Nontarget binding contributions are defined as instances of gene expression in which the binding of at least one factor is nontarget. (B) Fraction of expression level (vertical axis) attributable to nontarget vs. target binding for chromatin (blue) and free DNA (red). Dots are individual genes pooled across target patterns with $S = 1,000$. (C) Expression levels attained by optimizing to reduce GEE (as in Fig. 2, gray) compared to optimizing for target binding and against nontarget binding (color). (D) GEE is minimized when intrinsic specificity is high, chromatin is present, and nontarget binding is co-opted (blue circles). Nevertheless, multiple combinations of intrinsic specificity and regulatory scheme will often suffice to yield a very similar GEE (e.g., the circled points; see text).

result, the GEE is substantially increased relative to the original scheme across the entire range of intrinsic specificities S ; this effect is particularly pronounced in the chromatin scenario, where penalizing nontarget binding can increase GEE by up to two orders of magnitude (Fig. 3D). Thus, optimizing for globally accurate gene expression patterning is *not* synonymous with maximizing regulation by target factors and minimizing the influence of nontarget factors. Rather, when intrinsic specificity is limited, co-opting nontarget binding is the most effective way to minimize global gene expression error.

Multiple regulatory strategies can achieve comparable expressive capacity. For example, a chromatin scheme that co-opts nontarget binding at intrinsic specificity $S = 500$ has a similar GEE as 1) a free-DNA scheme with $S \approx 2,000$ that co-opts nontarget binding; 2) a chromatin scheme with $S \approx 2,000$ that penalizes nontarget binding; and 3) a free-DNA scheme with $S \approx 5,000$ that penalizes nontarget binding (Fig. 3D). This observation suggests a route by which chromatin might have alleviated the evolutionary pressure to maintain high intrinsic specificity: starting from the free DNA scenario, chromatin might first emerge due to selective forces not directly related to gene regulation [e.g., to protect DNA from damage (70)], then evolve to be used in gene silencing. Once chromatin was involved in regulating clusters, mutations in binding sequences leading to a decrease in intrinsic specificity (of up to an order of magnitude in this example) would be more likely to be tolerated and fixed without incurring any increase in GEE, while at the same time accommodating the “spike-and-slab” distribution of target gene expression levels at lower cost of selection (71).

Combining Activation with Repression Enhances the Benefits of Chromatin-Based Regulation. Our analysis thus far has assumed that all TFs behave as activators. We now relax this assumption and allow each gene to be regulated by both activators and repressors. Specifically, we stipulate that each gene is regulated by three factors: one multitarget factor, always an activator; one single-target activating TF; and one single-target repressing TF. Thus, if we have M genes, we now have $2M$ total single-target TFs. Many eukaryotic TFs possess binding domains separate from the activating or repressing domains, therefore we consider the activating or repressing behavior to be inherent to the factor, not to the site to which it is bound. The single-target activating and repressing TFs target separate sites, and nontarget binding is permitted across all TF sites independent of the nature of their targeting factor. The dissociation rates for target and nontarget binding remain the same for all factors. We adopt a simple model for repression whereby a single bound repressor completely counteracts any activation. Thus, in free DNA, a gene is expressing only if the multitarget factor site is bound, at least one activator (which could be a multitarget factor) is bound to a single-target TF site, and no site (including the multitarget factor site) is bound by a repressor (Fig. 4A). In chromatin, a gene expresses only if chromatin is permissive, at least one activator is bound, and no repressors are bound.

We reasoned that adding repressors to either scenario may serve to counteract the overexpression of lowly expressing genes due to nontarget binding. To ensure such an effect would be clearly visible in our simulation results, we lowered the minimum target expression level for ON genes from 0.09 to 0.009, thereby extending the target dynamic range by an order of magnitude, from 10-fold analyzed in Fig. 2, to 100-fold.

Simulations for this 100-fold modulation task clearly show that adding repressors indeed improves the accuracy of expression in the low-expression regime for both chromatin and free DNA (Fig. 4B). While adding repressors improves the dynamic range for both free DNA and chromatin, it decreases GEE only in the chromatin scenario, and actually *increases* the GEE in the free DNA scenario (Fig. 4C). As suggested by previous analyses (9), this counterintuitive effect likely emerges due to the fact that adding new TFs requires new binding sites, which introduce new opportunities for nontarget interactions that limit expressive capacity. With the exception of the lowest tested intrinsic specificity ($S = 200$), chromatin compensates for this deleterious expansion of nontarget interactions by the orthogonal PF regulation and the PFs’ proofreading capacity. Thus, the opposing effects of adding repressor-mediated regulation to free DNA further widen the expressive capacity advantage for the chromatin scenario (Fig. 4C).

To understand how repressors contribute to improved regulation, we plotted the optimal concentrations of single-target regulatory factors for each gene at each target expression level (Fig. 4D). We observe a clear division of labor between activators and repressors: genes with high expression levels are regulated almost entirely by activators, while genes with low expression levels are regulated almost entirely by repressors. This observation implies that repressors act almost exclusively to counteract nontarget activation. The “crossover point” between regulation by repressors vs. regulation by activators—that is, the target expression level at which both repressor and activator concentration are nearly 0—indicates the baseline (leaky) expression level established by the combined nontarget binding activity of repressors and activators at a gene locus (Fig. 4D). Subsequent modulation of specific repressor or activator concentrations therefore serves to tune the expression away from this baseline to the desired level for each target gene.

Last, we explored the ability of the two regulatory scenarios to generate distributions of target expression levels for ON genes that differ from uniform, which has been our focus until now. When we bias this target distribution away from uniform to favor either low or high values of gene expression, the GEE in the chromatin scenario barely changes and stays very low, while it varies substantially and is generally much higher in the free DNA scenario (Fig. 4E). Thus, regulation that combines activation and repression in the chromatin scenario increases the flexibility of the regulatory apparatus: in addition to extending the target dynamic range, it can accommodate more general variations in the desired distribution of expression levels.

Robustness to Regulatory Factor Concentration Fluctuations.

In living cells, stochastic fluctuations in regulator concentrations may introduce variation in downstream gene expression levels. To examine the effect of regulator variability on expressive capacity, we start from the optimized solutions developed in the previous sections and incorporate random variation into the optimized regulator concentrations, then calculate the resulting GEE. Specifically, for each globally optimized solution that generates a target gene expression pattern, we add multiplicative noise to each optimal factor concentration c^* of regulatory factor as

$$c' = c^* (1 + \text{Normal}(0, \sigma^2)), \quad [2]$$

such that c' is normally distributed at c^* with SD $c^*\sigma$. In this way, single- and multitarget factors are perturbed relative to

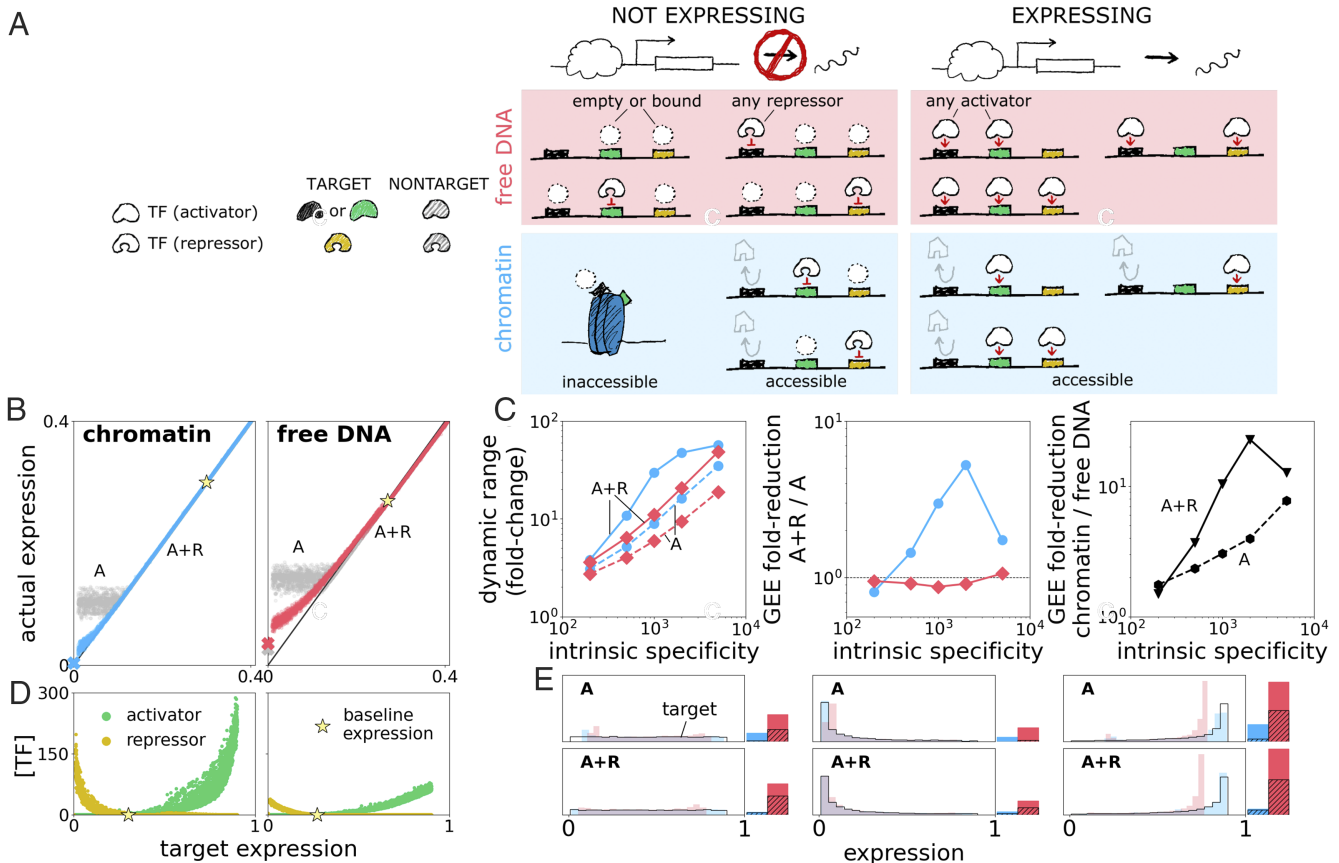


Fig. 4. Adding repressors boosts the expressive capacity advantage of chromatin over free DNA. (A) A schematic of bound factor combinations that do or do not lead to gene expression. (B) Target vs. actual expression for simulations with activators only (as in Fig. 2, “A,” gray) and with repressors (“A+R,” color) show that repressors improve accuracy at low target expression levels. (C) Repressors improve the expressive capacity of chromatin but have mixed effects on free DNA. *Left*, the dynamic range under both scenarios is larger when repressors are present (A+R, solid) than when they are not (A, dashed), but increases more for chromatin (blue) than for free DNA (red). *Middle*, the fold-reduction in error when using activators and repressors vs. the activators only is typically large for chromatin (blue) but may even lead to a GEE increase for free DNA (red). *Right*, the fold-reduction in error upon switching to chromatin is greater when repressors are present (A+R, solid) than when only activators are used for regulation (A, dashed). (D) Concentrations of single-target activator (green) and single-target repressor (gold) vs. target expression level of the corresponding gene. Both concentrations reach 0 around the same target expression level, or “baseline,” where nontarget binding by activators and repressors fully accounts for expression in the corresponding genes. Expression of individual genes is modulated above this level by activators, and below it by repressors. (E) A regulatory architecture employing both activators and repressors (A+R, bottom three panels, vs. activator-only architecture A, top three panels) enables chromatin scenario (blue) to accommodate various target distributions of expression levels (black outlined distributions) with minimal GEE, compared to free DNA (red), whose error is significantly higher and strongly dependent on the desired gene expression level distribution. GEE with concentrations optimally adjusted to each target distribution is shown at right for each distribution, following plotting conventions of Fig. 2B. Histograms are plotted over at least 3,750 target gene expression levels.

their individual optima and we can vary the fluctuation level by changing the parameter σ (see *SI Appendix, section S2.3* for details).

Fig. 5 shows the results for fluctuations introduced to the 100-fold modulation task with a uniform distribution of target ON expression levels when both activators and repressors are present. As expected, larger fluctuations significantly decrease the accuracy of regulation and thus increase the GEE (Fig. 5A, *Top*). While chromatin still allows for smaller expression errors, it is affected more strongly by concentration fluctuations at high intrinsic specificity of regulatory factors. Surprisingly, we find that there is an intermediate level of specificity at which chromatin maximizes its advantage over free DNA (Fig. 5A, *Bottom*).

Do multitarget or single-target factors contribute more to the degradation in accuracy upon introducing fluctuations? We added fluctuations either only to multitarget or only to single-target factors and compared the GEE between the two cases. Fig. 5B shows that, across most of the specificity range, fluctuations in single-target factor concentrations result in larger expression errors than fluctuations in multitarget factor

concentrations (see also *SI Appendix, Fig. S8*). A detailed analysis reveals the reason for this observation (*SI Appendix, section S2.3*). In optimal solutions, the multitarget factors are generally tasked with either keeping the genes entirely OFF or permitting expression. This means that the multitarget factors (and especially the PF system in chromatin scenario) operate closer to saturation, where fluctuations in the factor concentration do not propagate to gene expression. In contrast, the single-target TF system in both scenarios needs to fine-tune the expression levels of ON genes and is thereby more sensitive to fluctuations (*SI Appendix, Figs. S6 and S7*).

Taken together, our results suggest that chromatin is well suited when selection for extremely reliable ON/OFF control over gene expression is essential. Precise expression control is possible but seems to require a corresponding precision in TF expression levels. As a future research direction, we note that the concentrations themselves could have been optimized in the presence of fluctuations (rather than in the absence, as in the setup we report here), possibly leading to an identification of noise-robust chromatin regulatory regimes.

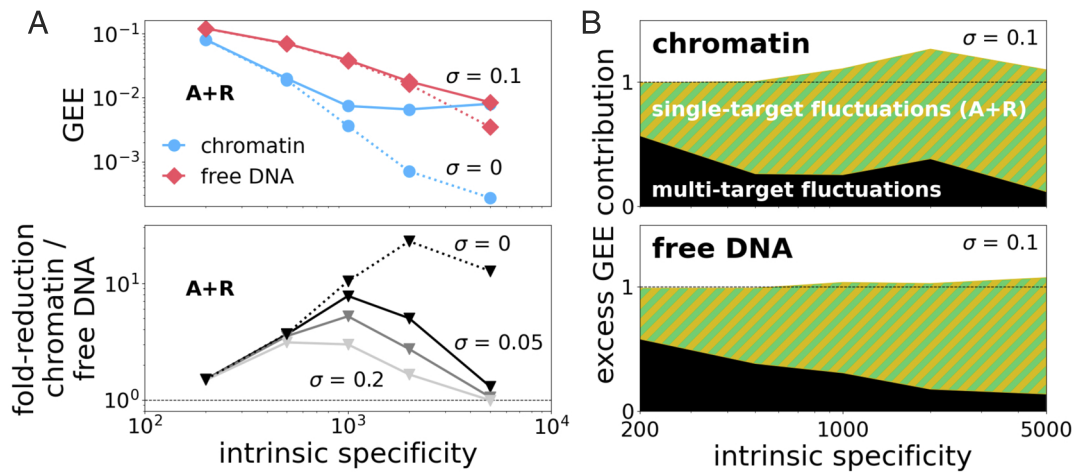


Fig. 5. Robustness to regulatory factor concentration fluctuations. (A) Adding fluctuations to all single-target and multitarget factor concentrations deteriorates GEE (Top) at higher levels of intrinsic specificity, more dramatically so for chromatin (blue) than for free DNA (red). Increasing fluctuation strength σ shifts the peak fold-reduction in GEE of chromatin vs. free DNA toward lower levels of intrinsic specificity (Bottom; original noise-free results shown in dashed line). (B) Relative increase in GEE due to fluctuations can be attributed primarily to single-target factors. Area plots show GEE when only single-target factors (striped) or only multitarget factors (black) fluctuate with $\sigma = 0.1$, normalized to the GEE when all factors fluctuate simultaneously (dashed line). Results for all panels are shown for the 100-fold modulation task using activators and repressors (A+R), with uniform distribution of target ON expression levels.

Discussion

Chromatin is often regarded as a central innovation in eukaryotes, but much remains to be understood concerning its implications for the global structure and strategy of gene regulatory programs. Here, we argue that chromatin, when understood as an integral and active component of gene regulation, allows for a substantially more accurate and scalable control of gene expression than is possible using TFs alone.

We motivated our study with the paradoxical observation that eukaryotes generally have much longer genomes than prokaryotes, but binding sites for TFs in metazoans appear to be less specific than in prokaryotes (12). This lack of specificity could lead to highly deleterious “crosstalk” in metazoan gene regulation: theory suggests that cooperative or combinatorial regulation schemes operating at equilibrium might be insufficient to ensure reliable expression control (9). Changes to effective genome size due to chromatinized sequence inaccessibility also cannot fully explain this paradox (6). Its possible resolution is a two-layer gene regulatory architecture that we propose in our work. The first layer of the architecture, namely proofreading-based modulation of chromatin accessibility, sets the discrete, ON/OFF state of the gene; the precise expression levels of ON genes are subsequently quantitatively modulated by activator and repressor TFs, acting as the second layer of the architecture. This paradigm of an “on-off switch followed by a rheostat” is well suited to generate the “spike-and-slab” distribution of expression levels across genes (Fig. 1C), with strong clustering of OFF genes across cell types and states. It is also consistent with recent experimental results showing that random stretches of DNA inserted into the genome are significantly more likely to express a reporter if a known PF binding site is included in the sequence (49).

An essential part of our approach is a global, systems-level analysis of end-to-end regulatory performance of both layers. This reveals a key insight responsible for a large boost in performance of chromatin-based regulation: Because OFF genes can be kept reliably silent via proofreading-based PF regulation, the regulatory system can indeed operate with TFs of lesser specificity and productively co-opt their nontarget binding. This sets a leaky baseline expression level for all genes that

targeted TF binding then up- or down- modulates to a precise desired level. Without reliable silencing of expression by PFs via chromatin, such co-opting necessarily leads to a possibly deleterious unwanted induction of OFF genes, as can be seen in our free DNA scenario.

A possible evolutionary interpretation of our results identifies the statistical structure of the target expression patterns as the key factor that ultimately shapes the regulatory network topology and architecture. By statistical structure, we mean any regularities that various gene expression levels should fulfill during the course of normal cell function: which genes need to be coexpressed, coregulated, or silenced together across all cell states and types. In our simulation, this statistical structure is made explicit by positing that target expression levels are drawn from a distribution where some genes are OFF and some ON but graded; as we find, this favors the two-layer regulation with chromatin. Specifically, the correlation in ON/OFF state between genes in the same cluster is captured in our model by shared regulation via multitarget factors. This correspondence between the statistics of the desired target expression patterns and a matching optimal regulatory architecture could be taken further. For example, if exact expression levels of the ON genes were highly correlated among certain genes, we hypothesize that allowing those genes to share a “single-target” TF would likely lead to a cost-efficient way to further lower the GEE, compared to a regulatory architecture that maintains a separate single-target factor for each gene. This reasoning would apply in the case where genes tend to express at similar levels to each other, but also in the situation where some genes are never ON simultaneously, as the single-target TF could then effectively be reused in different clusters. Some experimental evidence does in fact suggest that single TFs often activate mutually exclusive sets of genes in different cell types (72), implying that these genes belong to clusters that never turn ON together. Similarly, if some genes were expressed on average at higher levels than other genes, or with a larger dynamic range, we might imagine that the matching optimal regulatory architecture would involve systematic variation in intrinsic specificity as a third “layer” in addition to free DNA and chromatin. Indeed, this seems extremely likely to occur in nature given that average gene expression levels may vary by four orders

of magnitude across genes in the same organism under different experimental conditions (73). On an abstract level, this matching between an optimal gene regulatory architecture and the statistics of the desired target patterns that must be generated is highly reminiscent of the matching between the optimal structure of sensory systems and the statistical structure of incoming stimuli that must be encoded, as posited by the “efficient coding” theory in neuroscience (74).

Our conclusions are largely predicated on chromatin’s ability to silence genes and to proofread PF binding before opening. We propose two experiments to test these assumptions. First, to check whether proofreading is indeed responsible for keeping gene expression levels OFF, one could mutate a binding site for a known PF to lower its intrinsic specificity, then see how effectively increasing the concentration of the corresponding PF leads to chromatin opening and expression. In the absence of proofreading, higher concentration will necessarily correlate with a higher degree of chromatin accessibility; with proofreading, this is not the case if the binding site is weak; even with higher concentration, binding to weaker sites can be “rejected” as erroneous (75, 76). Second, to evaluate the effectiveness of silencing, one could introduce synthetic activating TFs targeted to endogenous regulatory regions that are active in some tissues and inactive in others due to chromatin conformation, e.g., the *eve* or *rho* enhancers in embryonic *Drosophila* (77). Quantitative measurements of differential expression across tissues could then indicate how much activation can be induced by TF binding in open vs. closed regions.

Our qualitative conclusions concerning the benefits of chromatin are not specific to a particular molecular pathway, but apply broadly to any kinetic proofreading mechanism whereby increasing concentrations of regulatory factors favor transcriptionally permissive states. The parameter values we selected for our simulations were based on the kinetics of nucleosome repositioning in response to remodeler recruitment. However, sensitivity analysis shows these values can be further tuned to optimize chromatin’s advantages over free DNA (*SI Appendix, Fig. S5*). Such values might be achieved by alternative mechanisms to the exact one we consider: higher levels of chromatin organization also show dose-dependent responses to regulator concentrations commensurate with our key assumptions. For example, experiments and modeling indicate that compacted chromatin can decompact with increasing concentrations of targeted activator (42, 43)—though whether compacted chromatin is truly less accessible than noncompacted chromatin is a subject of continuing debate (40, 78). Absent from our model is explicit regulation of chromatin closing or the maintenance of transcriptionally repressive states, for which experimental evidence suggests repressor complexes such as Polycomb may be essential (36, 79–81). As in the case of TF repressors, we expect these elements to change the quantitative but not the qualitative conclusions of the study.

We chose to focus on the accuracy of gene expression as the fundamental problem of regulation. In real cells, however, additional concerns such as energy consumption (which is higher for chromatin remodeling) and resource allocation (e.g., of polymerases) would also be expected to exert evolutionary pressure on regulatory systems. Dynamics are an area of particular relevance to the question of chromatin regulation, as the potential mechanisms (discussed above) differ substantially in their rates: chromatin decompaction occurs on a timescale that is rapid compared to the length of a cell cycle (1.5 h vs. 20 h) (43), but it is still orders of magnitude slower than nucleosome repositioning (on the order of s or min). Similarly, TF binding

and unbinding events as well as transcription initiation are faster than most chromatin remodeling processes (82). Therefore, in systems where speed is more important than accuracy in absolute expression levels, we may expect TFs to remain in use despite the reduced global error facilitated by chromatin-based regulation.

Because we evaluate expression levels at steady state, our results most directly apply to genes that must stably maintain a constant expression level over a long timespan relative to the rates of transcription and protein decay. The most salient biological analogs include “housekeeping” genes that carry out basal functions in all cells, as well as tissue- or cell-type-specific genes that show little cell-to-cell or temporal variation in transcript levels (83–85). In addition to constant expression, three features make stably expressed genes (SEGs) particularly suitable for analysis with our paradigm. First, SEGs vary in the absolute expression levels they maintain, with one analysis revealing a three- to fourfold difference between the 25th and 75th percentiles of expression, with an absolute dynamic range of >50-fold overall (86). Researchers have also hypothesized means for graded control of housekeeping genes specifically, for example, through the activity of corepressors (87). Second, chromatin remodeling may silence some sets of SEGs during the development of particular cell lineages (88). Third, experimental evidence from embryonic fruit flies suggests that housekeeping genes (expressed in all cells) rely on chromatin remodelers to position nucleosomes downstream from transcriptional start sites, rather than to regulate which TFs can bind. In contrast, developmental genes rely on a different set of remodelers to maintain accessibility at enhancers (89). Thus, housekeeping genes and cell-type-specific SEGs may offer a natural comparison between two alternative regulatory schemes, one of which (housekeeping) more closely resembles the free DNA scenario, and the other of which (developmental) more closely resembles the chromatin scenario.

Whether in chromatin or free DNA, an essential feature of our model is nontarget binding by regulatory factors, i.e., binding weakly to enhancers that do not regulate the factors’ central, “functional” targets. The extent to which weakly binding factors are essential, as opposed to merely consequential, for transcriptional regulation remains a subject of open discussion (10, 13, 90–92). Evidence for biologically significant weak binding emerges from studies on TF co-occupancy at enhancers, which show high prevalence in regions of accessible chromatin even between TFs with divergent binding site preferences and disparate functions (93). Weak but functional regulatory interactions have also been inferred in developmental networks, which appear to rely to a high degree on “noncanonical” TF binding interactions to establish cell type identity (94).

Our model suggests that genes may cope with unwanted weak binding by co-opting it to establish baseline expression levels for genes. Though the leaky expression in our model arises from uniform low-rate binding of nontarget TFs to enhancer binding sites, in practice TFs will vary along a continuum of specificity in their binding to different DNA sequences. Furthermore, it is likely that molecular steps between TF binding and transcription initiation can filter out signals emerging from binding of sufficiently low specificity (15, 75, 76) or of TFs lacking the appropriate cofactors. One prediction from these observations is that individual genes may establish baseline levels from subsets of the TFs that weakly bind their enhancer sequences, leading to a continuous distribution of baseline expression levels across genes. At least some data do indicate such distributions for genes expressed at less than one transcript per cell (i.e., for putatively OFF genes) (95), although we cannot rule out either

natural stochasticity or binding of specific TFs in erroneously open chromatin as the cause.

How does the notion of baseline expression we propose relate to previous discussions of biologically significant weak binding? We postulate that whether a given weak interaction is functional may be a more nuanced question than whether large changes to TF concentration noticeably modulate a gene's expression. In particular, some weak interactions may be *interchangeable*, in the sense that it may not matter to an enhancer which weakly binding TFs are present so long as some are. Such agnosticism to TF identity could conceivably improve the robustness of expression to uncorrelated local variations in the concentrations of weakly binding TFs.

An example of interchangeability comes from evidence that collectives of TFs can compete with nucleosomes to bind DNA and potentially activate genes (59). In particular, genomic analysis of yeast indicates that regions of DNA predicted to have high levels of nonspecific DNA binding are depleted of nucleosomes, including at the majority of promoters (96). Furthermore, DNA footprinting studies in mouse enhancers suggest simultaneous binding at regions of nucleosome competition is largely independent of TF identity and generally does not require direct TF–TF interaction (97). It would be interesting to couple such assays with measurements of the resulting gene expression driven by enhancers of interest, particularly if their strongest-binding TFs could be knocked out. Under a hypothesis of baseline expression driven by weak binding, we would expect identity-independent binding to induce some level of residual regulatory activity after knockout. A caveat to this approach is that it would not detect situations in which the activity of interchangeable weakly binding TFs is negligible or absent unless specific TFs are also present, for example, if the specific TF is an activator and the weakly binding TFs are repressors.

Nontarget binding may be co-opted or tolerated when relatively few genes are actively regulated. As more enhancers and corresponding genes become accessible, however, the unwanted effects of nontarget binding become harder to avoid. Some stem cells, for example, express genes specific to multiple possible differentiated lineages simultaneously, without performing the functions of cells belonging to these lineages (98). Such promiscuous expression might arise simply through the accessibility of regulating enhancers even if their specific binding factors are absent. Interestingly, many cancers also express embryonic stem cell markers or pluripotency factors, including PFs like OCT4 (99, 100). This may help to explain some of the phenotypic plasticity of cancerous cells as follows (101): suppose a differentiated cell escapes into a cancerous state simultaneously with the erroneous expression of PFs. At the moment of escape, the regulatory concentrations of TFs present in the cell were optimized for the differentiated cell—that is, in an environment

in which the appropriate genes were silenced. If these genes are suddenly made accessible to regulation, we might expect a higher probability of spuriously activating an unwanted transcriptional program than in an endogenous stem cell, where the regulatory concentrations would be optimized assuming these genes were accessible. Indeed, stem cells may maintain appropriate function despite susceptibility to high levels of nontarget binding due to a reliance on signaling factors secreted by other stem cells; the tumor-like behavior of cancer cells can be suppressed by these same signaling factors (102).

Chromatin may not have emerged for the purpose of regulating genes, but subsequent evolution has dramatically expanded its capacity to do so. With our work, we have aimed to provide comparative, quantitative arguments to help interpret the qualitative differences in regulatory schemes observed between prokaryotes and their eukaryotic descendants. It is our hope that further research will eventually explain, not merely the varying capacities of these genetic regulatory systems, but how their interaction with evolutionary pressures eventually gave rise to the complex, multicellular organisms that have so long excited the scientific imagination.

Materials and Methods

We use numerical optimization to minimize GEE defined in Eq. 1 with respect to all regulatory factor concentrations. Unless stated otherwise, we report results averaged over sets of between $N_T = 60$ and $N_T = 100$ target patterns that are shared between the “free DNA” and “chromatin” scenarios, for fixed intrinsic specificity $S = 1,000$, with the number of genes M fixed at 250. Multitarget factors each regulate 10 genes, and the number of ON clusters is fixed at 8 (80 active genes). We commonly analyze 5-fold variation around $S = 1,000$ to visualize trends. We start with simulations where the expression levels of ON genes should vary across a 10-fold range, on par with in-gene expression variation observed across a variety of experimental conditions in yeast (73), and consider larger expression range later. The remaining parameters can be found in *SI Appendix, Table S1*. Mathematical description of the entire model and the optimization procedure can be found in *SI Appendix*.

Data, Materials, and Software Availability. Simulation and analysis code has been deposited in GitHub (https://github.com/officerredshirt/network_crosstalk) (103).

ACKNOWLEDGMENTS. M.L.P. was supported by the European Molecular Biology Laboratory (EMBL) Interdisciplinary Postdoc Programme (EIPOD4 fellowships), cofunded by Marie Skłodowska-Curie Actions (Grant Agreement No. 847543). J.C. and M.L.P. were supported by EMBL Core Funding and Theory@EMBL. This work is supported by European Research Council Grant DynaTrans (101118866) to G.T. We would like to thank the members of the J.C. and G.T. groups, especially Natalia Misunou, Michal Hledík, and Réka Borbély, for helpful feedback and discussion. We also thank EMBL IT Services for the use of high performance computing resources.

1. P. H. V. Hippel, A. Revzin, C. A. Gross, A. C. Wang, Non-specific DNA binding of genome regulating proteins as a biological control mechanism: 1. The lac operon: Equilibrium aspects. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4808–4812 (1974).
2. J. Elf, G. W. Li, X. S. Xie, Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
3. S. J. Maerkl, S. R. Quake, A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315**, 233–237 (2007).
4. F. M. Weinert, R. C. Brewster, M. Rydenfelt, R. Phillips, W. K. Kegel, Scaling of gene expression with transcription-factor fugacity. *Phys. Rev. Lett.* **113**, 258101 (2014).
5. E. E. Swinestead *et al.*, Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
6. Z. Wunderlich, L. A. Mirny, Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* **25**, 434–440 (2009).
7. R. S. Mann, K. M. Lelli, R. Joshi, “Chapter 3: Hox specificity” in *Current Topics in Developmental Biology*, O. Porquie, Ed. (Elsevier, 2009), vol. 88, pp. 63–101.
8. A. L. Todeschini, A. Georges, R. A. Veitia, Transcription factors: Specific DNA binding and specific gene regulation. *Trends Genet.* **30**, 211–219 (2014).
9. T. Friedlander, R. Prizak, C. C. Guet, N. H. Barton, G. Tkačik, Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.* **7**, 12307 (2016).
10. A. Tanay, Extensive low-affinity transcriptional interactions in the yeast genome. *Genome Res.* **16**, 962–972 (2006).
11. J. Crocker *et al.*, Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
12. J. F. Kribelbauer, C. Rastogi, H. J. Bussemaker, R. S. Mann, Low-affinity binding sites and the transcription factor specificity paradox in eukaryotes. *Annu. Rev. Cell Dev. Biol.* **35**, 357–379 (2019).
13. A. Shaheini *et al.*, Systematic analysis of low-affinity transcription factor binding site clusters in vitro and in vivo establishes their functional relevance. *Nat. Commun.* **13**, 5273 (2022).
14. J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4135–4139 (1974).

15. S. A. Cepeda-Humerez, G. Rieckh, G. Tkačik, Stochastic proofreading mechanism alleviates crosstalk in transcriptional regulation. *Phys. Rev. Lett.* **115**, 248101 (2015).
16. H. Schiessel, R. Blossey, Pioneer transcription factors in chromatin remodeling: The kinetic proofreading view. *Phys. Rev. E* **101**, 040401 (2020).
17. N. C. Lammers, A. I. Flaholz, H. G. Garcia, Competing constraints shape the nonequilibrium limits of cellular decision-making. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2211203120 (2023).
18. H. Boeger, Kinetic proofreading. *Annu. Rev. Biochem.* **91**, 423–447 (2022).
19. S. Balaji, M. M. Babu, L. M. Iyer, N. M. Luscombe, L. Aravind, Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.* **360**, 213–227 (2006).
20. V. Charoensawan, D. Wilson, S. A. Teichmann, Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* **38**, 7364–7377 (2010).
21. A. B. Georges, B. A. Benayoun, S. Caburet, R. A. Veitia, Generic binding sites, generic DNA-binding domains: Where does specific promoter recognition come from? *FASEB J.* **24**, 346–356 (2010).
22. F. Reiter, S. Wienerroither, A. Stark, Combinatorial function of transcription factors and cofactors. *Curr. Opin. Genet. Dev.* **43**, 73–81 (2017).
23. X. Grau-Bové *et al.*, A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. *Nat. Ecol. Evol.* **6**, 1007–1023 (2022).
24. F. H. Lam, D. J. Steger, E. K. O'Shea, Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).
25. F. Zhu *et al.*, The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
26. B. Fierz, M. G. Poirier, Biophysics of chromatin dynamics. *Annu. Rev. Biophys.* **48**, 321–345 (2019).
27. Y. Lorch, J. W. LaPointe, R. D. Kornberg, Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**, 203–210 (1987).
28. K. Struhl, Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1–4 (1999).
29. H. Boeger, J. Griesenbeck, R. D. Kornberg, Nucleosome retention and the stochastic nature of promoter chromatin remodeling for transcription. *Cell* **133**, 716–726 (2008).
30. A. S. Hansen, E. K. O'Shea, Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.* **9**, 704 (2013).
31. T. Kujirai, H. Kurumizaka, Transcription through the nucleosome. *Curr. Opin. Struct. Biol.* **61**, 42–49 (2020).
32. E. Li, Y. Zhang, DNA methylation in mammals. *Cold Spring Harb. Perspect. Biol.* **6**, a019133 (2014).
33. G. E. Zentner, S. Henikoff, Regulation of nucleosome dynamics by histone modifications. *Nat. Struct. Mol. Biol.* **20**, 259–266 (2013).
34. J. S. Becker, D. Nicetto, K. S. Zaret, H3K9me3-dependent heterochromatin: Barrier to cell fate changes. *Trends Genet.* **32**, 29–41 (2016).
35. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
36. A. L. Hughes, J. R. Kelley, R. J. Klose, Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochim. Biophys. Acta (BBA), Gene Regul. Mech.* **1863**, 194567 (2020).
37. K. Maeshima *et al.*, The physical size of transcription factors is key to transcriptional regulation in chromatin domains. *J. Phys.: Condens. Matter* **27**, 064116 (2015).
38. K. S. Zaret, Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
39. A. Mota *et al.*, FRET-FISH probes chromatin compaction at individual genomic loci in single cells. *Nat. Commun.* **13**, 6680 (2022).
40. K. Maeshima, S. Iida, M. A. Shimazoe, S. Tamura, S. Ide, Is euchromatin really open in the cell? *Trends Cell Biol.* **34**, 7–17 (2023).
41. M. G. Poirier, E. Oh, H. S. Tims, J. Widom, Dynamics and function of compact nucleosome arrays. *Nat. Struct. Mol. Biol.* **16**, 938–944 (2009).
42. F. Erdel *et al.*, Mouse heterochromatin adopts digital compaction states without showing hallmarks of HP1-driven liquid-liquid phase separation. *Mol. Cell* **78**, 236–249.e7 (2020).
43. N. A. Pease *et al.*, Tunable, division-independent control of gene activation timing by a polycomb switch. *Cell Rep.* **34**, 108888 (2021).
44. G. O. Bryant *et al.*, Activator control of nucleosome occupancy in activation and repression of transcription. *PLoS Biol.* **6**, e317 (2008).
45. P. Trojer, D. Reinberg, Facultative heterochromatin: Is there a distinctive molecular signature? *Mol. Cell* **28**, 1–13 (2007).
46. A. Gaspar-Maia, A. Alajem, E. Meshorer, M. Ramalho-Santos, Open chromatin in pluripotency and reprogramming. *Nat. Rev. Mol. Cell Biol.* **12**, 36–47 (2011).
47. S. M. Foo *et al.*, Zeldin potentiates morphogen activity by increasing chromatin accessibility. *Curr. Biol.* **24**, 1341–1346 (2014).
48. G. Fernandes *et al.*, Synthetic reconstruction of the hunchback promoter specifies the role of Bicoid, Zeldin and Hunchback in the dynamics of its transcription. *eLife* **11**, e74509 (2022).
49. R. Galupa *et al.*, Enhancer architecture and chromatin accessibility constrain phenotypic space during *Drosophila* development. *Dev. Cell* **58**, 51–62.e4 (2023).
50. K. S. Zaret, J. S. Carroll, Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
51. J. L. Hansen, K. J. Loell, B. A. Cohen, A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife* **11**, e73358 (2022).
52. R. Blossey, H. Schiessel, Kinetic proofreading of gene activation by chromatin remodeling. *HFSP J.* **2**, 167–170 (2008).
53. R. Blossey, H. Schiessel, Kinetic proofreading in chromatin remodeling: The case of ISWI/ACF. *Biophys. J.* **101**, L30–L32 (2011).
54. G. Brysbaert, M. F. Lensink, R. Blossey, Regulatory motifs on ISWI chromatin remodelers: Molecular mechanisms and kinetic proofreading. *J. Phys.: Condens. Matter* **27**, 064108 (2015).
55. D. Porcelli, B. Fischer, S. Russell, R. White, Chromatin accessibility plays a key role in selective targeting of Hox proteins. *Genome Biol.* **20**, 115 (2019).
56. K. J. Polach, J. Widom, Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254**, 130–149 (1995).
57. T. Raveh-Sadka, M. Levo, E. Segal, Incorporating nucleosomes into thermodynamic models of transcription regulation. *Genome Res.* **19**, 1480–1496 (2009).
58. O. Igoshin, J. Narula, Thermodynamic models of combinatorial gene regulation by distant enhancers. *IEE Syst. Biol.* **4**, 393–408 (2010).
59. L. A. Mirny, Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 22534–22539 (2010).
60. L. Bintu *et al.*, Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720–724 (2016).
61. R. Shelansky, H. Boeger, Nucleosomal proofreading of activator-promoter interactions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2456–2461 (2020).
62. S. Bruno, R. J. Williams, D. Del Vecchio, Mathematical analysis of the limiting behaviors of a chromatin modification circuit. *Math. Control Signal. Syst.* **35**, 399–432 (2023).
63. M. Hledik, "Genetic information and biological optimization," PhD thesis, Institute of Science and Technology, Austria (2024).
64. H. A. Pliner *et al.*, Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol. Cell* **71**, 858–871.e8 (2018).
65. Y. A. Medvedeva *et al.*, EpiFactors: A comprehensive database of human epigenetic factors and complexes. *Database* **2015**, bav067 (2015).
66. S. A. Lambert *et al.*, The human transcription factors. *Cell* **172**, 650–665 (2018).
67. J. Cao *et al.*, Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* **361**, 1380–1385 (2018).
68. T. C. Voss, G. L. Hager, Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* **15**, 69–81 (2014).
69. R. J. White *et al.*, A high-resolution mRNA expression time course of embryonic development in zebrafish. *eLife* **6**, e30860 (2017).
70. I. A. Drinnenberg *et al.*, Towards a synthesis of chromatin biology and evolution. *Development* **146**, dev178962 (2019).
71. M. Hledik, N. Barton, G. Tkačik, Accumulation and maintenance of information in evolution. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2123152119 (2022).
72. A. Kamal *et al.*, GRaNE and GRaNP: Inference and evaluation of enhancer-mediated gene regulatory networks. *Mol. Syst. Biol.* **19**, e11627 (2023).
73. J. Zrimec *et al.*, Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. *Nat. Commun.* **11**, 6141 (2020).
74. M. Chalk, O. Marre, G. Tkačik, Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 186–191 (2018).
75. R. Grah, B. Zoller, G. Tkačik, Nonequilibrium models of optimal enhancer function. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31614–31622 (2020).
76. B. Zoller, T. Gregor, G. Tkačik, Eukaryotic gene regulation at equilibrium, or non? *Curr. Opin. Syst. Biol.* **31**, 100435 (2022).
77. J. Crocker, D. L. Stern, TALE-mediated modulation of transcriptional enhancers in vivo. *Nat. Methods* **10**, 762–767 (2013).
78. R. V. Chereji, P. R. Eriksson, J. Ocampo, H. K. Prajapati, D. J. Clark, Accessibility of promoter DNA is not the primary determinant of chromatin-mediated gene regulation. *Genome Res.* **29**, 1985–1995 (2019).
79. K. Sneppen, L. Ringrose, Theoretical analysis of Polycomb-Trithorax systems predicts that poised chromatin is bistable and not bivalent. *Nat. Commun.* **10**, 2133 (2019).
80. J. Reining, F. Ruge, M. Howard, L. Ringrose, A theoretical model of Polycomb/Trithorax action unites stable epigenetic memory and dynamic regulation. *Nat. Commun.* **11**, 4782 (2020).
81. N. P. Blackledge, R. J. Klose, The molecular principles of gene regulation by Polycomb repressive complexes. *Nat. Rev. Mol. Cell Biol.* **22**, 815–833 (2021).
82. N. C. Lammers, Y. J. Kim, J. Zhao, H. G. Garcia, A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting. *Curr. Opin. Cell Biol.* **67**, 147–157 (2020).
83. E. Eisenberg, E. Y. Levanon, Human housekeeping genes, revisited. *Trends Genet.* **29**, 569–574 (2013).
84. Y. Lin *et al.*, Evaluating stably expressed genes in single cells. *GigaScience* **8**, giz106 (2019).
85. L. Qiu *et al.*, Identification of human global, tissue and within-tissue cell-specific stably expressed genes at single-cell resolution. *Int. J. Mol. Sci.* **23**, 10214 (2022).
86. C. J. Joshi, W. Ke, A. Drangowska-Way, E. J. O'Rourke, N. E. Lewis, What are housekeeping genes? *PLoS Comput. Biol.* **18**, e1010295 (2022).
87. A. Mitra, A. M. Raicu, S. L. Hickey, L. A. Pile, D. N. Arnosti, Soft repression: Subtle transcriptional regulation with global impact. *BioEssays* **43**, 2000231 (2021).
88. L. Thorrez *et al.*, Tissue-specific disallowance of housekeeping genes: The other face of cell differentiation. *Genome Res.* **21**, 95–105 (2011).
89. O. Hendy *et al.*, Developmental and housekeeping transcriptional programs in *Drosophila* require distinct chromatin remodelers. *Mol. Cell* **82**, 3598–3612.e7 (2022).
90. M. D. Biggin, Animal transcription networks as highly connected, quantitative continua. *Dev. Cell* **21**, 611–626 (2011).
91. W. W. Fisher *et al.*, DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21330–21335 (2012).
92. J. Crocker, E. Preger-Ben Noon, D. L. Stern, "The soft touch" in *Current Topics in Developmental Biology*, P. M. Wassarman, Ed. (Elsevier, 2016), vol. 117, pp. 455–469.
93. S. MacArthur *et al.*, Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10**, R80 (2009).
94. A. R. Sonawane *et al.*, Understanding tissue-specific gene regulation. *Cell Rep.* **21**, 1077–1088 (2017).
95. D. Hebenstreit *et al.*, RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **7**, 497 (2011).
96. A. Afek, I. Sela, N. Musa-Lempel, D. B. Lukatsky, Nonspecific transcription-factor-DNA binding influences nucleosome occupancy in yeast. *Biophys. J.* **101**, 2465–2475 (2011).
97. C. Sönmez *et al.*, Molecular co-occupancy identifies transcription factor binding cooperativity in vivo. *Mol. Cell* **81**, 255–267.e6 (2021).
98. D. Zipori, The nature of stem cells: State rather than entity. *Nat. Rev. Genet.* **5**, 873–878 (2004).

99. M. Monk, C. Holding, Human embryonic genes re-expressed in cancer cells. *Oncogene* **20**, 8085-8091 (2001).
100. M. Schoenhals *et al.*, Embryonic stem cell markers expression in cancers. *Biochem. Biophys. Res. Commun.* **383**, 157-162 (2009).
101. E. Vlashi, F. Pajonk, Cancer stem cells, cancer cell plasticity and radiation therapy. *Semin. Cancer Biol.* **31**, 28-35 (2015).
102. L. M. Postovit *et al.*, Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4329-4334 (2008).
103. M. L. Perkins, J. Crocker, G. Tkačik, Computational code from "Chromatin enables precise and scalable gene regulation with factors of limited specificity." GitHub. https://github.com/officerredshirt/network_crosstalk. Accessed 8 December 2024.