Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Review

Long-term evolution of regulatory DNA sequences. Part 2: theory and future challenges

Elia Mascolo[1], Réka Borbély[1], Noa O Borst[2], Nicholas H Barton[1], Justin Crocker[2] and Gašper Tkačik[1]

Promoters and enhancers are *cis*-regulatory elements (CREs), DNA sequences that bind transcription factor (TF) proteins to up- or down-regulate target genes. Decades-long efforts yielded TF-DNA interaction models that predict how strongly an individual TF binds arbitrary DNA sequences and how individual binding events on the CRE combine to affect gene expression. These insights can be synthesized into a global, biophysically realistic, and quantitative genotype-phenotype (GP) map for gene regulation, a 'holy grail' for the application of evolutionary theory. A global map provides a rare opportunity to simulate the long-term evolution of regulatory sequences and pose several fundamental questions: How long does it take to evolve CREs *de novo*? How many non-trivial regulatory functions exist in sequence space? How connected are they? For which regulatory architecture is CRE evolution most rapid and evolvable? In this article, the second of a two-part series, we review the application of evolutionary concepts — epistasis, robustness, evolvability, tunability, plasticity, and bet-hedging — to the evolution of gene regulatory sequences. We then evaluate the potential for a unifying theory for the evolution of regulatory sequences and identify key open challenges.

Addresses
[1] Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg AT-3400, Austria
[2] Developmental Biology Unit, European Molecular Biology Laboratory, Heidelberg DE-69117, Germany

Corresponding author: Tkačik, Gašper (Gasper.Tkacik@ist.ac.at)

Current Opinion in Genetics & Development 2026, **98**:102472

This review comes from a themed issue on **Genome Architecture and Expression**

Edited by **Luca Giorgetti** and **Daniel Jost**

Available online xxxx

https://doi.org/10.1016/j.gde.2026.102472

0959–437X/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

In Part 1 of this review series [1], we argued that regulatory DNA allows us to reconstruct genotype-phenotype (GP) maps that are both *quantitative* and *global*. Here, 'global' means that the map assigns a regulatory phenotype to all 4^L possible sequences at a cis-regulatory element (CRE) of length L [1]. This assignment is conditioned on a given molecular context. Mutations in trans-elements can change this context, for example, by changing transcription factors' (TFs) binding preferences or expression profiles, thus reshaping the GP map itself. Therefore, trans-acting elements may be treated either as fixed parameters or, when appropriate, as additional evolving variables that can alter the GP map of the CRE.

We reasoned that global and quantitative GP maps for regulatory DNA offer the opportunity to advance beyond the trade-off between empirical (yet local and thus incomplete) maps and global (yet idealized and thus unrealistic) maps. They open the door to studying 'long-term evolution', by which we mean evolutionary trajectories that can start from arbitrary DNA sequences (including random non-functional sequences) and proceed through an indefinite number of mutations. We contrasted this with 'short-term evolution', which only accounts for a few mutations ($\lesssim 10$) around an observed wild-type for which phenotype or fitness measurements are available.

In Part 2 of this review series, we discuss how — in addition to biology and physics-informed modeling and simulations — a general *theory* for the long-term evolution of regulatory sequences and architecture may be within reach. By mapping genotypes into phenotypes, regulatory GP maps act as information-theoretic codes. We review recent developments that bridge information theory and evolutionary theory. These developments begin to address an overarching question: which *regulatory architectures* optimize theoretical objectives? When these architectures themselves can evolve, what do they evolve toward and under what conditions? For an illustration and description of the key notions of *regulatory architecture*, *function*, *grammar*, and *code*, we point the reader to the first Figure of Part 1 of this review series [1].

Epistasis, robustness, and evolvability for regulatory sequences

In this section, we briefly review the application of the general evolutionary concepts of epistasis, robustness, and evolvability to GP maps for *cis*-regulatory sequences. The motivation for this is twofold. First, because such

GP maps can be both quantitative and global, these evolutionary concepts can be elevated from qualitative descriptors to computable quantities. Second, as a consequence, these quantities can serve as key ingredients for the development of a more general theory, as we explain in Section II.

Epistasis refers to non-additive effects of individual mutations on phenotype or fitness. Because the biophysics of TF-DNA interactions necessitates a nonlinear mapping between the binding energy and the resulting site occupancy, a non-epistatic, additive expectation would be a straw-man null model; thermodynamic models provide a more appropriate baseline by capturing the 'global epistasis' in all CREs due to the sigmoid binding nonlinearity [2–4]. Identifying 'specific epistasis' [5–7], that is, the extra, site-specific epistatic interactions beyond the global ones induced by a single sigmoid, is far more challenging. Specific epistasis could arise because the loci within a single TF binding site contribute to binding energy non-additively [8–10], or because the occupancies of multiple TFs themselves interact non-linearly to control gene expression [11]. Empirical work provides some evidence that specific epistasis can generate rugged landscapes with multiple fitness peaks [12]. However, apparent ruggedness must be interpreted cautiously, as it can reflect noise due to statistical and technical errors (e.g. sampling noise, assay variability, batch effects, etc.) rather than true epistasis.

When classical toy models of epistasis ignore the sigmoidal transformation from binding energy to gene expression, even an additive system can masquerade as one with hopelessly complex high-order specific epistasis. Explicitly incorporating known physical constraints and the resulting global epistasis substantially reduces the apparent complexity of specific epistasis. This approach has been successfully applied to TFs and CREs, and is sufficient to explain most of the observed phenotypic variance [5,13,14].

Robustness enables the organism to maintain its function in the face of perturbations, and it is expected to be selectable insofar as it provides a mechanism for preserving fitness. Mutational robustness denotes robustness specifically to DNA mutations. In regulatory sequences, the phenotype to be maintained depends on the level of organization: for individual TF binding sites, it is the TF-binding affinity; for an entire promoter or enhancer, it is the *regulatory function*, that is, the dependency of gene expression on regulatory inputs [1]. Cooperativity and combinatorial control can make the regulatory function robust even when the affinity of individual binding sites is perturbed.

While in idealized asexual models, the equilibrium 'mutation load' (loss of fitness due to mutation) depends only on deleterious mutation rate [15], in the presence of epistasis and recombination, selection can favor more mutationally robust GP maps that buffer mutational effects [16]. For TF-DNA interactions, the intrinsic binding sigmoid ensures that at sufficiently high TF concentrations, sites with one or a few mismatches can still be occupied, buffering the immediate fitness consequences of point mutations and thereby increasing mutational robustness [17].

Beyond mutational robustness, gene regulation may also be selected for robustness to environmental and consequent physiological variability. In developmental contexts, the presence of multiple weaker enhancers could ensure proper spatio-temporal gene expression patterns even when the upstream TF concentrations are perturbed [18].

Evolvability is the ability of the system to generate phenotypic variation that is heritable and adaptive [19]. Robustness (maintaining existing function) and evolvability (accessing new functions) are often put into contrast, but [20] demonstrated that this need not be the case on a typical TF-DNA GP map, where the neighborhoods of large neutral networks in sequence space are flanked by regulatory sequences binding alternative TFs: this endows the landscapes with local robustness, while larger mutational distances suddenly unlock new phenotypes. These ideas have been applied to landscapes derived from TF-DNA interactions in bacteria [17] and eukaryotes [20–22]. More generally, fitness landscapes where large neutral plateaus connect domains with differing phenotypes can provide high evolvability because cryptic neutral variation enables the population to access various phenotypes without crossing fitness valleys [23]. Alternatively, when gene expression is viewed as a quantitative trait — as opposed to a binary, ON/OFF one — one can ask about the 'tunability' of a landscape, that is, the ability of mutations to fine-tune regulatory function. This property has been reported for both prokaryotic [17] and eukaryotic regulatory logic [21]. Moreover, the evolvability of any specific phenotype can depend strongly on the properties of the GP map. Reconstruction and experimental mapping of an ancient GP map of a TF-DNA regulatory module showed that some DNA specificities are intrinsically far more likely to arise by mutation, and that these biases shaped subsequent evolution of the regulatory module [24]. Both evolvability and robustness are therefore context-dependent properties, affected by the specific TF-CRE interactions involved [17].

Selection for evolvability and robustness might shape regulatory architecture. Evidence from mutagenesis during early metazoan development suggests that promoters are more mutationally robust, with most mutations affecting overall expression level but not their

essential function, whereas enhancer mutations readily alter spatiotemporal expression patterns [25,26]. If confirmed, such an 'evolutionary division of labor' would represent a candidate architectural strategy that jointly accommodates robustness and evolvability.

Evolvability is distinct from phenotypic plasticity. While evolvability implies the ability to nimbly adapt to environmental changes via genetic mutations, plasticity refers to non-genetic, yet adaptive (and sometimes rapid and quantitative) responses to environmental change. Plasticity is often used only descriptively, but can be made quantitative in the context of gene regulation, where it naturally maps to the already-introduced concept of a multi-dimensional 'regulatory phenotype': the set of 'environments' (represented by different combinations of TF concentrations) paired with their corresponding gene expression responses [14]. Regulation is therefore a form of plasticity where molecular mechanisms adjust gene expression to the environment via a 'regulatory function' (see Fig. 1B in [1]). In contrast, gene expression noise generates uncontrolled phenotypic variability and could plausibly be adaptive via 'bet hedging' [27], a strategy extensively studied in bacteria [28–30].

Toward an evolutionary theory of genetic regulatory architecture

The concepts introduced in the previous section can influence the evolution of regulatory architectures, but they are local by construction, being measured around given (often wild-type) sequences. As such, they naturally inform what we have referred to as short-term evolution. A general theory of long-term regulatory evolution must instead capture these notions using quantitative descriptors of the entire GP map. Such a 'general theory' would be a quantitative framework that links (i) biophysical constraints on molecular interactions, (ii) evolutionary dynamics in genotype space, and (iii) GP map properties. The objective is not a complete predictive description of gene regulation across all cellular contexts, but to understand which classes of regulatory solutions are biophysically possible, evolutionarily accessible, or selectively favored under realistic population-genetic regimes. With this perspective in mind, we now turn to what such a general theory should look like.

On the one hand, physics and control theory are essential to realistically model *how* regulatory responses are physically generated by molecular components. Regulatory phenotypes are not qualitative traits, but quantitative input-output mappings that describe how gene expression responds to environmental changes. Such mappings are constrained at the systems level by biophysical facts: all concurrently expressed TFs compete for binding to all accessible CREs, and due to limited TF-DNA specificity, any (even non-cognate) TF can bind any CRE, albeit with a sequence-dependent probability, giving rise to some level of unavoidable non-cognate, off-target binding (or 'crosstalk') [31]. Biophysical constraints further dictate that regulatory sequences map into regulatory phenotypes through nonlinear transformations (e.g. due to the sigmoid that maps binding energy onto TF occupancy), so that global epistasis is not an exception but a necessary baseline expectation. Within all these physical and structural constraints, the theory may suggest optimal regulatory strategies with respect to specified information-theoretic criteria, such as error minimization (reducing deviations from optimal gene-expression output) [32], rate-distortion optimization (balancing regulatory accuracy against molecular cost) [33], channel-capacity maximization (maximizing information transmission from regulatory inputs to gene-expression outputs) [34], or description-length minimization (minimizing the effective number of DNA bases devoted to encoding regulatory programs) [35]. Such optima should be understood not as predictions of realized biological solutions, but rather as benchmarks for rationalizing what should in principle be biophysically possible with existing molecular components, and what would be functionally desirable [36,37], regardless of whether such architectures can in fact be produced and maintained by natural selection.

On the other hand, population genetics imposes an entirely orthogonal set of evolutionary constraints: setting limits to how closely and rapidly any such regulatory architectures can actually be approached given available adaptation mechanisms, or maintained in the face of mutation and drift. Existing, well-developed theory for evolutionary dynamics [38] allows us to compute or simulate evolutionary outcomes on given GP maps for gene regulation that incorporate said biophysical constraints. At the same time, the quantitative language for expressing these limits and the resulting GP map properties must be provided via the measures of evolvability, robustness, tunability, etc., as discussed in the previous section.

Yet there is a wide epistemological divide between documenting, defining, and quantifying desirable properties of existing GP maps, and between suggesting a first-principles, predictive theory of what gene regulatory GP maps should look like and how they could evolve. There is indeed a venerable history of combining mechanistic constraints and evolutionary arguments to explain gene regulatory architectures; as an early example, we point to Savageau's demand rules formulated in the 1970s, and the follow-up work on the role of activation vs repression [39–41]. Nevertheless, one might argue that, due to the central role of historical contingency in evolution, we cannot hope for any overarching predictive

theory for regulatory architectures: we will always be confined to description, perhaps quantitative if we are lucky; yet forever barred from true prediction, even when it is only qualitative.

We strongly disagree. We also point to the similarly optimistic disposition of colleagues pursuing a theory-driven understanding of evolution [42]. Rather than argue the point of principle, we highlight promising lines of inquiry toward a tentative evolutionary theory of gene regulation.

Direction 1. Evolvable genotype-phenotype maps direct long-term evolution of regulatory sequences

If molecular context is fixed (e.g. TF binding preferences immutable), evolution under selection and drift is described by drift-barrier theory [43]: molecular recognition approaches but typically fails to fully settle at an optimum. But when GP map features can evolve (e.g. via mutations in trans-regulatory elements), new evolutionary pathways emerge. As an example, a transient increase in TF promiscuity can evolve and reshape the GP map, opening new pathways to neofunctionalization, before specificity re-evolves [44].

Often, multiple CREs share the same molecular context (e.g. targets of the same TF machinery). We recently formalized this setting as 'replicated GP maps', in which molecular parameters shaping the GP map (e.g. TF concentrations, binding preferences, and cooperativities) are shared across multiple CRE loci (hence 'replicated') in an organism [11,45]. Replicated GP maps evolve toward states that maximize the fraction of fit genotypes while minimizing the information that selection must accumulate in the CREs [46,47]. This principle holds, surprisingly, even when selection on any individual CRE is not overwhelmingly stronger than genetic drift [45]. Under realistic biophysical constraints, it enables first-principles predictions of key regulatory parameters, such as TF concentrations, binding site lengths, and TF cooperativity. The prediction of features of regulatory architecture that were previously taken as given highlights the potential of this theoretical development.

This information-theoretic approach also identifies conditions under which entirely new regulatory mechanisms, such as chromatin-based gene silencing, are expected to evolve: when the information cost of encoding the novel mechanism in the genome is smaller than the information savings it affords across all regulatory sites. This framework treats bits as a kind of evolutionary currency, allowing the feasibility and eventual fixation of novel mechanisms to be predicted rather than merely described [45].

Lastly, simulations show that the architectures predicted by this fitness-information tradeoff [45] may also be the most evolvable ones, that is, those where regulatory function can emerge most rapidly from random sequences [11] (Figure 1). This link between evolution and information theory frames evolvability as a global and quantifiable feature of the GP map. It also has the potential to explain long-standing puzzles in gene regulation, including the 'specificity paradox' [11]: why eukaryotes employ short, cooperatively acting binding sites of diverse strength.

Direction 2. Accounting for mutational robustness yields quantitative predictions for regulatory architecture

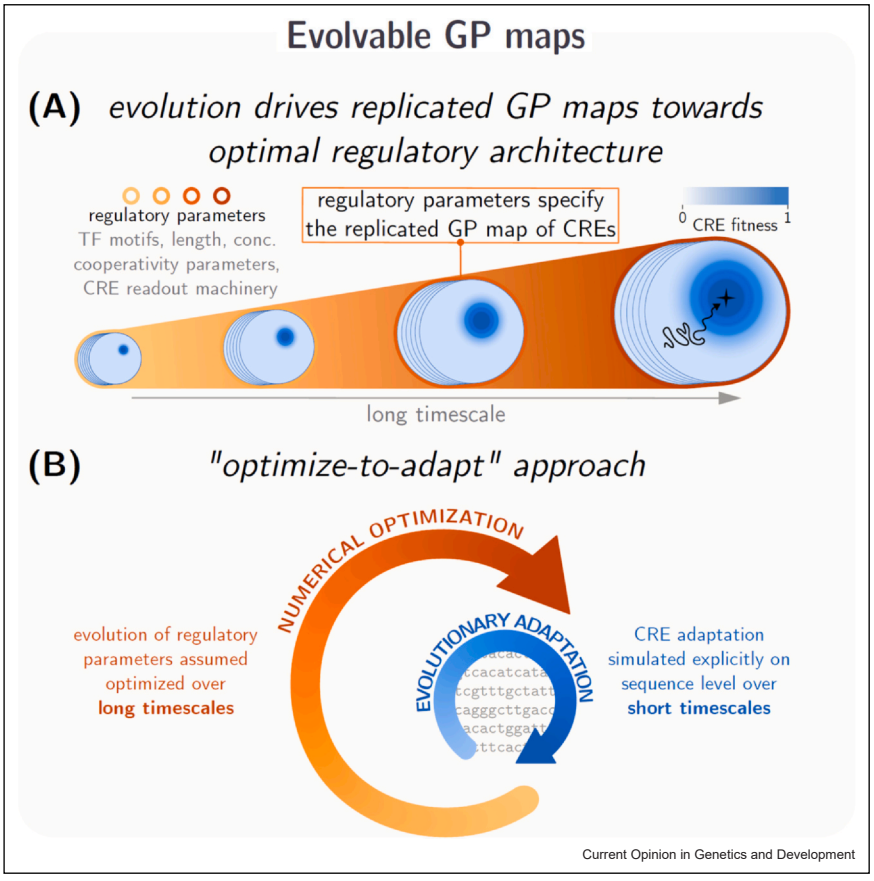
Most theoretical and simulational work on the evolution of regulatory sequences and architectures has focused on the weak-mutation regime, where populations are typically fixed for a single genotype and evolve toward a drift-selection balance that can be formalized analogously to the energy-entropy tradeoff of statistical physics. But what happens if the mutation rate is high enough to make populations polymorphic at CRE loci (a regime most likely to be relevant for prokaryotes)? The link between evolution and information theory naturally extends to this high-mutation regime: mutational flux acts as an additional entropic force, and selection must now accumulate a predictable amount of extra information to select genotypes that are not just functional but also more mutationally robust [45]. The action of selection for robustness thus becomes a quantifiable global metric, measurable in bits. Much work remains to flesh out such theoretical extensions and their implications.

An early indication that such extensions are worthwhile is provided by the elegant work by Stewart and colleagues, who argue why TF motifs are around 10 bp long [48]. This work incorporates many attributes we consider essential for a more general theory: it takes into account biophysical constraints and nonlinearities, it is formulated at the systems-level to model crosstalk, and it considers not only the evolution of CREs but also of the GP maps themselves (e.g. motif lengths can evolve). The ~10 bp TF length sweet spot is identified as a balance between mutational robustness and the TF specificity required for reliable gene regulation. While this argument does not explain differences between prokaryotes and eukaryotes and the 'specificity paradox', it showcases that including mutational robustness within a broader theoretical framework may be essential.

Future challenges

We conclude this review series by highlighting several outstanding theoretical and simulational challenges that now appear increasingly tractable. For brevity, we do not dwell on the (perfectly valid!) efforts to scale up all existing directions, from collecting more data on regulatory GP maps and analyzing them with better dimensionality-reduction methods, to refining mechanistic models with ever more detailed knowledge. Rather, we highlight higher-level arguments concerning the evolution of the regulatory system as a whole.

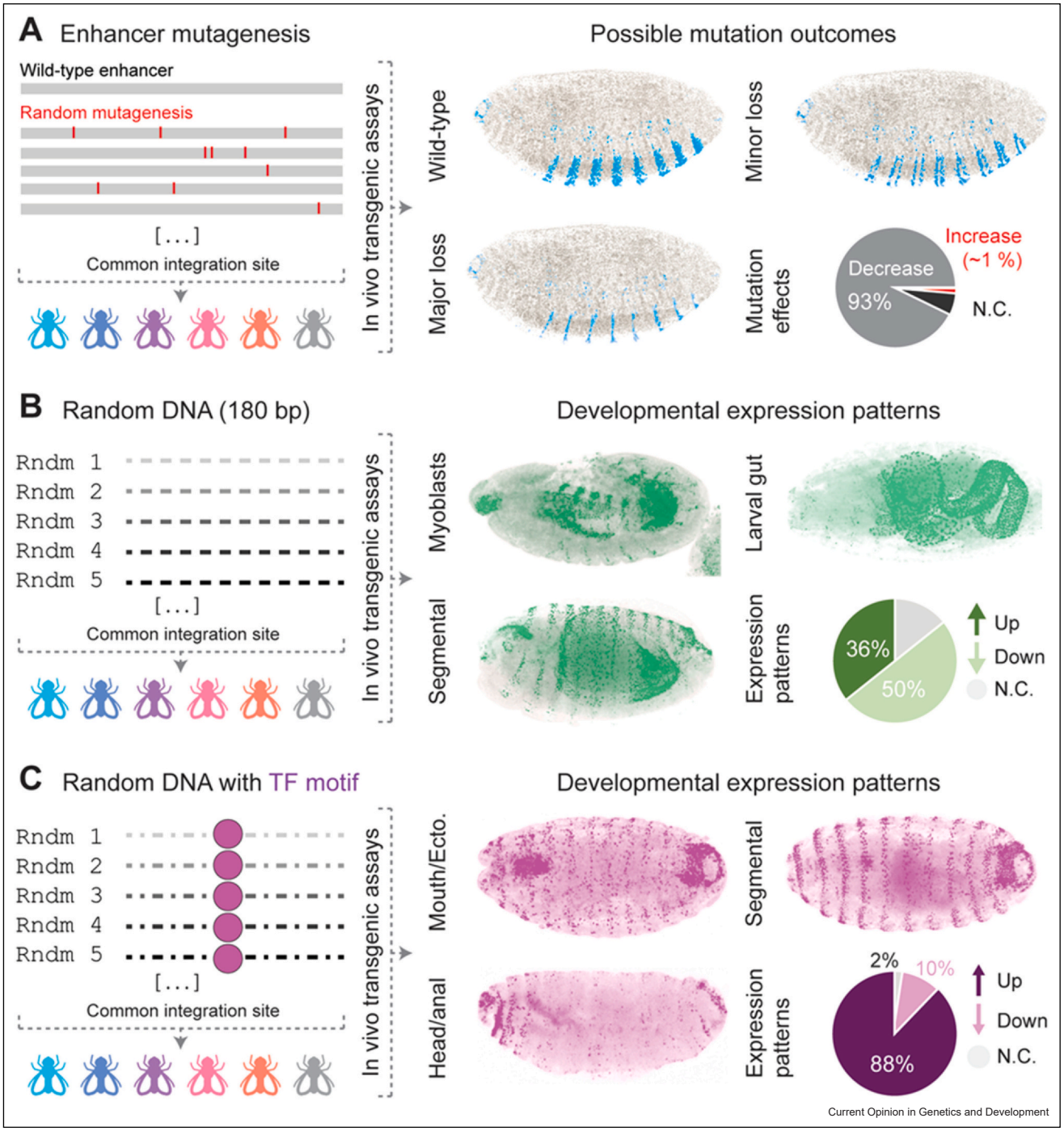
Figure 1



Co-evolution of CREs and of replicated GP maps for regulatory sequences. **(a)** Regulatory parameters (TF concentrations, binding specificities, cooperativity parameters, properties of the non-linear mapping from sequence to expression) are shared among the CREs in the genome, even when individual CREs are selected for different regulatory phenotypes. CREs can thus be seen as evolving on a 'replicated GP map' (piled blue discs, with color representing fitness over genotype space). Over long timescales, evolution favors regulatory parameters (left-to-right progression on the x-axis, with older maps on the left depicted as smaller, 'further away' disks) that maximize the fraction of fit phenotypes on the replicated GP map for regulatory sequences (the fraction of dark blue increases); at evolutionary equilibrium, this can be understood as a tradeoff between required genetic information to specify individual CREs and mean fitness. Surprisingly, this evolutionary optimization of genetic architecture for replicated maps can occur even when selection acting on individual mutations in CREs is not much stronger than drift. The optimization can predict equilibrium values for regulatory parameters (orange to red) *ab initio* [45]. **(b)** To tractably explore the outcomes of GP map and CRE sequence co-evolution, the 'optimize-to-adapt' approach assumes time-scale separation [11]. Shared regulatory parameters change much more slowly than individual CREs evolve. Assuming arguments from **(a)** hold, we can numerically optimize the regulatory parameters *in silico* (outer orange loop), so that the fraction of fit genotypes in the GP maps is maximized. These equilibrium parameters coincide with those for which evolutionary adaptation of individual CRE sequences (simulated via mutation, selection, and drift; inner blue loop) is most rapid and, in a sense that can be made mathematically precise, most evolvable.

Challenge 1. What are the most informative sequence ensembles to pursue, theoretically and experimentally? In Part 1 of this review series, we examined how classical mutagenesis studies on wild-type regulatory sequences can be complemented by a more recent focus on fully random libraries, which are closely linked to evolutionary theory [14,47]. These experimental strategies provide a way to infer quantitative GP maps, and thus can offer quantitative estimates of robustness (Figure 2a) and evolvability (Figure 2b, c) as described in this article, that is, as global rather than local features of GP maps. But are there library designs that interpolate between these two extremes? Are there clever ways to use phylogeny or theory to guide genotype construction for massively parallel assays? The iterative and directed enrichment of initially random libraries has already been used for synthetic promoter design [49], in SELEX-inspired methods to derive TF binding preferences [50], or for optimizing gene regulatory phenotypes [28]. Such directed (supervised) iterative approaches could be extended into the evolutionary domain, where more exploratory (unsupervised) yet still model-informed methods could allow experiments to efficiently navigate the vast sequence space to better extract global, quantitative GP maps.

Figure 2



Regulatory potential revealed by enhancer mutagenesis and random DNA in *Drosophila melanogaster*. Data from [51]. **(a)** Low mutational robustness of developmental enhancers. Random mutagenesis of native enhancers cloned upstream of a reporter gene, followed by site-specific integration and reporter assays in stage 15 embryos, shows that most mutations reduce enhancer output. For the shavenbaby *E3N* enhancer, most variants drive decreased expression, highlighting the sensitivity of developmental enhancers to sequence perturbations. **(b, c)** Expression is scored relative to a minimal reporter as increased (up), decreased (down), or unchanged (N.C.). **(b)** Latent regulatory activity in random DNA. Fully random 180 bp sequences cloned upstream of a reporter gene, when integrated at a common genomic site, frequently drive reproducible and spatially patterned reporter gene expression in late embryos, demonstrating that regulatory activity is widespread in sequence space even in the absence of evolved enhancers. **(c)** TF motifs enhance the evolvability of new regulatory function from random sequences. Random DNA libraries biased toward binding sites for the pioneer factor Zelda show a marked increase in both the frequency and magnitude of developmental expression, indicating that incorporation of key TF motifs strongly biases random sequence toward enhancer-like behavior.

Challenge 2. Isolating the impact of regulatory mechanisms on long-term evolution

Are the molecular mechanisms (and corresponding GP map features) with predictive power near the extant wild-type sequences the same that govern long-term regulatory sequence evolution? For instance, promiscuous RNAP binding with variable spacer length in bacterial promoters was found essential for predicted long-term dynamics, accelerating *de novo* regulatory function emergence by orders of magnitude [8]. Yet in extant promoters, this feature appears far less important: omitting it only slightly reduces predictive power. In other words, features that are essential for modeling regulatory sequences sampled around existing wild-type CREs may differ from those required to recapitulate long-term adaptive dynamics. Seen in a positive light, this consideration puts forward the simulated evolutionary dynamics as a potential 'arbiter' for distinguishing essential from incremental GP map features.

Challenge 3. Can information theory quantify regulatory code optimality?

Shannon's information theory piqued the interest of evolutionary biologists very early [52], but has remained on the sidelines of the mainstream, in contrast to its role in other branches of life science [53,54]. Recently established theoretical links suggest that, absent mechanistic constraints and under strong selection, regulatory architecture would evolve toward CREs that implement a Shannon-optimal code for desired phenotypes [45,47], that is, one that assigns information proportional to the negative logarithm of phenotype frequency, so that common phenotypes require less genotypic information and the overall genotypic information that must be maintained by selection is minimized. While more work is needed to grasp the impact and breadth of this connection, the simple fact that it links two fundamental theories, that is, evolution by natural selection and information theory, suggests that notions thought to be merely descriptive for genomes (e.g. sequence information, redundancy, compression, coding efficiency, etc.) may provide deeper quantitative insights.

Challenge 4. Evolution of gene regulatory circuits for emergent phenotypes

Gene regulation is a strongly interacting, feedback-rich, networked system, encoded by all CREs. Consequently, the function and evolutionary fate of an individual CRE may not be understandable when analyzed as an isolated input/output 'device'. While it is not unreasonable to assume, as a first approximation, that high fitness should correlate with precise and timely response of such devices to environmental (i.e. TF concentration) changes, it is ultimately the output of the entire gene regulatory network that gives rise to biological function, and that is selected for. In this review series, we focused on efforts to bridge the gap between CRE sequence and regulatory (gene expression) phenotype; Challenge 4 argues for further linking gene expression phenotypes (plural!) of multiple interacting CREs to collective, higher-order phenotypic functions. Examples here might include: the emergence of tissue or organismal shape and geometry, controlled by multiple genes and mediated by tissue mechanics; or cell-cell communication modulating gene regulatory networks across multiple cells that give rise to a reproducible body plan during early organismal development; or the gene regulatory networks in bacteria controlling enzyme levels to keep cellular metabolism running efficiently across different environments.

Impressive work has been carried out, especially in the evo-devo and biophysics communities, on the evolution of circuits giving rise to observed tissue- or organism-scale organization [55–60]. Yet most of this work abstracted away regulatory sequences, focused on mutational moves taking place directly at the level of gene network parameters, and often replaced evolutionary with numerical optimization. While this drastically reduces the dimensionality of the search space and makes such problems tractable simulation-wise [61], it also closes the door to the rigorous and quantitative application of evolvability, robustness, and genetic information concepts, as well as to realistic estimates of adaptation speed, which are only truly calibrated with reference to the actual genotypic space and mutational moves taking place therein. Therefore, a major goal for future work could be to connect at least one emerging tissue- or organismal-level phenotype (e.g. tissue patterning or morphology) that has measurable fitness proxies to regulatory DNA sequences through intermediate phenotypes (e.g. gene expression levels, regulatory network parameters, etc.). This would enable the explicit study of that phenotype's evolution via mutations in regulatory DNA. This program is ambitious because key links remain poorly understood, including long-range and 3D regulatory interactions, enhancer-promoter specificity, and the propagation of expression changes through developmentally integrated, pleiotropic, and polygenic trait architectures. In particular, long-distance effects, such as enhancer-promoter interactions typical of multicellular eukaryotes, complicate GP maps inference by making CREs' effects interdependent. At the same time, their unequal prevalence across taxa raises interesting questions about the evolution of regulatory grammars and architectures [1], similarly to the above-mentioned 'specificity paradox'. Due to these difficulties, progress will likely require focusing on biological systems where both mechanistic knowledge and experimental tractability are unusually strong.

Data Availability

No data were used for the research described in the article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Calin Guet and Santiago Herrera-Álvarez for essential contributions to this manuscript.

E.M. acknowledges support from the APART-USA fellowship, jointly funded by the Austrian Academy of Sciences (ÖAW) and the Institute of Science and Technology Austria (ISTA). N.B. acknowledges funding from the ERC Advanced Grant 101055327 "HaplotypeStructure".

This study was also supported by the European Molecular Biology Laboratory (N.O.B., J.C.).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Mascolo E., Borbély R., Herrera-Álvarez S., Guet C.C., Crocker J., Tkačik G. Long-term evolution of regulatory DNA sequences. Part 1: Simulations on global, biophysically-realistic genotype-phenotype maps. arXiv; [cited 2026 Jan 29]. Available from: http://arxiv.org/abs/2601.19681 doi:⟨10.48550/arXiv.2601.19681⟩.

2. • Otwinowski J, McCandlish DM, Plotkin JB: **Inferring the shape of global epistasis**. *Proc Natl Acad Sci* 2018, **115**:E7550-E7558, https://doi.org/10.1073/pnas.1804015115.
This work introduces a statistical framework to infer global epistasis from high-throughput mutational data, showing that much of the apparent epistasis in GP maps can be explained by nonlinear transformations of an underlying additive genetic trait.

3. Johnson MS, Reddy G, Desai MM: **Epistasis and evolution: recent advances and an outlook for prediction**. *BMC Biol* 2023, **21**:120, https://doi.org/10.1186/s12915-023-01585-3

4. Lagator M, Paixão T, Barton NH, Bollback JP, Guet CC: **On the mechanistic nature of epistasis in a canonical cis-regulatory element**. *eLife* 2017, **6**:e25192, https://doi.org/10.7554/eLife.25192

5. •• Metzger BPH, Park Y, Starr TN, Thornton JW: **Epistasis facilitates functional evolution in an ancient transcription factor**. *eLife* 2024, **12**:e88737, https://doi.org/10.7554/eLife.88737.2.
By reconstructing a transcription factor GP map, this study shows that pairwise epistasis and global nonlinearities are sufficient to explain functional evolution, that epistasis can strongly influence what evolutionary paths are accessible, and that it can favor the evolvability of regulatory function.

6. Starr TN, Thornton JW: **Epistasis in protein evolution**. *Protein Sci* 2016, **25**:1204-1218, https://doi.org/10.1002/pro.2897

7. Muñiz-Trejo R, Patton JE, Herrera-Álvarez S, Thornton JW: **Epistatic drift in protein evolution**. *Curr Opin Genet Dev* 2025, **95**:102412, https://doi.org/10.1016/j.gde.2025.102412

8. •• Lagator M, Sarikas S, Steinrueck M, Toledo-Aparicio D, Bollback JP, Guet CC, *et al.*: **Predicting bacterial promoter function and evolution from random sequences**. *eLife* 2022, **11**:e64543, https://doi.org/10.7554/eLife.64543.
This study uses random libraries to reveal and quantify the high evolvability of promoter function from random sequences, and shows how accounting for the biophysical properties of the RNA polymerase can provide significant insights into *de novo* promoter evolution.

9. Omidi S, Zavolan M, Pachkov M, Breda J, Berger S, Nimwegen E van: **Automated incorporation of pairwise dependency in transcription factor binding site prediction using dinucleotide weight tensors**. *PLoS Comput Biol* 2017, **13**:e1005176, https://doi.org/10.1371/journal.pcbi.1005176

10. Lagator M, Igler C, Moreno AB, Guet CC, Bollback JP: **Epistatic interactions in the arabinose cis-regulatory element**. *Mol Biol Evol* 2016, **33**:761-769, https://doi.org/10.1093/molbev/msv269

11. • Borbély R, Tkačik G: **Regulatory architectures optimized for rapid evolution of gene expression**. *bioRxiv* 2025, https://doi.org/10.1101/2025.06.10.658850 [cited 2026 Jan 12]. p. 2025.06.10.658850. Available from: https://www.biorxiv.org/content/10.1101/2025.06.10.658850v1.
This work shows how information theory can link evolutionary dynamics with biophysical constraints to rationalize — and possibly even predict — the features of regulatory architectures, and identifies which ones lead to high evolvability by linking them to global properties of the underlying GP map.

12. Westmann CA, Goldbach L, Wagner A: **The highly rugged yet navigable regulatory landscape of the bacterial transcription factor TetR**. *Nat Commun* 2024, **15**:10745, https://doi.org/10.1038/s41467-024-54723-y

13. • Park Y, Metzger BPH, Thornton JW: **The simplicity of protein sequence-function relationships**. *Nat Commun* 2024, **15**:7953, https://doi.org/10.1038/s41467-024-51895-5.
This work shows that the GP maps of proteins are often far simpler than expected, with most phenotypic variance explained by low-order interactions combined with global nonlinearities (global epistasis).

14. Grah R, Guet CC, Tkačik G, Lagator M: **Linking molecular mechanisms to their evolutionary consequences: a primer**. *Genetics* 2025, **229**:iyae191, https://doi.org/10.1093/genetics/iyae191

15. Haldane JBS: **The effect of variation of fitness**. *Am Nat* 1937, **71**:337-349, https://doi.org/10.1086/280722

16. Kondrashov AS: **Deleterious mutations and the evolution of sexual reproduction**. *Nature* 1988, **336**:435-440, https://doi.org/10.1038/336435a0

17. Igler C, Lagator M, Tkačik G, Bollback JP, Guet CC: **Evolutionary potential of transcription factors for gene regulatory rewiring**. *Nat Ecol Evol* 2018, **2**:1633-1643, https://doi.org/10.1038/s41559-018-0651-y

18. Tsai A, Alves MR, Crocker J: **Multi-enhancer transcriptional hubs confer phenotypic robustness**. *eLife* 2019, **8**:e45325, https://doi.org/10.7554/eLife.45325

19. Payne JL, Wagner A: **The causes of evolvability and their evolution**. *Nat Rev Genet* 2019, **20**:24-38, https://doi.org/10.1038/s41576-018-0069-z

20. Payne JL, Wagner A: **The robustness and evolvability of transcription factor binding sites**. *Science* 2014, **343**:875-877, https://doi.org/10.1126/science.1249046

21. Aguilar-Rodríguez J, Payne JL, Wagner A: **A thousand empirical adaptive landscapes and their navigability**. *Nat Ecol Evol* 2017, **1**:0045, https://doi.org/10.1038/s41559-016-0045

22. Aguilar-Rodríguez J, Payne JL: **Robustness and evolvability in transcriptional regulation**. In *Evolutionary Systems Biology: Advances, Questions, and Opportunities*. Edited by Crombach A. Springer International Publishing; 2021:197-219, , https://doi.org/10.1007/978-3-030-71737-7_9 [cited 2026 Jan 13].

23. Draghi JA, Parsons TL, Wagner GP, Plotkin JB: **Mutational robustness can facilitate adaptation**. *Nature* 2010, **463**:353-355, https://doi.org/10.1038/nature08694

24. •• Herrera-Álvarez S, Patton JEJ, Thornton JW: **The structure of an ancient genotype–phenotype map shaped the functional evolution of a protein family**. *Nat Ecol Evol* 2025, **9**:1656-1669, https://doi.org/10.1038/s41559-025-02777-6.
This study reconstructs an ancient GP map and shows how its structure biased the long-term evolution of an entire protein family.

25. • Li XC, Fuqua T, van Breugel ME, Crocker J: **Mutational scans reveal differential evolvability of Drosophila promoters and enhancers**. *Philos Trans R Soc B Biol Sci* 2023, **378**:20220054, https://doi.org/10.1098/rstb.2022.0054.

This work shows that promoters and enhancers differ systematically in robustness and evolvability, revealing a potential evolutionary division of labor in regulatory DNA.

26. Villar D, Berthelot C, Aldridge S, Rayner TF, Lukk M, Pignatelli M, *et al.*: **Enhancer evolution across 20 mammalian species**. *Cell* 2015, **160**:554-566, https://doi.org/10.1016/j.cell.2015.01.006

27. Grimbergen AJ, Siebring J, Solopova A, Kuipers OP: **Microbial bet-hedging: the power of being different**. *Curr Opin Microbiol* 2015, **25**:67-72, https://doi.org/10.1016/j.mib.2015.04.008

28. Wolf L, Silander OK, van Nimwegen E: **Expression noise facilitates the evolution of gene regulation**. *eLife* 2015, **4**:e05856, https://doi.org/10.7554/eLife.05856

29. Rieckh G. Studying the Complexities of Transcriptional Regulation [Thesis]; 2016 [cited 2026 Jan 13]. Available from: ⟨https://research-explorer.ista.ac.at/record/1128⟩.

30. • Vlková M, Silander OK: **Gene regulation in *Escherichia coli* is commonly selected for both high plasticity and low noise**. *Nat Ecol Evol* 2022, **6**:1165-1179, https://doi.org/10.1038/s41559-022-01783-2.
This work shows that gene regulation in *E. coli* is commonly shaped by selection for both high environmental responsiveness (plasticity) and low stochastic noise in gene expression, suggesting that promoter sequences and regulatory architectures have evolved to achieve both properties simultaneously.

31. Friedlander T, Prizak R, Guet CC, Barton NH, Tkačik G: **Intrinsic limits to gene regulation by global crosstalk**. *Nat Commun* 2016, **7**:12307, https://doi.org/10.1038/ncomms12307

32. Tkačik G, Gregor T, Bialek W: **The role of input noise in transcriptional regulation**. *PLoS One* 2008, **3**:e2774, https://doi.org/10.1371/journal.pone.0002774

33. Facchetti G, Iacono G, De Palo G, Altafini C: **A rate-distortion theory for gene regulatory networks and its application to logic gate consistency**. *Bioinformatics* 2013, **29**:1166-1173, https://doi.org/10.1093/bioinformatics/btt116

34. Tkačik G, Callan CG, Bialek W: **Information capacity of genetic regulatory elements**. *Phys Rev E* 2008, **78**:011910, https://doi.org/10.1103/PhysRevE.78.011910

35. Grünwald PD: **The Minimum Description Length Principle**. The MIT Press; 2007, https://doi.org/10.7551/mitpress/4643.001.0001

36. Zoller B, Gregor T, Tkačik G: **Eukaryotic gene regulation at equilibrium, or non?** *Curr Opin Syst Biol* 2022, **31**:100435, https://doi.org/10.1016/j.coisb.2022.100435

37. Perkins ML, Crocker J, Tkačik G: **Chromatin enables precise and scalable gene regulation with factors of limited specificity**. *Proc Natl Acad Sci* 2025, **122**:e2411887121, https://doi.org/10.1073/pnas.2411887121

38. Ewens WJ: **Mathematical population genetics**. In *Interdisciplinary Applied Mathematics*. Edited by Antman SS, Marsden JE, Sirovich L, Wiggins S. Springer; 2004, , https://doi.org/10.1007/978-0-387-21822-9 [cited 2026 Jan 13].

39. Savageau MA: **Design of molecular control mechanisms and the demand for gene expression**. *Proc Natl Acad Sci* 1977, **74**:5647-5651, https://doi.org/10.1073/pnas.74.12.5647

40. Gerland U, Hwa T: **Evolutionary selection between alternative modes of gene regulation**. *Proc Natl Acad Sci* 2009, **106**:8841-8846, https://doi.org/10.1073/pnas.0808500106

41. Grah R, Friedlander T: **The relation between crosstalk and gene regulation form revisited**. *PLoS Comput Biol* 2020, **16**:e1007642, https://doi.org/10.1371/journal.pcbi.1007642

42. Lässig M, Mustonen V, Walczak AM: **Predicting evolution**. *Nat Ecol Evol* 2017, **1**:0077, https://doi.org/10.1038/s41559-017-0077

43. Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M: **Drift-barrier hypothesis and mutation-rate evolution**. *Proc Natl Acad Sci* 2012, **109**:18488-18492, https://doi.org/10.1073/pnas.1216223109

44. Friedlander T, Prizak R, Barton NH, Tkačik G: **Evolution of new regulatory functions on biophysically realistic fitness landscapes**. *Nat Commun* 2017, **8**:216, https://doi.org/10.1038/s41467-017-00238-8

45. •• Barton NH, Tkačik G: **Evolution and information content of optimal gene regulatory architectures**. *bioRxiv* 2025, https://doi.org/10.1101/2025.06.10.657849 [cited 2026 Jan 12]. p. 2025.06.10.657849. Available from: https://www.biorxiv.org/content/10.1101/2025.06.10.657849v1.
This work develops an information-theoretic framework linking regulatory architecture and information with evolutionary dynamics, putting forward a general optimality principle for GP maps, and it introduces the concept of 'replicated genotype-phenotype maps'.

46. • Wagner A: **Information theory, evolutionary innovations and evolvability**. *Philos Trans R Soc B Biol Sci* 2017, **372**:20160416, https://doi.org/10.1098/rstb.2016.0416.
This work shows how an information-theoretic perspective can shed light on GP maps and their properties, including their ability to promote evolutionary innovation by enabling populations to access novel phenotypes while maintaining existing functions.

47. Hledík M, Barton N, Tkačik G: **Accumulation and maintenance of information in evolution**. *Proc Natl Acad Sci* 2022, **119**:e2123152119, https://doi.org/10.1073/pnas.2123152119

48. • Stewart AJ, Hannenhalli S, Plotkin JB: **Why Transcription factor binding sites are ten nucleotides long**. *Genetics* 2012, **192**:973-985, https://doi.org/10.1534/genetics.112.143370.
This work shows that the typical length of transcription factor binding sites (~10 bp) can be explained by an evolutionary tradeoff between the information required for binding specificity and the mutational load imposed by longer sites.

49. Wang X, Xu K, Huang Z, Lin Y, Zhou J, Zhou L, *et al.*: **Accelerating promoter identification and design by deep learning**. *Trends Biotechnol* 2025, **43**:3071-3087, https://doi.org/10.1016/j.tibtech.2025.05.008

50. Sahu B, Hartonen T, Pihlajamaa P, Wei B, Dave K, Zhu F, *et al.*: **Sequence determinants of human gene regulatory elements**. *Nat Genet* 2022, **54**:283-294, https://doi.org/10.1038/s41588-021-01009-4

51. Galupa R, Alvarez-Canales G, Borst NO, Fuqua T, Gandara L, Misunou N, *et al.*: **Enhancer architecture and chromatin accessibility constrain phenotypic space during Drosophila development**. *Dev Cell* 2023, **58**:51-62.e4, https://doi.org/10.1016/j.devcel.2022.12.003

52. Kimura M: **Natural selection as the process of accumulating genetic information in adaptive evolution**. *Genet Res* 1961, **2**:127-140, https://doi.org/10.1017/S0016672300000616

53. Tkačik G, Bialek W: **Information processing in living systems**. *Annu Rev Condens Matter Phys* 2016, **7**:89-117, https://doi.org/10.1146/annurev-conmatphys-031214-014803

54. Tkačik G, Wolde PR ten: **Information processing in biochemical networks**. *Annu Rev Biophys* 2025, **54**:249-274, https://doi.org/10.1146/annurev-biophys-060524-102720

55. • Sokolowski TR, Gregor T, Bialek W, Tkačik G: **Deriving a genetic regulatory network from an optimization principle**. *Proc Natl Acad Sci* 2025, **122**:e2402925121, https://doi.org/10.1073/pnas.2402925121.
This work shows how regulatory architecture can be derived from explicit optimization principles, connecting biophysical constraints to the evolution of regulatory codes.

56. François P, Hakim V: **Design of genetic networks with specified functions by evolution in silico**. *Proc Natl Acad Sci* 2004, **101**:580-585, https://doi.org/10.1073/pnas.0304532101

57. François P, Hakim V, Siggia ED: **Deriving structure from evolution: metazoan segmentation**. *Mol Syst Biol* 2007, **3**:MSB4100192, https://doi.org/10.1038/msb4100192

58. François P, Siggia ED: **Predicting embryonic patterning using mutual entropy fitness and in silico evolution**. *Development* 2010, **137**:2385-2395, https://doi.org/10.1242/dev.048033

59. Salazar-Ciudad I, Marín-Riera M: **Adaptive dynamics under development-based genotype–phenotype maps**. *Nature* 2013, **497**:361-364, https://doi.org/10.1038/nature12142

60. Mosby LS, Bowen AE, Hadjivasiliou Z: **Morphogens in the**
• **evolution of size, shape and patterning**. *Development* 2024, **151**:dev202412, https://doi.org/10.1242/dev.202412.
This work highlights how combining theoretical approaches with experiments can shed light on how changes in the spatiotemporal expression of morphogens drive the evolution of organismal phenotypes such as tissue and organ size, shape, and spatial patterning.

61. Sharpe J: **Computer modeling in developmental biology: growing today, essential tomorrow**. *Development* 2017, **144**:4214-4225, https://doi.org/10.1242/dev.151274