

## Review Article

# Phenotypic Nonspecificity as the Result of Limited Specificity of Transcription Factor Function

Anthony Percival-Smith 

Department of Biology, The University of Western Ontario, London, ON, Canada N6A 1B7

Correspondence should be addressed to Anthony Percival-Smith; [aperciva@uwo.ca](mailto:aperciva@uwo.ca)

Received 12 June 2018; Accepted 9 October 2018; Published 28 October 2018

Academic Editor: Martin Kupiec

Copyright © 2018 Anthony Percival-Smith. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Drosophila* transcription factor (TF) function is phenotypically nonspecific. Phenotypic nonspecificity is defined as one phenotype being induced or rescued by multiple TFs. To explain this unexpected result, a hypothetical world of limited specificity is explored where all TFs have unique random distributions along the genome due to low information content of DNA sequence recognition and somewhat promiscuous cooperative interactions with other TFs. Transcription is an emergent property of these two conditions. From this model, explicit predictions are made. First, many more cases of TF nonspecificity are expected when examined. Second, the genetic analysis of regulatory sequences should uncover *cis*-element bypass and, third, genetic analysis of TF function should generally uncover differential pleiotropy. In addition, limited specificity provides evolutionary opportunity and explains the inefficiency of expression analysis in identifying genes required for biological processes.

## 1. Introduction: The Specific World of TFs

In Biology gene expression is a major concern and is focused on how a single genotype exhibits multiple phenotypes. The classical example is an *E. coli* cell grown in glucose has very low levels of beta-galactosidase activity, whereas, the same cell will have high levels when grown in lactose [1]. Developmental biology provides nice examples of cells with the same DNA sequence, or genotype, giving rise to specialized transparent lens cells and red blood cells packed with hemoglobin. These distinct phenotypes are the result of differential programs of gene expression. The regulation of gene expression occurs at many different points along the flow of genetic information. One major mechanism of regulation is the control of the initiation of transcription, which is mediated by TFs. Eukaryotic TFs are generally thought to be composed of multiple functional domains: a DNA binding domain that recognizes a specific DNA sequence, a transcriptional regulation domain, and allosteric regulation domain(s) [2, 3]. This multipartite organization confers multiple levels for the regulation of TF specificity. Mechanisms of specificity include specific DNA sequence recognition mediated by the DNA binding domain and regulation of the activity of the TF

prior to binding DNA or after binding DNA. In this paper, I will assume that the initial binding of TFs sets in motion the recruitment of RNA polymerase and all subsequent gene and chromatin modifications and, therefore, although important, these epigenetic mechanisms are not discussed.

The first mechanism regulating TF specificity is specific DNA sequence recognition by the DNA binding domain. The evidence for this mechanism is reflected in the thousand or so structures of protein DNA complexes solved to date. Visualizing amino acid base interactions of a protein DNA complex was a watershed moment in the study of gene expression [4]. From the first structures of lambda repressors to the structure of many more protein DNA complexes, clear, specific amino acid base interactions were observed (Figure 1(a)). Perhaps the most beautiful mechanism of specific DNA recognition is the plant virulence TFs, TAL effectors [5]. The DNA binding domain of TAL TFs is a series of repeats of 34 amino acids with each repeat recognizing a specific base (Figure 1(b)). This one to one repeat to base recognition allows the engineering of proteins that can specifically recognize any DNA sequence in a genome [6]. Indeed, the idea that specific amino acid base contacts were made in the protein DNA complex leads to the isolation or design of change of specificity mutants [7–11]. The

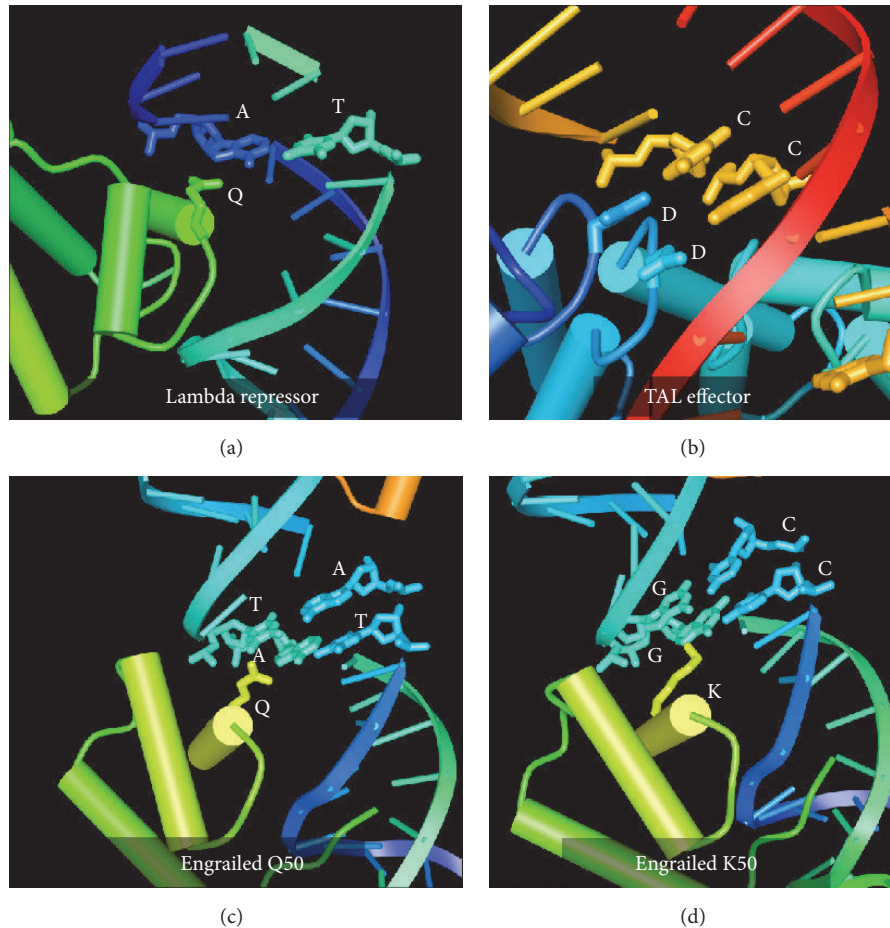


FIGURE 1: The specific amino acid base interactions of DNA binding protein domains with their recognition site. The interaction of a glutamine (green) of the 434 lambda repressor with a specific A:T base pair (blue:teal) (a) [4]. The interaction of the aspartic acids (blue) of two HD TAL repeats with cytosine bases (brown) (b) [5]. The interaction of the glutamine (yellow) of the wild-type Engrailed Q50 homeodomain with TA:AT base pairs (teal:blue) (c). The interaction of the lysine (yellow) of the change of specificity mutant Engrailed K50 homeodomain with GG:CC base pairs (teal:blue) (d) [12]. All images were generated with the Cn3D rendering program from the coordinates in the NCBI database [13].

structural basis for a change of specificity in a homeodomain (HD) has been determined [12] (Figures 1(c) and 1(d)). The studies of change of specificity mutants reinforce the idea that there are mechanisms of specificity required for the regulation of gene expression. Zooming out from the specific interactions at the interface of a TF DNA complex reveals the TF DNA complex on the promoter regulating the expression of a gene. Important for genetic analysis of gene expression, mutations in the DNA recognition site result in misregulation of gene expression identifying the sequence as a *cis*-acting regulatory element.

Although the structural studies beautifully illustrate how TFs recognize specific DNA sequences, there are long-standing questions about how TFs find the specific targets in complex genomes. Mathematical analysis of the kinetics of the lac repressor operator interaction in the *E. coli* was a major topic 40 years ago. One hypothesis proposed for the lac repressor system is that the lac repressor slides on one-dimensional DNA to find the recognition site, and this hypothesis is supported by image analysis in living cells [16].

In addition, analysis of the binding constants and kinetics of the protein DNA interactions point out two interesting properties. First the difference between the dissociation constant ( $K_d$ ) for recognition of specific DNA sequences and the nonspecific sequences can be large in the case of the lac repressor ( $10^{-7}$ - $10^{-3}$ ) or small in the case of the HD ( $10^{-2}$ ) [17, 18]. In cases of small differences, it does not take much nonspecific DNA to compete for binding with the recognition site. Second, for a dynamic system the dissociation constant needs to be large enough such that the TF dissociates from the DNA recognition site; otherwise, the TF is bound irreversibly. The half-life of the lac repressor DNA and HD DNA complexes are in the order of 20-30 minutes [18, 19]. In order to have high specificity for binding site recognition and a dynamic system, weak interactions between TFs in prokaryotes and eukaryotes occur and are a mechanism of cooperativity [20]. In prokaryotes many TFs oligomerize and the complex recognizes a sequence of 6-8 nucleotides and these oligomers interact on DNA using weak interactions to effectively increase the size of the target site recognized to

12-16 nucleotides. This is also observed in eukaryotes with the interactions between the Homeotic selector (HOX) proteins and Extradenticle (EXD) [21]. The cooperative interaction between HOX proteins and EXD increases the target size recognized and confers distinct target site recognition to many HOX proteins when in the complex with EXD. With the addition of Homothorax to the HOX EXD complex the target site becomes larger. However, in vitro analyses identifying the sequence of high affinity sites identifies many sequences that are not identified in the genome as occupied in a ChIP seq analysis [22]. Despite decades of work, how TFs find target sequences is still an interesting and open issue.

The application of information theory to the sequence information recognized by the DNA-binding domain of TFs establishes a clear difference between bacterial and eukaryotic TFs [23]. The DNA binding sites recognized by bacterial transcription factors have higher information content than the bacterial genome to be searched harboring these sequences; whereas, the DNA binding sites recognized by eukaryotic transcription factors have much lower information than the genome being searched and significantly lower information than bacterial TF DNA binding sites. Bacterial TFs recognize specific DNA elements in the genome because the DNA binding sites present enough information for specific binding; whereas, eukaryotic TFs do not recognize enough information in the DNA binding sites resulting in a large amount of spurious binding throughout the genome. This problem with the information present in the interaction between a eukaryotic transcription factor and its DNA binding sites is the heart of the Futility Theorem which asserts that essentially all predicted TF binding sites generated with models for the binding of individual TFs have no functional role [24]. This Futility Theorem still holds today for the analysis of eukaryotic TFs and particularly human TFs [25].

A second mechanism of specificity is the regulation of TF activity. TFs are regulated either prior to the formation of a DNA protein complex or after. During development, a simple mechanism of regulation prior to complex formation is whether the TF is expressed in the specific cell or not. Expression is not the only mechanism of regulation, and other mechanisms that regulate TF factor activity prior to DNA binding include regulation of subcellular localization. Nuclear receptor activity is regulated by whether it is in the nucleus or not, a mechanism shared by the NFkB family of TFs [26, 27]. In addition, TF activity is regulated after the formation of the DNA protein complex. An example of this is the yeast Gal4p transcription factor being bound to the *cis*-regulatory sequences in the presence or absence of galactose [28]. Gal4p activity is controlled by whether it is bound to Gal80p or not. The analysis of the last two mechanisms often involves the genetic dissection of the functional domains of the TF to identify the domains required for specific subcellular localization, binding specific regulatory proteins or allosteric effectors. The best examples of the effects of allosteric effectors on TF function are allolactose on the lac repressor interaction with DNA and steroids on nuclear factors. The specific world of transcription factor function is presently the major model for describing TF function. In the proceeding section, I propose an alternate model for

eukaryotic TF function. Although in describing this model, I do not incorporate or discuss much from the specific world; I do acknowledge that there are specific mechanisms regulating TF function, which would have to be incorporated into a more complete model for the regulation of the initiation of transcription of all genes in the genome.

## 2. A World of Limited Specificity

*2.1. Hypothesis: Limited Specificity of TFs.* Extensive phenotypic nonspecificity of TF function has been observed in *Drosophila* [14, 29–34]. In these observations, phenotypic nonspecificity is defined as one phenotype being induced or rescued by multiple TFs (Figure 2). Phenotypic nonspecificity is observed for HD containing TFs in the induction of wingless, eyeless, arista to tarsus transformations, ectopic thoracic beards, *Hox*-mediated control of autophagy, suppression of *spalt* expression phenotypes following ectopic expression and the rescue of neuromere development and mesoderm formation. This phenotypic nonspecificity was somewhat expected because the HOX DNA binding domain (the HD) recognizes the same recognition sequence. But surprisingly, the phenotypic nonspecificity is not just restricted to HD containing TFs. Both HD containing and non-HD containing TFs when ectopically expressed induce wingless, eyeless, ectopic thoracic beard phenotypes, and the reduced maxillary palp phenotype of *proboscipedia* is rescued by expression of the TF Doublesex male [14]. In the examples of phenotypic nonspecificity (Figure 2), four TFs that induce wingless and eyeless phenotypes have distinct DNA recognition sites of low information content and three are HD TFs and one is a Zn-finger TF. These observations by themselves do not prove a particular mechanism. Here I propose a mechanism and use this model to make explicit testable predictions that would support the hypothesis.

Three different models for TF function can be imagined. A true nonspecific model where every TF has no DNA sequence preference for binding which would lead to an absence of differential gene expression, and an extreme model of specificity where TFs find the promoters of specific sets of genes relying on a high degree of specific DNA sequence recognition which is the combined result of the DNA recognition of binding sites by individual DNA binding domains of TFs and coordinated with very specific protein::protein interactions between only a few TFs for cooperative binding to increase the information content of the DNA protein complex interaction. In the third model, I propose that the DNA sequence recognition and cooperative interactions are limited in specificity and not sufficient to target the expression of a very limited set of genes required for a specific phenotype. This model of limited specificity of TFs is supported by TF occupancy observed in a genome browser (ChIP seq) [35] at two different scales: at the small scale of a few kilobases TFs look to recognize specific regions (TF binding sites) of a promoter, but at a genome scale of megabases each TF is seemingly randomly distributed, and each random distribution of a TF is specific to each individual TF. I propose the distribution of TFs is dependent

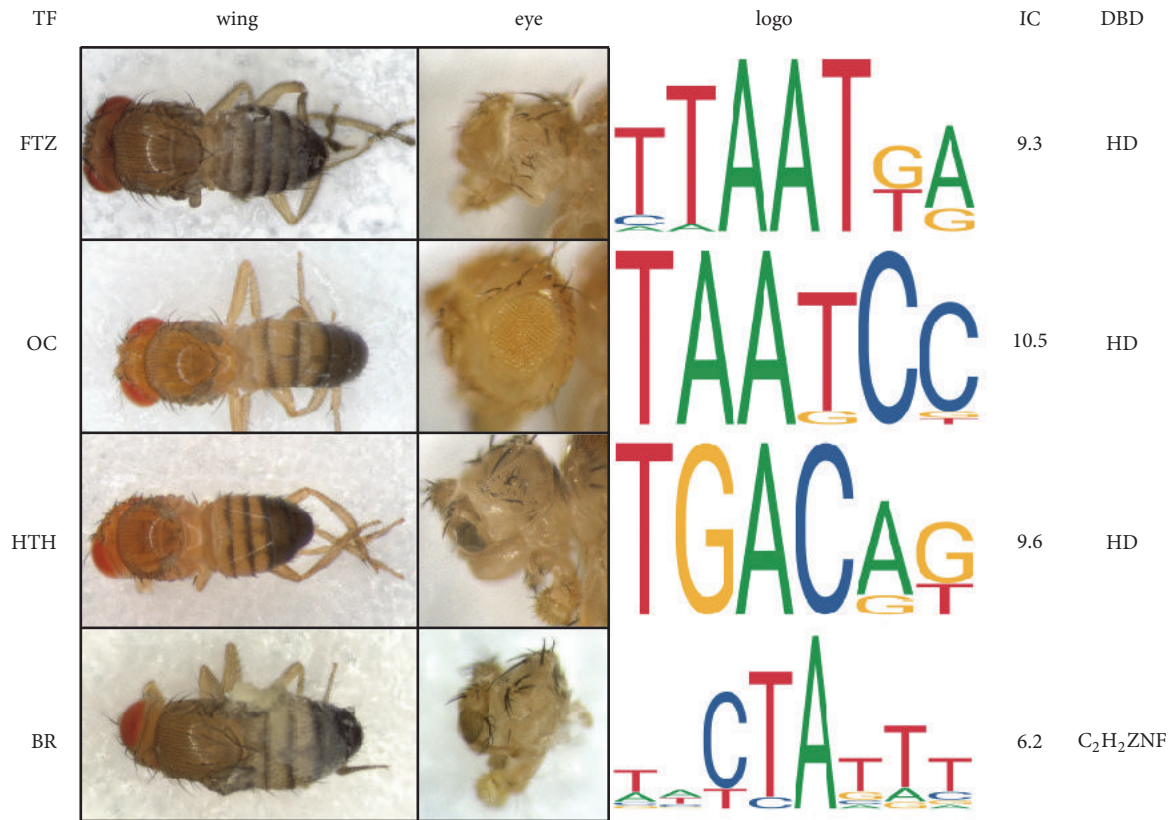


FIGURE 2: Phenotypic nonspecificity in *Drosophila*. The transcription factors are Fushi tarazu (FTZ), Ocelliless (OC), Homothorax (HTH), and Broad (BR). Wing and eye refer to the wingless and eyeless phenotypes induced by ectopic expression of the respective TF by either the rhomboid GAL4 or eyeless GAL4 driver [14]. The sequence logos for the recognition sequences of the four TFs are from the JASPAR database [15]. The information content (IC) for the recognition site is listed along with the DNA binding domain (DBD) of the TFs either the homeodomain (HD) or the C<sub>2</sub>H<sub>2</sub> Zn fingers (ZNF).

on the recognition of DNA sequences by the DNA binding domain and cooperative interactions between TFs. The low information content of the DNA sites leads to multiple binding of TFs all along the genome (for a TF that recognizes five bases, this would be a binding site on average every kilobase), and these distributions are refined by cooperative interactions. The cooperative interactions between TFs are also of limited specificity: some TFs interact homophilically, some TFs do not interact homophilically, and all TFs have a very large set of heterophilic interactions. Figure 3 shows the distribution of three TFs, the proportion of genes that have 0, 1, 2, ... TFs conforms to a Poisson distribution. (For simplicity's sake I have assumed each TF recognizes a binding site of the same length, each TF has the same ability to recruit RNA polymerase, and all genes are the same size). Transcription is an emergent property of these random distributions of TFs. To keep it simple in Figure 3, I assume that a gene with two TFs or more is expressed; however, there is not a good estimate of the real number required. Later I will modify and extend the model to explain default repression. The major point is that random distributions of TFs result in large sets of genes being transcribed. In the model of limited specificity, the genes required for completion of a particular process are a small proportion of the genes regulated by the TF and most genes have no role.

Limited specificity explains phenotypic nonspecificity because it allows the substitution of one TF with another. For example, a process that requires the expression of four genes, and the four genes, plus a hundred more, are regulated by TF<sub>a</sub>. All that is required for phenotypic nonspecificity is that the expression of TF<sub>b</sub> in place of TF<sub>a</sub> is able to result in the expression of the four genes required for the process; even though, the total set of regulated genes are quite distinct between the two TFs apart from these four genes.

I predict that there are many more cases of phenotypic nonspecificity. Indeed, there may be cases reported in the literature that have not been recognized as phenotypic nonspecificity. One interesting potential observation of phenotypic nonspecificity is that the initial set of four TFs identified that transform fibroblasts to pluripotent stem cells is not the only set of four TFs that induce pluripotency [36, 37]. Future analysis of the induction of pluripotency may uncover an even larger set of TFs capable of inducing this phenotype. In addition, in an analysis of the neuron cell fates of the *Drosophila* optic lobe, phenotypic convergence was observed where a neural trait is regulated independently by different combinations of transcription factors [38]. Phenotypic convergence like phenotypic nonspecificity may be a consequence of limited specificity. Because phenotypic nonspecificity is not often tested directly, the following set

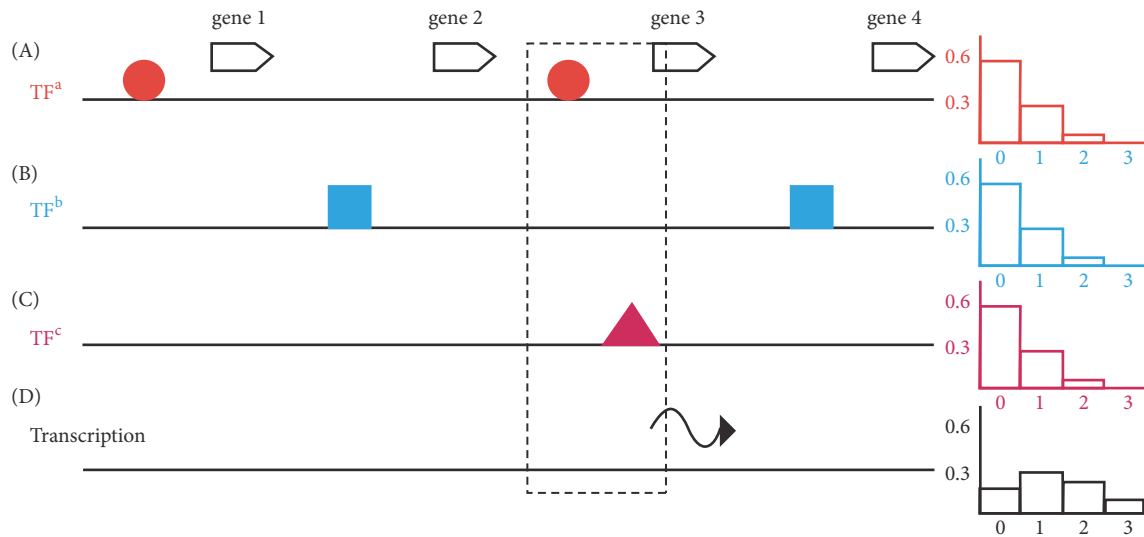


FIGURE 3: A simplified representation for the nonspecific model for transcriptional regulation. The pattern of binding of three TFs (TF<sup>a</sup> red circle, TF<sup>b</sup> blue square, and TF<sup>c</sup> magenta triangle) to a region of a genome with four genes (A-C) is shown. To the right of each distribution is the expected proportion of genes in the genome with no, 1, 2, and 3 TFs bound. The dotted box indicates a gene with two TFs bound, which in this simplified representation, is sufficient to recruit RNAP for transcription. Transcription (wiggly line) is an emergent property of random TF distributions. To the right is the expected proportion of genes in the genome with no, 1, 2, and 3 of these three TFs bound.

of predictions is made. First, I predict that many more TF combinations induce pluripotency. Second, the hierarchy of segmentation along the anterior posterior axis of *Drosophila* uses two large sets of TFs, the Gap proteins and the Pair-rule proteins [39]. Based on my previous observations of frequencies of phenotypic nonspecificity in *Drosophila* [14], I predict in 30-50% of these genes that the coding region can be substituted by 10-20% of *Drosophila* TFs and result in normal segmentation. In addition, this pattern of phenotypic rescue will also apply to the set of TFs expressed along the Dorsal ventral axis of the nerve cord in vertebrates, and the set of TFs temporally expressed during the development of the *Drosophila* brain [40, 41]. Based on the small size of the regulatory regions of yeast genes relative to *Drosophila* genes, I also predict that 10-20% of yeast transcription factor loci can have their coding regions substituted by 5-10% of yeast transcription factors. If any of these are shown, then these processes require the expression of a set of TFs rather than just the specific set identified, suggesting that TFs do not have intrinsic properties associated with a specific process. Basically, apart from some DNA sequence binding preferences and some preferences for TF partners to interact with cooperatively, TFs are very similar functionally, that is they lack unique and identifying leg, heart or muscle determining protein domains and functions.

**2.2. A Language Change Associated with Phenotypic Nonspecificity.** The language for discussing mechanisms of transcriptional regulation is biased. For example, the formulation of a common question in the study of HOX proteins “Given the high degree of sequence conservation of the HD between HOX proteins and therefore conservation of

the DNA binding sites recognized by these HDs, how do the HOX proteins function to specifically direct specific segmental identities during development?” asserts that there is phenotypic specificity of TF function; a mechanism for specificity is assumed and sought at the outset. The question of whether the phenotype regulated by a TF is phenotypically specific or nonspecific is neither discussed nor tested at the outset. The assumption of phenotypic specificity is down to the level of the syntactic articles used. For example, prior to my observation of extensive phenotypic nonspecificity in *Drosophila* [14], I would have confidently and definitively written that the TF Proboscipedia directs maxillary palp development; however, now I write a TF directs maxillary palp development, because the TF Doublesex Male rescues the proboscipedia phenotype. Phenotypic specificity uses the definite article “the”, and phenotypic nonspecificity uses the indefinite article “a”. In a further example of bias, in the discussion of multiple genes regulated by a TF, the genes are often divided into specific targets and nonspecific, off-targets suggesting that there is an intrinsic difference between these genes regulated by a TF and further implying that off-targets are safe to ignore. This use of language is arbitrarily dismissing a set of regulated genes as unimportant. In the model of limited specificity, more genes regulated by a TF are not required than are required for a process, and therefore, there are no such things as a limited set of confounding off-targets.

**2.3. Promoter Analysis in a World of Limited Specificity: cis-Element Bypass.** The mutational analysis of a promoter is a common approach for identifying *cis*-acting regulatory sequences controlling expression of a gene. In a world with limited specificity this approach is still valid, but the

interpretation of the results is restricted to an analysis of the present wild-type promoter and not an analysis of the potential of the promoter. When the promoter analysis is being performed, the set of TFs required for regulation of the gene does not change, and hence, does not address the potential of the promoter. If TFa can be substituted for another TFb then the phenomenon of *cis*-element bypass is expected, where the promoter with the mutation in the binding site for TFa is now expressed in the presence of TFb because TFb binds to its own distinct *cis*-element. Conversely, a promoter with a mutation in the *cis*-element binding TFb would be expressed in the presence of TFa, but not when TFa is substituted by TFb. *Cis*-element bypass is an explicit prediction of genetic analysis in the world of limited specificity and would suggest that the *cis*-elements are sufficient but not always necessary. The substitution of one TF for another uncovers the potential of the regulatory region. In a world of limited specificity, arguments about the importance of the organization of *cis*-elements in the promoter for regulation of gene expression are not very important.

**2.4. Analysis of the Functional Domains of TFs: Differential Pleiotropy.** The aim of mutational analysis of TF loci is determining the functional organization of the proteins by identifying functional domains and motifs important for establishing the wild-type pattern of gene expression. This approach is particularly important in the determination of the modular structure of TFs like Gal4p and nuclear receptors [2, 3]. However, the genetic analysis of other TFs like the *Drosophila* HD containing TFs Ultrabithorax, Sex combs reduced and Antennapedia, the yeast HD containing TF Pho2p, and Human HOX proteins has uncovered differential pleiotropy, the nonuniform behavior of mutant alleles across phenotypes suggesting motifs make small tissue specific contributions to overall TF activity [42–48]. The differential pleiotropy of mutations in TF loci has been attributed to an ensemble nature to allostery of intrinsically disordered proteins [49], but alternatively differential pleiotropy may also be an expected outcome in the model of limited specificity. In the model, the TF is randomly distributed, and each TF is not much different from another apart from its distribution. The TFs also interact cooperatively with limited specificity resulting in tissue specific TF distributions and activation of transcription. The mechanism of cooperativity is not restricted to two small highly structured protein domains making very specific amino acid interactions with each other such that only a limited number of TF::TF interactions occur. Rather the TF::TF interactions are mediated by more diffuse, nonspecific protein surfaces, and the way the surfaces of TFs interact may change from tissue to tissue and gene to gene. Therefore, the genetic analysis may not identify separate, specific interaction surfaces indicating a high degree of modularity. Rather the diffuse nonspecific surfaces and tissue specific and gene specific deployment of these surfaces show up in genetic analysis as differential pleiotropy where specific regions of the TF seem to make small tissue specific contributions to overall TF activity. Differential pleiotropy may be evidence supporting limited specificity.

**2.5. Phenotypic Nonspecificity as an Evolutionary Opportunity.** The study of evolution and development has identified Toolkit genes [50]. Toolkit genes share four major conserved characteristics: structure (basically amino acid sequence), expression, requirement, and function. Many of the Toolkit genes encode TFs. Unfortunately, in the experimental set up testing conservation of function, it is assumed that the function of the proteins is highly specific that is only that specific protein/function can generate or rescue the phenotype. This is particularly true in the analysis of Toolkit TFs. The experimental test is to substitute a TF with an orthologous TF from another organism and determine whether the ortholog can induce or rescue a particular developmental phenotype [51–56]. Observation of phenotypic nonspecificity renders this test uninformative because multiple nonorthologous/nonparalogous TFs can induce or rescue the phenotype [14, 29–34]. Therefore, some toolkit functions are now not functionally conserved, with structurally unrelated TFs able to substitute, severely weakening the Toolkit gene hypothesis at both the levels of conservation of function and structure leaving just conservation of expression and requirement.

In the evolutionary history of animals and plants it is hard to explain how a complex multicellular organism of multiple cell types arises as a result of development. A common assertion is that animals and plants have acquired the genes during evolution to accomplish this feat. A significant proportion of the genes required for determination of the body plan encodes TFs, and hypotheses have proposed that the increase in complexity is associated with the increase in the number of TFs and families of TFs [57]. In a sense, it is proposed that the acquisition of TFs recapitulates phylogeny [58]. This model has two problems. First, phenotypic nonspecificity suggests that TFs are not tailored by evolution to have a restricted role in a specific process. Second, the multiple body plan simplifications observed during evolution confounds the march to complexity. Phylogenomic analysis shows that between divergence of ctenophores and cnidarians, the simple porifera and placozoa phyla branch off; this body plan simplification is also observed in divergence of bilaterian phyla [59–61]. Phenotypic nonspecificity and body plan simplification suggests that a simple addition of genes does not explain or can be used to track the rise of complexity.

Also, an interesting consideration is the effect of the proliferation of transcription factors during evolution in the specific and limited specificity models. In the specific model, a transcription factor like Eyeless/PAX6 becomes associated with the regulation of genes required for eye development and once established is conserved during evolution. The origin of this specificity is difficult to explain. For example, when the gene encoding the TF arose what was the function of the TF at that point? Does it have no function and spends time this way, even though it can bind to DNA and affect transcription, until during evolution it becomes employed for a specific role? Does the TF have an initial function that during evolution is moved to another process and the original role replaced by another TF that has somehow changed its specificity? These are not considerations encountered with limited specificity. From the first time it is expressed, the TF is functional and as long as it does not result in a pattern of gene

expression that is detrimental to survival, it is maintained in the genome. During evolution expression of the TF or the genes it regulates may change such as upon genetic analysis in the extant organism it seems to be the essential TF regulator of a process but in reality, it is one of many TFs that have the potential to regulate the process. An expectation of limited specificity is the existence of TF loci that when mutant result in phenotypically viable organisms with no difference in fitness; however, analysis of differential gene expression shows large changes in the pattern of gene expression.

The phenomenon of phenotypic nonspecificity may be good for evolution and development, as phenotypic specificity is an evolutionary constraint, whereas, phenotypic nonspecificity is an evolutionary opportunity. In recently evolved systems involving regulation of multiple genes like the Bicoid and Dorsal morphogens, it would take a large number of random sequence changes in *cis*-elements during evolution in a specific world to bring about the specific pattern of gene expression required. Whereas, in a world of limited specificity, all that is required is a sequence change that results in a change in the expression pattern of a TF that has the potential of bringing about the pattern of gene expression required. Large-scale changes in gene expression can occur in the mechanism of limited specificity with one mutational change because the recognition sites (*cis*-elements) for the TF already exist in the genes required for that phenotypic change. Distalless normally required for limb development can be seconded into a different situation to pattern eye-spots in butterfly wings because the *cis*-elements for Distalless already exist in the genes required for eye spot formation [62]. The set of Dorsal regulated genes, which pattern the early dorsal ventral axis of *Drosophila* could be a consequence of a simple change in Dorsal expression [63]. *Cis*-element bypass uncovers the potential of regulatory regions, which is why it is an important prediction with ramifications for evolution. The world of limited specificity may be a powerful model to explain large-scale changes in gene expression that may have occurred during evolution.

**2.6. Phenotypic Nonspecificity and the Analysis of Patterns of Gene Expression.** In the world of limited specificity, a TF factor is required for the regulation of large sets of genes, the size of which depends on how widely the TF is expressed. In addition, as long as the genes whose products are required for the development of a specific organ are expressed at the correct time, place, and level whether tissue specifically or not is all that matters. These two considerations explain why expression analysis is inefficient at identifying the components required for specific developmental/biological processes. The example I will use is the simple developmental program of sporulation in *S. cerevisiae*, which includes meiosis and the encapsulation of the four haploid products in a spore wall. The general belief in the early 1980s was that genes required for sporulation would be expressed specifically during sporulation; indeed, a few genes specifically expressed during sporulation are required but a large majority of genes specifically expressed during sporulation are not required for sporulation [64–67]. In addition, a systematic screen of genes not required for yeast viability showed that there are

more genes required for sporulation that are not specifically expressed during sporulation than genes that are specifically expressed during sporulation, and more surprisingly genes downregulated during sporulation are also required for sporulation [68, 69]. Therefore, identifying genes required for sporulation using an expression approach was inefficient due to two reasons: first a minority of genes specifically expressed during sporulation were required for sporulation, and second the set of genes required for sporulation but not preferentially expressed during sporulation is larger than the set specifically expressed during sporulation and required. This result is expected in a world of limited specificity as the pattern of gene expression is less important than whether the gene is just expressed at the correct time and place irrespective of other considerations and, therefore, questioning the value of performing expression analysis to screen for genes required for biological processes. And conversely, the expression pattern of a gene may also not provide much information on how it is required. For example, in *Drosophila* String (*Cdc-25*) expression is restricted to mitotically dividing cells of the embryo potentially suggesting an important regulatory role for the cell cycle to occur or not. However, expression of String in all cells rescues the string phenotype and does not induce ectopic mitosis, suggesting that String does not have an important role in initiating mitosis [70]. In addition, the TF Fushi tarazu is expressed and required in the even-number parasegments but low-level expression in all cells partially rescues the fushi tarazu phenotype [71]. In the world of limited specificity, the relationships between expression pattern and requirement are not straightforward and do not generally support “guilt by association”. Differential gene expression is a well-described phenomenon; however, limited specificity may change how differential expression is primarily viewed: from a mechanism that promotes the expression of genes that initiate a process to a mechanism that inhibits the expression of genes that might disrupt a process, which would be under selection.

**2.7. Default Repression.** When an experimental perturbation is applied, which includes the alteration in the expression of a TF(s), gene expression analysis identifies genes that are both activated and repressed [72]. Interestingly the random distribution of TFs in the genome proposed in the model of limited specificity results in some promoters with few TFs bound and other promoters with many TFs bound. Using this characteristic of the model of limited specificity, I explain the activation/repression phenomenon with the incorporation of the three habits model and poised RNA polymerase II (RNAP) [73–75]. The three habits model explains the behavior of the terminal TFs of major cell-cell signaling pathways. The three habits are: activator insufficiency, where a ligand activated signaling pathway response element binding TFs (SPRE TF) is insufficient for activating transcription of a target gene alone; cooperativity, where a ligand activated SPRE TF with other local activators (TFs) cooperates and activates transcription; and default repression, where in the absence of a ligand the SPRE TF and local TFs repress transcription of the target gene. Although the three habits

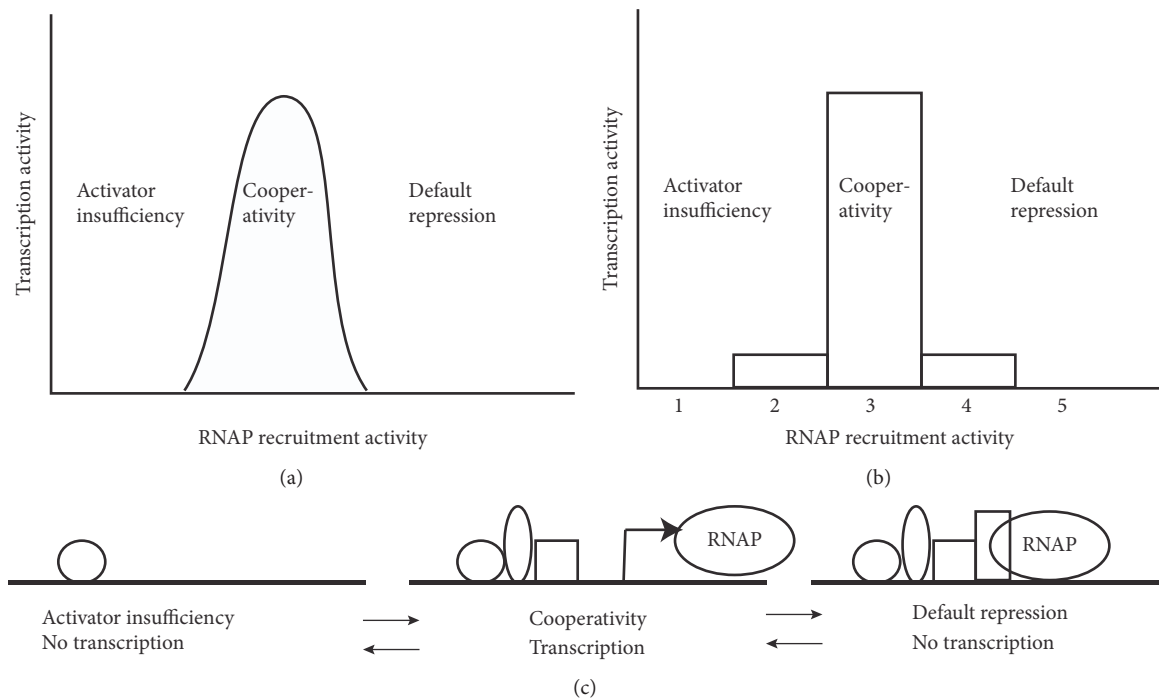


FIGURE 4: The model for default repression. Transcription activity is the rate of transcription initiation of a gene and RNAP recruitment activity is related to the amount of time RNAP is around the start of transcription of a gene (a and b). In (a) RNAP recruitment activity is assumed to be continuous, but in (b) it is assumed that each TF bound contributes an integer amount of TF recruitment activity that interacts cooperatively to recruit RNAP. It is assumed for simplicity of representation that when three TFs are bound, transcription is at its maximum. The three habits of the gene expression are shown as three interchangeable states (c). When one TF is bound transcription does not occur because of activator insufficiency. When three TFs are bound they cooperate and transcription occurs at the maximum rate. When four TFs are bound RNAP is strongly recruited to the promoter and does not engage in transcription and the gene is in default repression.

were proposed to explain observations on the expression of target genes at the end of cell-cell signaling pathways, I will extend it to all genes. I propose that transcriptionally active promoters are in a sweet spot or Goldilocks zone for the number of TFs bound. Starting from no bound TFs (Figure 4), as more TFs bind the amount of RNAP recruitment activity (the term RNAP recruitment is more vague than how it is defined by [76] which is the formation of a closed complex and more reflects time spent around the start of transcription), increase using cooperative interactions and transcription proceeds; however, there can be too much of a good thing where the RNAP recruiting activity is too high and RNAP is conformationally stuck or poised at the promoter. In Figure 4(b), I have shown a simplified example where all TFs have the same recruitment activity. At 1 or 2 TFs bound there is no or little expression (activator insufficiency), at three TFs bound there is the highest level of transcription (cooperativity), but above that number, RNAP is poised on the promoter and less transcription occurs (default repression). This model proposes two mechanisms for the activation of transcription: the gain of RNAP recruitment activity by the addition of TFs and the loss of RNAP recruitment activity by the loss or inactivation of TFs. Most models of transcription put transcription as the end point of RNAP recruitment activity such that poised RNAPs are a prelude to transcription; here poised RNAPs are the consequence of

too much RNAP recruitment activity. This explains nicely why in a genome wide analysis of expression, genes are both repressed and activated when a transcription factor is either added (ectopic expression) or taken away (loss of function). When a TF is added many genes now have enough RNAP recruitment activity to be transcribed and, in addition, now other genes have too much RNAP recruitment activity and are not transcribed as highly. Likewise, when a TF is taken away many genes now do not have enough RNAP recruitment activity to be transcribed while other genes have lost enough RNAP recruitment activity to now be transcribed. In this model, the TF is neither a repressor nor an activator, only the number of TFs bound at the promoter sets whether a gene is transcribed or not.

One prediction of this model is the existence of TF loci that reciprocally oscillate between high levels of transcription and high levels of TF accumulation due to both an autoregulatory element with many binding sites for the TF and the TF having a short half-life. The oscillation occurs because at low TF concentration the gene is highly transcribed but as the TF concentration increases and more TFs are bound to the autoregulatory element transcription decreases due to default repression resulting in a subsequent decrease in the concentration of the TF as it is rapidly degraded. New techniques in the live imaging of TF accumulation and active transcription may uncover examples of this [77].



I have no knowledge of experimental examples in eukaryotes of the phenomenon that reduction in RNAP recruitment results in the expression of a gene. However, there may be a bacterial example in the most classical of systems, the lac repressor [78, 79]. The lac repressor interacts with RNAP when not bound to allolactose. The RNAP in this complex with the lac repressor cycles through abortive initiation making short transcripts. The addition of allolactose to the lac repressor RNAP complex results in transient activation of transcription. The exact mechanism of this has been debated and investigated extensively [80]. The heat shock factor and RNAP are bound to (poised) the heat shock promoter in non-heat shocked cells. Poised RNAP II in eukaryotic cells also initiate short transcripts [81]. It is not difficult to imagine that the heat shock response is just temperature sensitive RNAP recruitment, as temperature increases recruitment decreases and transcription is initiated. The model also explains why the mutation of a silencer site results in activation of gene transcription, as the mutation just reduces the number of TFs that bind. In addition, the model explains how a TF can be both an activator and repressor in the same cell as each promoter has a different number of TFs bound. Another prediction of this model is that genes with poised RNA polymerases have more TFs bound than when they are transcribed. To test this prediction, the unbiased detection of the number of TFs bound to a promoter when transcribed or not is the major technical issue that needs to be overcome.

**2.8. Limited Specificity and the Mathematical Modeling of Gene Expression.** One of the interesting aspects of limited specificity is the potential to simplify mathematical modeling of eukaryotic gene expression. In the world of limited specificity all TFs have a similar intrinsic potential, differing only in the relative random distributions of binding in the genome. By applying a few rules to TF behavior, such as the DNA-binding preference and set of interacting TFs, the transcription profile of a cell expressing a defined set of transcription factors could be determined. It is easy to imagine doing this theoretically in a simulation, but it may be possible to take known transcription profiles and TF occupancy profiles to inform such a model and predict the outcome of a perturbation in vivo. The gene expression profile would be represented as a product of a matrix of genes, DNA-binding sites and TFs. Mathematical modeling is made easier in the world of limited specificity because TFs have a smaller number of specific attributes (variables).

### 3. Conclusions

To explain phenotypic nonspecificity of TFs in the induction or rescue of a phenotype, I propose a model where TFs are not inherently much different from one another in terms of intrinsic function and where transcription is an emergent property of the random distributions of these TFs which is influenced by DNA sequence binding preference and somewhat promiscuous TF:TF interactions. Using this model, I make three predictions. First further examination will uncover many more cases of phenotypic nonspecificity.

Second, in cases where multiple TFs can function to determine a phenotype, *cis*-element bypass will be observed. To my knowledge, no cases of *cis*-element bypass have been reported because it has yet to be directly tested. Third, genetic analysis of TFs in the nonspecific world will uncover differential pleiotropy. Differential pleiotropy is observed for the HD containing *Drosophila*, Human and yeast TFs: Ultrabithorax, Fushi tarazu, Sex combs reduced, Antennapedia, HOXA7, HOXB3, and Pho2p; therefore, in order to make a more general statement, observation of differential pleiotropy needs to be extended to other TF families. If the world of limited specificity exists, it has far reaching ramifications for evolution, expression analysis, and mathematical modeling of gene expression.

Although I have been outlining genetic analysis strictly in a world of limited specificity, highly specific mechanisms exist. One example is the specificity conferred to nuclear receptors by the ligand binding domains [82], and another example is the specific and modular interaction between Gal4p with Gal80p [83]. Therefore, if limited specificity exists, the next big problem would be what proportion of the observed patterns of gene expression does it explain? For example, if limited specificity explains 90% of gene expression where specificity only explains 10% then it would have to be considered the dominant model.

### Conflicts of Interest

The author has no conflicts of interest.

### Acknowledgments

This work was supported by the National Science and Engineering Research Council (NSERC) Discovery and Accelerator grants to Anthony Percival-Smith.

### References

- [1] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *Journal of Molecular Biology*, vol. 3, pp. 318–356, 1961.
- [2] R. Brent and M. Ptashne, "A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor," *Cell*, vol. 43, no. 3, pp. 729–736, 1985.
- [3] S. Rusconi and K. R. Yamamoto, "Functional dissection of the hormone and DNA binding activities of the glucocorticoid receptor," *EMBO Journal*, vol. 6, no. 5, pp. 1309–1315, 1987.
- [4] A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison, "Recognition of a DNA operator by the repressor of phage 434: A view at high resolution," *Science*, vol. 242, no. 4880, pp. 899–907, 1988.
- [5] D. Deng, C. Yan, X. Pan et al., "Structural basis for sequence-specific recognition of DNA by TAL effectors," *Science*, vol. 335, no. 6069, pp. 720–723, 2012.
- [6] M. Cristian, T. Cermak, E. L. Doyle, C. Schmidt, and et al., "DNA double-strand breaks with TAL effectors," *Genetics*, vol. 186, pp. 757–761, 2010.

- [7] P. Youderian, A. Vershon, S. Bouvier, R. T. Sauer, and M. M. Susskind, "Changing the DNA-binding specificity of a repressor," *Cell*, vol. 35, no. 3, pp. 777–783, 1983.
- [8] M. Hollis, D. Valenzuela, D. Pioli, and R. Wharton, "Specific recognition of a chimeric operator by a repressor heterodimer," *Proc. Natl. Acad. Sci. USA*, vol. 85, pp. 5834–5838, 1988.
- [9] S. D. Hanes and R. Brent, "DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9," *Cell*, vol. 57, no. 7, pp. 1275–1283, 1989.
- [10] A. Percival-Smith, M. Muller, M. Affolter, and W. J. Gehring, "The interaction with DNA of wild-type and mutant fushi tarazu homeodomains," *EMBO Journal*, vol. 9, no. 12, pp. 3967–3974, 1990.
- [11] S. D. Hanes and R. Brent, "A genetic model for interaction of the homeodomain recognition helix with DNA," *Science*, vol. 251, no. 4992, pp. 426–430, 1991.
- [12] L. Tucker-Kellogg, M. A. Rould, K. A. Chambers, S. E. Ades, R. T. Sauer, and C. O. Pabo, "Engrailed (Gln50→Lys) homeodomain-DNA complex at 1.9 Å resolution: Structural basis for enhanced affinity and altered specificity," *Structure*, vol. 5, no. 8, pp. 1047–1054, 1997.
- [13] Y. Wang, L. Y. Geer, C. Chappay, J. A. Kans, and S. H. Bryant, "Cn3D: sequence and structure views for Entrez," *Trends in Biochemical Sciences*, vol. 25, no. 6, pp. 300–302, 2000.
- [14] A. Percival-Smith, "Non-specificity of transcription factor function in *Drosophila melanogaster*," *Development Genes and Evolution*, vol. 227, no. 1, pp. 25–39, 2017.
- [15] A. Khan, O. Fornes, A. Stigliani et al., "JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework," *Nucleic Acids Research*, vol. 46, no. 1, pp. D260–D266, 2018.
- [16] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf, "The lac repressor displays facilitated diffusion in living cells," *Science*, vol. 336, no. 6088, pp. 1595–1598, 2012.
- [17] S.-Y. Lin and A. D. Riggs, "Lac repressor binding to DNA not containing the lac operator and to synthetic poly dAT," *Nature*, vol. 228, no. 5277, pp. 1184–1186, 1970.
- [18] M. Affolter, A. Percival-Smith, M. Muller, W. Leupin, and W. J. Gehring, "DNA binding properties of the purified Antennapedia homeodomain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 87, no. 11, pp. 4093–4097, 1990.
- [19] A. D. Riggs, R. F. Newby, and S. Bourgeois, "lac repressor-Operator interaction. II. Effect of galactosides and other ligands," *Journal of Molecular Biology*, vol. 51, no. 2, pp. 303–314, 1970.
- [20] A. D. Johnson, B. J. Meyer, and M. Ptashne, "Interactions between DNA-bound repressors govern regulation by the  $\lambda$  phage repressor," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 10, pp. 5061–5065, 1979.
- [21] H. D. Ryoo, T. Marty, F. Casares, M. Affolter, and R. S. Mann, "Regulation of Hox target genes by a DNA bound homothorax/Hox/extradenticle complex," *Development*, vol. 126, no. 22, pp. 5137–5148, 1999.
- [22] M. Slattery, T. Riley, P. Liu et al., "Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins," *Cell*, vol. 147, no. 6, pp. 1270–1282, 2011.
- [23] Z. Wunderlich and L. A. Mirny, "Different gene regulation strategies revealed by analysis of binding motifs," *Trends in Genetics*, vol. 25, no. 10, pp. 434–440, 2009.
- [24] W. W. Wasserman and A. Sandelin, "Applied bioinformatics for the identification of regulatory elements," *Nature Reviews Genetics*, vol. 5, no. 4, pp. 276–287, 2004.
- [25] S. A. Lambert, A. Jolma, L. F. Campitelli et al., "The Human Transcription Factors," *Cell*, vol. 175, no. 2, pp. 598–599, 2018.
- [26] D. Picard and K. R. Yamamoto, "Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor," *EMBO Journal*, vol. 6, no. 11, pp. 3333–3340, 1987.
- [27] P. A. Baeuerle and D. Baltimore, "I $\kappa$ B: A specific inhibitor of NF $\kappa$ B transcription factor," in *Science*, vol. 242, pp. 540–546, 1988.
- [28] E. Giniger, S. M. Varnum, and M. Ptashne, "Specific DNA binding of GAL4, a positive regulatory protein of yeast," *Cell*, vol. 40, no. 4, pp. 767–774, 1985.
- [29] S. Greig and M. Akam, "The role of homeotic genes in the specification of the *Drosophila* gonad," *Current Biology*, vol. 5, no. 9, pp. 1057–1062, 1995.
- [30] F. Hirth, T. Loop, B. Egger, D. F. B. Miller, T. C. Kaufman, and H. Reichert, "Functional equivalence of Hox gene products in the specification of the tritocerebrum during embryonic brain development of *Drosophila*," *Development*, vol. 128, no. 23, pp. 4781–4788, 2001.
- [31] A. Percival-Smith, W. A. Teft, and J. L. Barta, "Tarsus determination in *Drosophila melanogaster*," *Genome*, vol. 48, no. 4, pp. 712–721, 2005.
- [32] A. Banreti, B. Hudry, M. Sass, A. Saurin, and Y. Graba, "Hox Proteins Mediate Developmental and Environmental Control of Autophagy," *Developmental Cell*, vol. 28, no. 1, pp. 56–69, 2014.
- [33] K. M. Lelli, B. Noro, and R. S. Mann, "Variable motif utilization in homeotic selector (Hox)-cofactor complex formation controls specificity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 52, pp. 21122–21127, 2011.
- [34] A. Percival-Smith, L. Sivanantharajah, J. J. H. Pelling, and W. A. Teft, "Developmental competence and the induction of ectopic proboscises in *Drosophila melanogaster*," *Development Genes and Evolution*, vol. 223, no. 6, pp. 375–387, 2013.
- [35] D. Karolchik, R. M. Kuhn, R. Baertsch et al., "The UCSC genome browser database: 2008 update," *Nucleic Acids Research*, vol. 36, no. 1, pp. D773–D779, 2008.
- [36] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.
- [37] M. Maekawa, K. Yamaguchi, T. Nakamura, and R. Shibukawa, "Direct reprogramming of somatic cells is promoted by maternal transcription factor Glis1," *Nature*, vol. 474, pp. 225–229, 2011.
- [38] N. Konstantinides, K. Kapuralin, C. Fadil, and Barboza., "Phenotypic convergence: distinct transcription factors regulate common terminal features," *Cell*, vol. 174, pp. 622–635, 2018.
- [39] M. Akam, "The molecular basis for metameric pattern in the *Drosophila* embryo," *Development*, vol. 101, no. 1, pp. 1–22, 1987.
- [40] G. Le Dréau and E. Martí, "Dorsal-ventral patterning of the neural tube: a tale of three signals," *Developmental Neurobiology*, vol. 72, no. 12, pp. 1471–1481, 2012.
- [41] X. Li, T. Erclik, C. Bertet et al., "Temporal patterning of *Drosophila* medulla neuroblasts controls neural fates," *Nature*, vol. 498, no. 7455, pp. 456–462, 2013.
- [42] L. T. Bhoite, J. M. Allen, E. Garcia et al., "Mutations in the Pho2 (Bas2) transcription factor that differentially affect activation with its partner proteins Bas1, Pho4, and Swi5," *The Journal of Biological Chemistry*, vol. 277, no. 40, pp. 37612–37618, 2002.

- [43] E. Tour, C. T. Hittinger, and W. McGinnis, "Evolutionarily conserved domains required for activation and repression functions of the *Drosophila* Hox protein Ultrabithorax," *Development*, vol. 132, no. 23, pp. 5271–5281, 2005.
- [44] C. T. Hittinger, "Pleiotropic functions of a conserved insect-specific Hox peptide motif," *Development*, vol. 132, no. 23, pp. 5261–5270, 2005.
- [45] F. Prince, T. Katsuyama, Y. Oshima et al., "The YPWM motif links Antennapedia to the basal transcriptional machinery," *Development*, vol. 135, no. 9, pp. 1669–1679, 2008.
- [46] L. Sivanantharajah and A. Percival-Smith, "Analysis of the sequence and phenotype of *Drosophila* Sex combs reduced alleles reveals potential functions of conserved protein motifs of the Sex combs reduced protein," *Genetics*, vol. 182, no. 1, pp. 191–203, 2009.
- [47] L. Sivanantharajah and A. Percival-Smith, "Acquisition of a leucine zipper motif as a mechanism of antimorphy for an allele of the *Drosophila* Hox gene sex combs reduced," *G3: Genes, Genomes, Genetics*, vol. 4, no. 5, pp. 829–838, 2014.
- [48] A. Dard, J. Reboulet, F. JiaY Bleicher, and M. Duffraisse, "Human HOX protein use diverse and context-dependent motifs to interact with TALE class cofactors," *Cell Reports*, vol. 22, pp. 3058–3071, 2018.
- [49] L. Sivanantharajah and A. Percival-Smith, "Differential pleiotropy and HOX functional organization," *Developmental Biology*, vol. 398, no. 1, pp. 1–10, 2015.
- [50] S. B. Carroll, "Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution," *Cell*, vol. 134, no. 1, pp. 25–36, 2008.
- [51] J. Malicki, K. Schughart, and W. McGinnis, "Mouse Hox-2.2 specifies thoracic segmental identity in *Drosophila* embryos and larvae," *Cell*, vol. 63, no. 5, pp. 961–967, 1990.
- [52] J. J. Zhao, R. A. Lazzarini, and L. Pick, "The mouse Hox-1.3 gene is functionally equivalent to the *Drosophila* Sex combs reduced gene," *Genes & Development*, vol. 7, no. 3, pp. 343–354, 1993.
- [53] C. Hunter and C. Keynon, "Specification of anteroposterior cell fates in *Caenorhabditis elegans* by *Drosophila* Hox proteins," in *Nature*, vol. 377, pp. 229–232, 1995.
- [54] G. Halder, P. Callaerts, and W. J. Gehring, "Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*," *Science*, vol. 267, no. 5205, pp. 1788–1792, 1995.
- [55] B. Lutz, H.-C. Lu, G. Eichele, D. Miller, and T. C. Kaufman, "Rescue of *Drosophila* labial null mutant by the chicken ortholog Hoxb-1 demonstrates that the function of Hox genes is phylogenetically conserved," *Genes & Development*, vol. 10, no. 2, pp. 176–184, 1996.
- [56] A. Percival-Smith and J. A. Laing Bondy, "Analysis of murine HOXA-2 activity in *Drosophila melanogaster*," *Developmental Genetics*, vol. 24, no. 3-4, pp. 336–344, 1999.
- [57] J. F. Ryan, K. Pang, J. C. Mullikin, M. Q. Martindale, and A. D. Baxevanis, "The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa," *EvoDevo*, vol. 1, no. 1, 2010.
- [58] R. A. Raff and T. C. Kaufman, "Embryos, genes, and evolution," in *pages 251-261*, pp. 251–261, Indiana University Press, Bloomington, IN, USA, 1991.
- [59] J. F. Ryan, K. Pang, C. E. Schnitzler, A. D. Nguyen, and R. T. Moreland, "The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution," *Science*, vol. 342, Article ID 1232592, 2013.
- [60] L. L. Moroz, K. M. Kocot, M. R. Citarella, S. Dosung, and T. P. Norekian, "The ctenophore genome and the evolutionary origins of neural systems," *Nature*, vol. 510, pp. 109–114, 2014.
- [61] J. Aruga, Y. S. Odaka, A. Kamiya, and H. Furuya, "Dicyema Pax6 and Zic: Tool-kit genes in a highly simplified bilaterian," *BMC Evolutionary Biology*, vol. 7, no. 1, article no. 201, 2007.
- [62] B. Dhungel, Y. Ohno, R. Matayoshi et al., "Distal-less induces elemental color patterns in *Junonia* butterfly wings," *Zoological Letters*, vol. 2, no. 1, 2016.
- [63] A. Stathopoulos, M. Van Drenth, A. Erives, M. Markstein, and M. Levine, "Whole-genome analysis of Dorsal-ventral patterning in the *Drosophila* embryo," *Cell*, vol. 111, no. 5, pp. 687–701, 2002.
- [64] A. Percival-Smith and J. Segall, "Isolation of DNA sequences preferentially expressed during sporulation in *Saccharomyces cerevisiae*," *Molecular and Cellular Biology*, vol. 4, no. 1, pp. 142–150, 1984.
- [65] A. Percival-Smith and J. Segall, "Characterization and mutational analysis of a cluster of three genes expressed preferentially during sporulation of *Saccharomyces cerevisiae*," *Molecular and Cellular Biology*, vol. 6, no. 7, pp. 2443–2451, 1986.
- [66] A. T. Garber and J. Segall, "The SPS4 gene of *Saccharomyces cerevisiae* encodes a major sporulation-specific mRNA," *Molecular and Cellular Biology*, vol. 6, no. 12, pp. 4478–4485, 1986.
- [67] S. Chu, J. DeRisi, M. Eisen et al., "The transcriptional program of sporulation in budding yeast," *Science*, vol. 282, no. 5389, pp. 699–705, 1998.
- [68] A. M. Deutschbauer, R. M. Williams, A. M. Chu, and R. W. Davis, "Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 24, pp. 15530–15535, 2002.
- [69] A. H. Enyenihi and W. S. Saunders, "Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*," *Genetics*, vol. 163, no. 1, pp. 47–54, 2003.
- [70] B. A. Edgar and P. H. O'Farrell, "The three postblastoderm cell cycles of *Drosophila* embryogenesis are regulated in G2 by string," *Cell*, vol. 62, no. 3, pp. 469–480, 1990.
- [71] B. Argiropoulos, J. Ho, B. J. Blachuta, I. Tayyab, and A. Percival-Smith, "Low-level ectopic expression of Fushi tarazu in *Drosophila melanogaster* results in ftzUal/Rpl-like phenotypes and rescues ftz phenotypes," *Mechanisms of Development*, vol. 120, no. 12, pp. 1443–1453, 2003.
- [72] S. D. Hueber, D. Bezdán, S. R. Henz, M. Blank, H. Wu, and I. Lohmann, "Comparative analysis of Hox downstream genes in *Drosophila*," *Development*, vol. 134, no. 2, pp. 381–392, 2007.
- [73] D. S. Gilmour and J. T. Lis, "RNA polymerase II interacts with the promoter region of the noninduced hsp70 gene in *Drosophila melanogaster* cells," *Molecular and Cellular Biology*, vol. 6, no. 11, pp. 3984–3989, 1986.
- [74] S. Barolo and J. W. Posakony, "Three habits of highly effective signaling pathways: Principles of transcriptional control by developmental cell signaling," *Genes & Development*, vol. 16, no. 10, pp. 1167–1181, 2002.
- [75] J. Zeitlinger, A. Stark, M. Kellis et al., "RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo," *Nature Genetics*, vol. 39, no. 12, pp. 1512–1516, 2007.
- [76] B. Krummel and M. J. Chamberlin, "RNA Chain Initiation by *Escherichia coli* RNA Polymerase. Structural Transitions of the Enzyme in Early Ternary Complexes," *Biochemistry*, vol. 28, no. 19, pp. 7829–7842, 1989.

- [77] J. P. Bothma, M. R. Norstad, S. Alamos, and H. G. Garcia, "LlamaTags: A Versatile Tool to Image Transcription Factor Dynamics in Live Embryos," *Cell*, vol. 173, no. 7, pp. 1810–1822.e16, 2018.
- [78] A. Schmitz and D. J. Galas, "The interaction of RNA polymerase and lac repressor with the lac control region," *Nucleic Acids Research*, vol. 6, no. 1, pp. 111–137, 1979.
- [79] S. B. Straney and D. M. Crothers, "Lac repressor is a transient gene-activating protein," *Cell*, vol. 51, no. 5, pp. 699–707, 1987.
- [80] A. Sanchez, M. L. Osborne, L. J. Friedman, J. Kondev, and J. Gelles, "Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules," *EMBO Journal*, vol. 30, no. 19, pp. 3940–3946, 2011.
- [81] I. Jonkers and J. T. Lis, "Getting up to speed with transcription elongation by RNA polymerase II," *Nature Reviews Molecular Cell Biology*, vol. 16, no. 3, pp. 167–177, 2015.
- [82] S. Green and P. Chambon, "Oestradiol induction of a glucocorticoid-responsive gene by a chimeric receptor," *Nature*, pp. 75–78, 1987.
- [83] J. B. Thoden, L. A. Ryan, R. J. Reece, and H. M. Holden, "The interaction between an acidic transcriptional activator and its inhibitor: The molecular basis of Gal4p recognition by Gal80p," *The Journal of Biological Chemistry*, vol. 283, no. 44, pp. 30266–30272, 2008.