Review

CellPress

Transcription factors: specific DNA binding and specific gene regulation

Anne-Laure Todeschini[1,2], Adrien Georges[1,2], and Reiner A. Veitia[1,2]

[1] Institut Jacques Monod, Paris, France
[2] Université Paris Diderot, Paris, France

Specific recognition of *cis*-regulatory regions is essential for correct gene regulation in response to developmental and environmental signals. Such DNA sequences are recognized by transcription factors (TFs) that recruit the transcriptional machinery. Achievement of specific sequence recognition is not a trivial problem; many TFs recognize similar consensus DNA-binding sites and a genome can harbor thousands of consensus or near-consensus sequences, both functional and nonfunctional. Although genomic technologies have provided large-scale snapshots of TF binding, a full understanding of the mechanistic and quantitative details of specific recognition in the context of gene regulation is lacking. Here, we explore the various ways in which TFs recognizing similar consensus sites distinguish their own targets from a large number of other sequences to ensure specific cellular responses.

Specific recognition of DNA-binding sites and specific transcriptional responses: nontrivial problems

The specificity of transcription regulation in all organisms depends on the ability of TFs, the *trans* elements of the gene expression system, to interact with *cis*-regulatory regions (CRRs) of DNA. A productive transcription reaction can take place only if all the required (*cis* and *trans*) factors 'cooperate' to recruit the transcriptional machinery [1]. Prokaryotes and eukaryotes use different strategies to target TFs to specific genome locations. Bacterial TFs tend to recognize extended DNA sites, which is enough to ensure specificity in small genomes. By contrast, eukaryotic TFs typically recognize shorter sequences and therefore clustering of sites is often required to achieve specific recognition [2]. In addition, these shorter motifs can occur many times in large eukaryotic genomes, and TFs display widespread, nonfunctional binding. TF consensus binding sites can be determined either *in vitro* through a process (often called Selex) of selective enrichment starting from random double-stranded (ds)DNA oligonucleotide sequences [3,4] or *in vivo* through the identification and analysis of a series of *bona fide* CRRs [5]. The sequences in the CRRs recognized by the TFs are called response elements (REs) and their discovery *in vivo* has been fueled by the advent of genomic approaches such as chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq).

The ability of a TF to bind to a specific RE relies on biophysical interactions between the protein structure and DNA [6]. An understanding of the 'design' principles of several TF classes lies behind recent advances in precision genome editing, such as the engineering of Zn finger nucleases for genome manipulation [7]. However, another important facet of TF specificity, which we focus on here, is how a TF selects the 'appropriate' regulatory targets containing the consensus out of a very large number of identical or highly similar sequences, thus eliciting a specific cellular response to a signal.

A series of outstanding experiments over the past 40 years suggest that the majority of TF molecules in a cell are in contact with DNA/chromatin at any instant. They include classical studies involving the lacI repressor of *Escherichia coli* to more recent single molecule analyses [8–11]. The results of these studies delineate three types of binding events: (i) specific functional binding to CRRs with a direct impact on gene regulation; (ii) specific but nonfunctional binding (to CRRs or elsewhere in the genome); and (iii) nonspecific nonfunctional binding, where functionality is defined as transcription regulation. Although nonfunctional binding may influence gene expression, as suggested decades ago for the lacI repressor [8,9], in this review, we focus on how eukaryotic TFs interact with each other and with DNA to ensure specific transcriptional responses. Decades of biophysical work on prokaryotic TFs have probably led to the false perception that TF binding specificity is a 'solved problem'. In eukaryotes, this impression is strengthened by the availability of a wealth of ChIP-on-Chip and ChIP-seq data. However, we now have to combine these 'footprints' of TF binding with dynamic data to better understand the precise mechanism(s) of specific recognition of DNA involved in interaction with nucleosomes, in a crowded nuclear environment.

To bind or not to bind, that is the question

Genome-wide studies (i.e., ChIP-on-Chip and ChIP-seq) over tens of animal TFs show that they bind to a remarkably widespread, overlapping set of genomic regions. Although the most highly bound regions are generally known functional targets, many of the thousands of regions bound at lower levels may represent specific or nonspecific nonfunctional interactions. However, these results have to be tempered by the fact that ChIP provides a snapshot of CRR

Corresponding author: Veitia, R.A. (veitia.reiner@ijm.univ-paris-diderot.fr).
Keywords: response element; transcription factor; *cis*-regulation; gene expression; transcription; ChIP-seq.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2014.04.002

occupancy devoid of kinetic information. Such results depend not only on the parameters of the peak detection algorithms but also on the duration of chromatin exposure to the fixative agent, so that with long exposure times even highly transient mostly nonfunctional interactions can be captured. That said, however, published results suggest that the cellular concentrations of many TFs are high enough to drive extensive nonfunctional interactions within the accessible portions of the chromatin [12]. In line with this, several studies suggest that the numbers of TF molecules per cell in eukaryotes such as *Drosophila*, mouse, and human range between 250 and >300 000 [13,14]. Although 300 000 molecules/nuclear volume seems like a very high concentration, it is plausible that the effective concentration for productive interactions is much lower as nonfunctional binding (specific or not) may titrate the TF.

Matching the right CRR with the right TF in response to a signal is clearly a complex problem with many facets. As mentioned above, the simple presence of a binding site of a TF alone cannot account for a productive interaction in terms of transcription. A study of approximately 100 mouse TFs showed that almost half of these proteins can recognize several different sequences in addition to the known consensus sequences [15]. Thus, although ChIP-seq peaks are typically enriched in the cognate consensus motif for the TF in question, a significant proportion of peaks lack it [16–18]. For instance, sequences close to the consensus identified for the E2F family *in vitro* are present in less than 20% of the regions recognized *in vivo*. Interestingly, an intact DNA-binding domain (DBD) is required for the recognition of regions lacking the consensus [19]. This suggests the existence of indirect TF binding to DNA, either via other TFs or chromatin-bound proteins, indicating that the problem of CRR recognition is not a binary problem but rather involves several players.

The difficulty of achieving appropriate specificity is further exacerbated by the fact that many TFs belong to large families with many paralogs, which often recognize very close consensus DNA-binding sites. For instance, the Hox gene family in vertebrates is composed of approximately 40 members, all of which share a similar DBD and recognize the core sequence 5′-TAAT-3′ [20]. Of course, minor sequence differences in both the DBD and the target sites are important. For example, the replacement of the HOXA11 DBD by that of HOXA13 DBD in mice induces a homeotic transformation of a portion of the uterus into a vagina-like structure (where *Hoxa13* is normally expressed [21]). However, this phenomenon is not universal: exchanging the coding sequences of *Hoxa3* and *Hoxb3* [22] causes no major phenotypic abnormalities, suggesting that they target the same genes when expressed in the same context. This example shows that the existence of restricted (nonoverlapping) domains of expression can enhance target specificity simply by decreasing the number of potentially 'competing' factors. Matchmaking between the TF and appropriate RE can also be influenced by the chromatin environment, which is often developmentally regulated to enhance the effect of restricted domains of expression.

Thus, many if not most TFs not only interact with DNA through a consensus site but also recognize divergent sequences. Moreover, achieving TF specificity often involves interactions with other TFs binding to neighboring sites or binding to chromatin without directly contacting DNA [6]. An important lesson to be drawn from genome-wide studies is that only a small fraction of the consensus binding sequences occurring in the genome are in fact occupied by a TF in a given cell type at a given moment. One mechanism that reduces the total number of possible binding sites is through burying irrelevant consensus sequences within compact chromatin, rendering those sequences unavailable. Below, we discuss how the chromatin context modulates the repertoire of targets recognized *in vivo*.

Chromatin and DNA accessibility: reducing complexity

A recent study on five TF regulators of early embryonic anterior–posterior patterning in *Drosophila* showed that predictions based on consensus sequences and *in vitro* protein–DNA affinities poorly correlate with *in vivo* measurements. Not surprisingly, restricting the prediction of binding to regions of open chromatin dramatically improved the results [23]. The authors optimistically suggest that a combination of chromatin accessibility data and simple computational models should be able to successfully predict TF binding. However, several genome-wide studies have shown that there is a high degree of overlap in the genomic regions recognized by functionally unrelated TFs (see [13] and references therein). For instance, data on 21 *Drosophila* TFs, with DBDs belonging to 11 different families, show that approximately 90% of the 300 most highly bound regions lying within open chromatin are occupied with high confidence by eight or more TFs. This preference for accessible chromatin regions is probably due to the high intranuclear concentrations of some TFs coupled to broad DNA-binding specificities [24] (Figure 1).

Other evidence suggests that chromatin conformation and accessibility can regulate gene expression on much finer scales [25]. For instance, hepatic albumin (*Alb1*) gene expression is regulated by an element involving TF-binding sites occupied by FoxA and Gata factors [26,27]. Structural studies of this element have shown that, even when occupied by nucleosomes, it is recognized by FOXA1, which modulates nucleosome positioning *in vitro* [28]. The forkhead domain of FOXA1 has structural similarities with the histones H1 and H5, and FOXA1 can replace one of these histones in the *Alb1* regulatory element and prevent chromatin compaction [29,30]. Interestingly, removal of *Foxa1* and *Foxa2* genes is embryonic lethal and prevents liver development but their deletion in differentiated hepatocytes does not cause any loss of cell identity. This observation shows that FoxA factors are required for the activation but not for the maintenance of hepatocyte expression of a subset of genes [31,32]. Thus, it has been suggested that FOXA1 and FOXA2 could function as pioneer factors that mark and make accessible regulatory elements in early hepatocyte differentiation [25,33] (Figure 2).

In summary, the reduced complexity achieved by locking DNA in inaccessible heterochromatin contributes

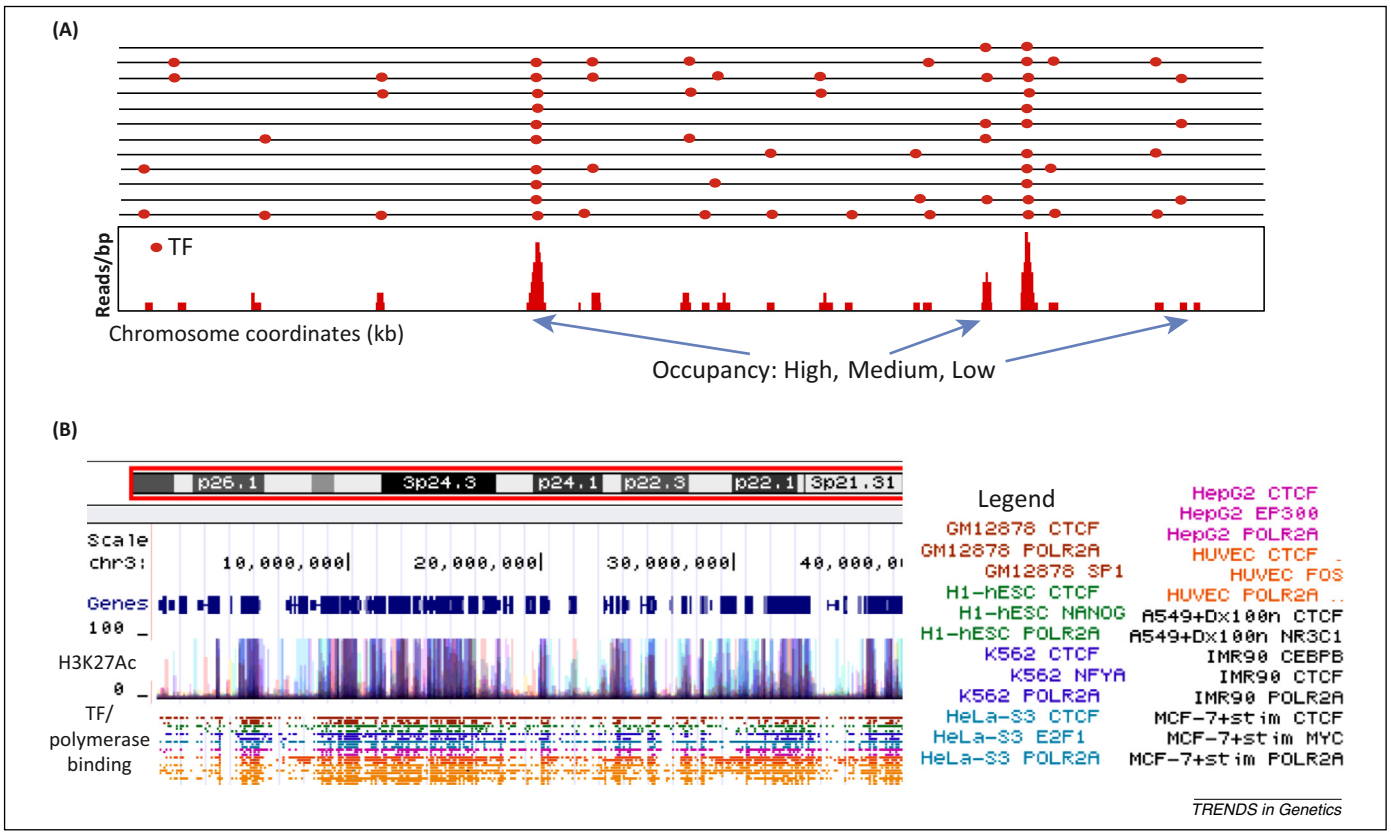

Figure 1. Open chromatin and TF binding. **(A)** Idealized representation of the interaction between a TF (red dots) and several DNA molecules (black horizontal lines), and the net result of this process is revealed as peaks in a ChIP-seq experiment (red histograms). Note the presence of DNA sites with different degrees of occupancy. In a ChIP-seq experiment, this depends on the thresholds retained for data analysis. **(B)** Typical ENCODE results summarizing ChIP data of accessible chromatin (histone H3K27 acetylation) and TF and polymerase binding for the human chromosome 3 in various cell lines of different origins (UCSC Genome Browser at http://genome.ucsc.edu/) [61]. Note the correlation between accessible chromatin and binding of various TFs (CTCF, MYC, NANOG, NFYA, E2F1, P300, FOS, CEBP, NR3C1/glucocorticoid receptor). This correlates with studies discussed in the text reporting a high degree of overlap in the genomic regions that are recognized by functionally related and unrelated TFs. Abbreviations: TF, transcription factor; ChIP-seq, chromatin immunoprecipitation coupled to high-throughput sequencing.

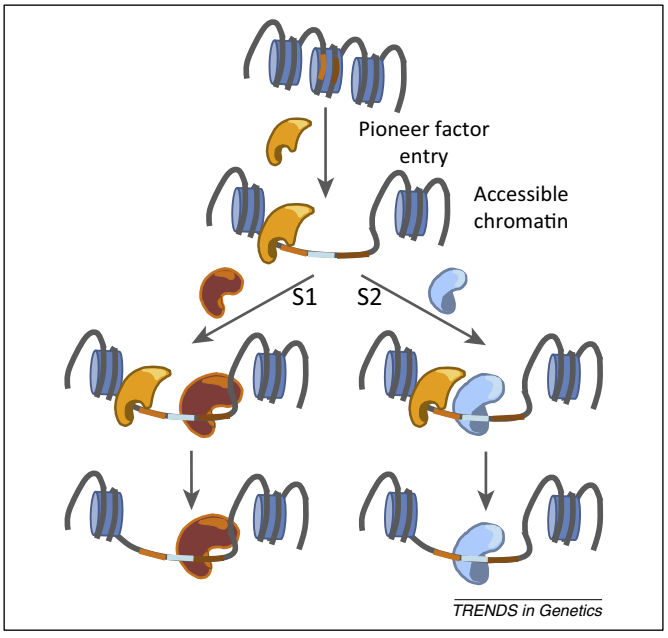

Figure 2. Pioneer factors at work. Several factors are required for the activation but not necessarily for the maintenance of gene expression by altering chromatin conformation (and accessibility) through moving or removing nucleosomes. Such factors can remain bound to or leave the relevant DNA. The outcome of this process is the exposure of REs for factors involved in mediating rapid responses to environmental or internal signals (S1 and S2). Abbreviation: RE, response element.

to, but cannot fully account for, specific CRR recognition in response to a particular signal. In the following section, we discuss biochemical principles relevant in this context.

Cooperation and cooperativity of transcription factors

As discussed above, highly expressed TFs can directly bind to DNA, although ultimately this depends on a trade-off among effective concentrations, interaction constants, and number of potential binding sites. To win on all accounts, some TFs contain multiple DBDs. This increases the length and diversity of the recognizing and recognized sequences [34]. In thermodynamic terms, the Gibbs free energy of the interaction between the TF and DNA involves the sum of the free energies of the interactions of each DBD with its site minus the energy spent in conformational changes on both the protein and DNA sides. That is, $\Delta G^{\circ}_{TF/DNA} = \Delta G^{\circ}_{DBD1/DNA} + \Delta G^{\circ}_{DBD2/DNA} - \delta$ (here, δ will be neglected compared to the binding energies). Using the known $\Delta G^{\circ} = -RT\ln K$ formula, we can estimate that the dissociation constant $K_{TF/DNA}$ is of the order of $K_{DBD1/DNA}K_{DBD2/DNA}$. The huge difference between $K_{TF/DNA}$ and $K_{DBD/DNA}$ is the elementary basis underlying the discrimination between a random DNA sequence resembling a half-site, which will be recognized with low affinity, and a true bipartite site. This general principle applies to

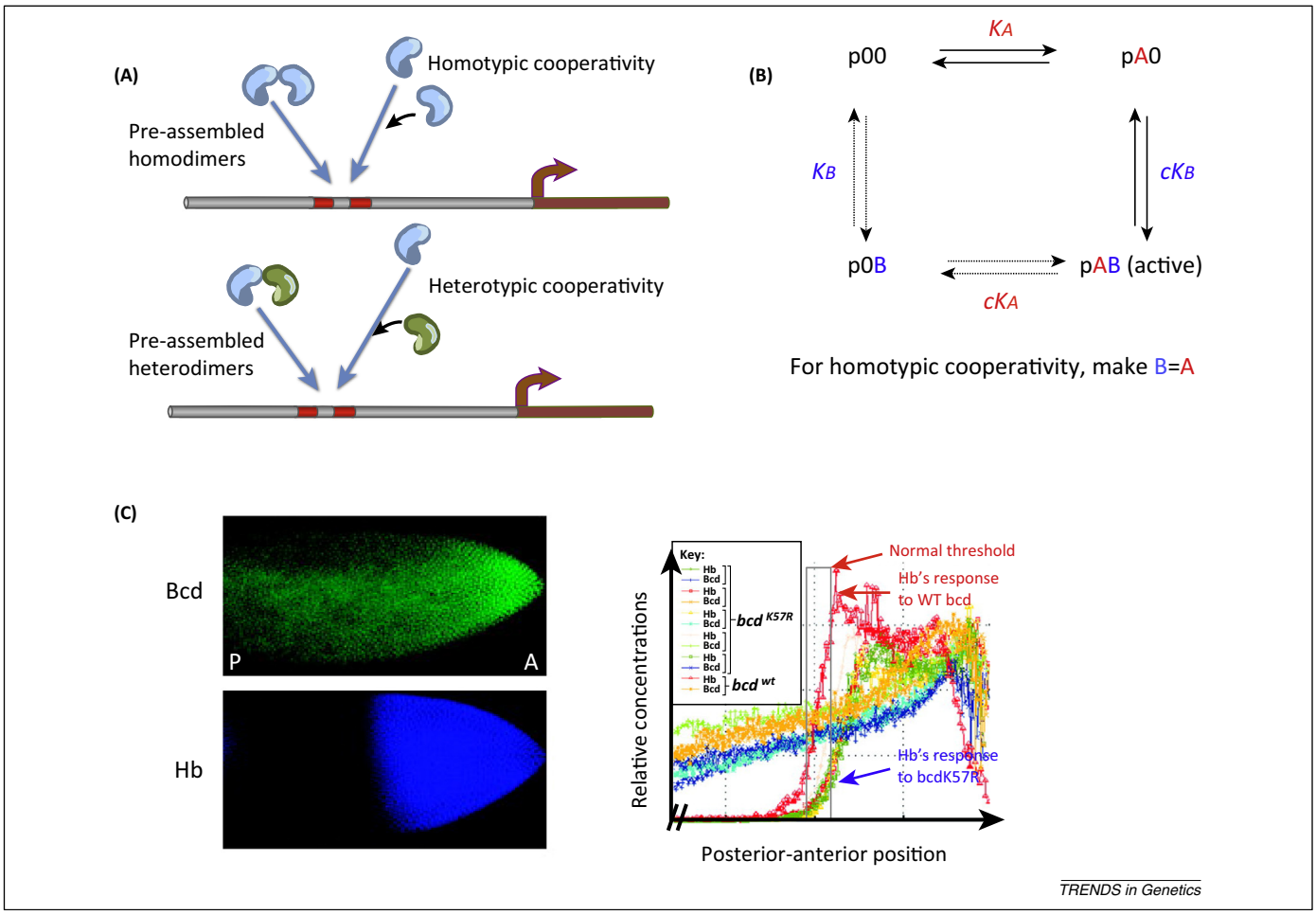

Figure 3. Specific DNA recognition and specific transcriptional response. **(A)** *Cis*-regulatory specificity provided by the recognition of the red-binding sites by a pre-assembled TF homodimer (blue proteins) or by homotypic cooperativity (which implies sequential binding and unbinding, upper panel). Specificity provided through recognition of a pre-assembled TF heterodimer or by heterotypic cooperativity (lower panel). **(B)** Mass action law representation of the recognition of a promoter with two binding sites for the TFs A and B. K_A and K_B are the individual dissociations constants of A and B and DNA, respectively. Factor c accounts for cooperation/cooperativity. Species p00, pA0, p0B, and pAB are promoter regions (of the same gene) bound by zero, one, or two TF molecules. By virtue of synergy, only the promoters with two occupied binding sites are assumed to contribute to transcription. To represent homotypic cooperativity and synergy, simply make A=B. **(C)** Fluorescence immunostaining for the TFs Bcd (bicoid) and Hb (hunchback) in control *Drosophila* embryos (displayed in the posterior–anterior direction). Note the sharp Hb expression boundary generated in response to the smooth Bcd gradient (the embryos displayed are truncated in their P part). In the right panel, the relative Hb signal is plotted as a function of the distance along the P–A axis. The curves represent expression data for individual embryos. Note the sigmoidal response of Hb transcription to the gradient of Bcd and how the bcd cooperativity mutant bcdK57R requires higher concentrations to elicit the same response level as the wild type. Reproduced and modified, with permission, from [46]. Copyright (2005) National Academy of Sciences, U.S.A. Abbreviation: TF, transcription factor.

the binding of atypical E2F proteins, which contain duplicated DBDs and Zn fingers (the number of which determines the strength of the binding) [7,35].

Another factor that contributes to the specific recognition of regulatory elements by a TF is the ability of the latter to interact with other proteins that bind to neighboring DNA/chromatin sites. The simplest illustration of this mechanism is provided by eukaryotic TFs that bind to DNA as dimers or higher order structures [36] (Figure 3). TF dimerization synergistically increases the binding affinity relative to the interaction with each of the monomers alone as well as increases the effective length of the RE recognized, which decreases the probability of its random occurrence in the genome [37]. This is seen for steroid and other nuclear receptors (NRs). In this family, the distance and relative orientation of the two half-sites recognized by the interacting monomers ensure discrimination among the targets of estrogen receptors (ERs), vitamin D receptors, or retinoic acid (RA) receptors. Specifically, the ERs can form homodimers or heterodimers that have the highest affinity for a palindromic sequence whose half-sites are spaced 3 bp apart [38]. By contrast, receptors of thyroid hormones, vitamin D, and RA mainly form heterodimers with the retinoid X receptor (RXR) and recognize direct repeats of the half-sites [39]. The spacing between the half-sites enhances target discrimination (3 bp for the vitamin D receptor, 4 bp for the thyroid hormone receptor, and 5 bp for the RA receptor) [40]. Interestingly, despite the fact that ERα and ERβ recognize similar consensus sites, genome-wide ChIP-on-Chip location experiments show that a large proportion of binding sites (30–70%) is specific to each receptor. This is probably achieved through interactions with alternative partners [41,42]. Recent studies have shown that, in the case of factors that usually form dimers, the presence of binding sites for other partners can compensate for the absence of a complete (bipartite) binding site. For instance, the genome-wide location of ERα shows the existence of a large number of targets with only

half-sites. An enrichment in consensus sequences of ERα partners such as activator protein 1 (AP-1) or FOXA1 are consistently found close to these half-sites [43], indicating that such partners compensate for the low affinity between ER and its half-sites.

Many TFs fail to specifically bind to DNA as monomers because of their low concentrations *in vivo* and the low concentration of CRRs, which render the process of finding the latter inefficient and lengthy. This problem is alleviated through canonical or homotypic cooperativity (Figure 3). Cooperativity refers to the facilitation of DNA binding of one TF molecule by an identical molecule already bound to DNA or, more likely in crowded nuclear conditions, to a slower release from DNA of a TF that is interacting with an identical molecule. However, even with cooperativity, lowly expressed TFs can lead to intermittent target gene expression and to stochastic fluctuations of the amount of gene product over time [44]. Cooperativity gives rise to threshold effects that help distinguish between random fluctuations of TF concentration (due to natural expression noise) and a response to a true environmental or developmental signal (increasing, e.g., TF synthesis). A classic example is provided by the sharp response of the transcription of *hunchback* to the TF Bicoid, whose expression level follows a concentration gradient along the anterior–posterior axis of the *Drosophila* embryo [45,46] (Figure 3). The idea of cooperativity can be extended to include interactions between different TFs (i.e., heterotypic cooperativity). Indeed, the interaction of a TF with cofactors upon binding to DNA brings the same advantages as discussed above. Similar physicochemical principles also hold for the interaction between DNA-bound TFs and the transcriptional machinery. If only a single TF-binding site is occupied, the interaction between the TF and the transcriptional machinery will be disproportionately weaker than if several bound TF molecules attract the transcriptional machinery. Note that in this case the TFs do not necessarily have to interact with each other. This phenomenon, called transcriptional synergy, also helps discriminate a CRR (loaded with all the required factors) from randomly occurring isolated binding sites (Figure 4) [47,48]. To illustrate these points, we can consider the simple situation of a promoter containing two binding sites for the TFs A and B. We can describe its transcriptional response to the concentrations of A and B with the bivariate Hill function in Equation 1,

$$\text{Normalized Transcription} = \frac{[A][B]}{cK_AK_B + [A][B]} \qquad [1]$$

where cK_AK_B is a product of the individual dissociations constants of A and B and DNA, respectively, and c ($0<c\leq1$) is an interaction constant accounting for cooperation/cooperativity. The higher K is the weaker the interaction is. The exponents of A and B are 1 because there is only one binding site for each factor per promoter. Equation 1 involves two approximations: (i) [A] and [B] refer, properly speaking, to nonspecifically bound and truly free TF, but they will be considered as good estimates of the respective total concentrations (see Figure 6 in [48] for the validity of this approximation); and (ii) by virtue of transcriptional synergy, only the promoter with two occupied binding sites

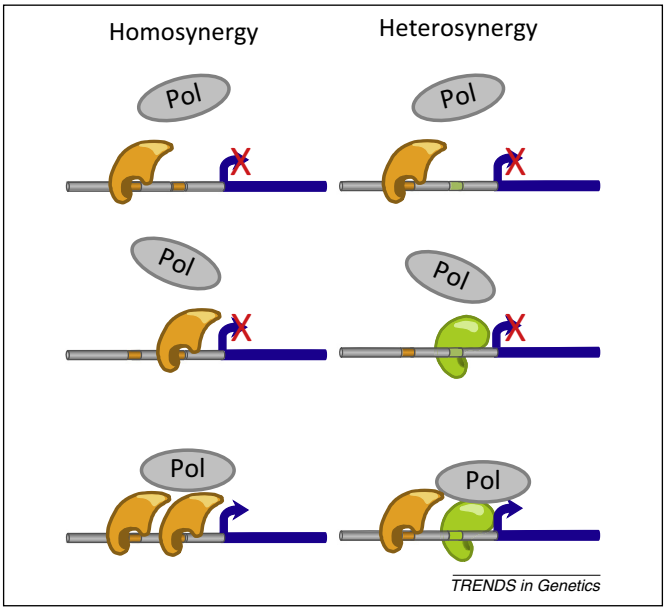

Figure 4. Transcriptional synergy. When only one TF-binding site in a CRR is occupied, its interaction with the transcriptional machinery is much weaker than when several bound TF molecules attract the transcriptional machinery. This phenomenon is called transcriptional synergy: homosynergy if the cooperating TFs are identical or heterosynergy if they are different. This phenomenon also helps discriminate between functional CRRs and randomly occurring isolated binding sites that, occupied or not, will not lead to productive transcription. Abbreviations: TF, transcription factor; CRR, *cis*-regulatory region.

is assumed to contribute to transcription (see [48] for details). That is, the interaction between DNA-bound A or B and the transcriptional machinery can be neglected compared to the attraction exerted by both A and B together on the latter (i.e., linear terms disappear from the equation). By construction, in the absence of one factor (concentration = 0), the normalized transcription will be 0. This equation shows why the cooperation between A and B is simply a form of (heterotypic) cooperativity. Indeed, when both factors A and B have similar concentrations and dynamics (i.e., A = B), Equation 1 reduces to the classical Hill function that represents textbook cooperativity for a promoter with two identical binding sites.

$$\text{Normalized Transcription} = \frac{[A]^2}{cK^2 + [A]^2} = \frac{[A]^2}{K'^2 + [A]^2} \qquad [2]$$

The Hill function written in this form is an approximation that can be generalized to any number of binding sites by changing the exponent. According to Equation 2, the concentration of A required to elicit 50% of the normalized transcription is K′ (which is an average dissociation constant). This shows that the 'dynamic range' of the sigmoidal transcriptional response involves concentrations not too far below or above K′. Thus, if K′ is in the nM range, the concentrations of A should also be in the nM range. pM or μM concentrations would be either too low (no CRR occupancy) or too high (saturated CRR), respectively. This demonstrates the need for the proportionality between binding affinity and TF concentration suggested above. Equation 2 also helps understand how small differences in the interaction between a CRR with its cognate binding factor, say *A*, or with a competitor paralog *a* lead to strong

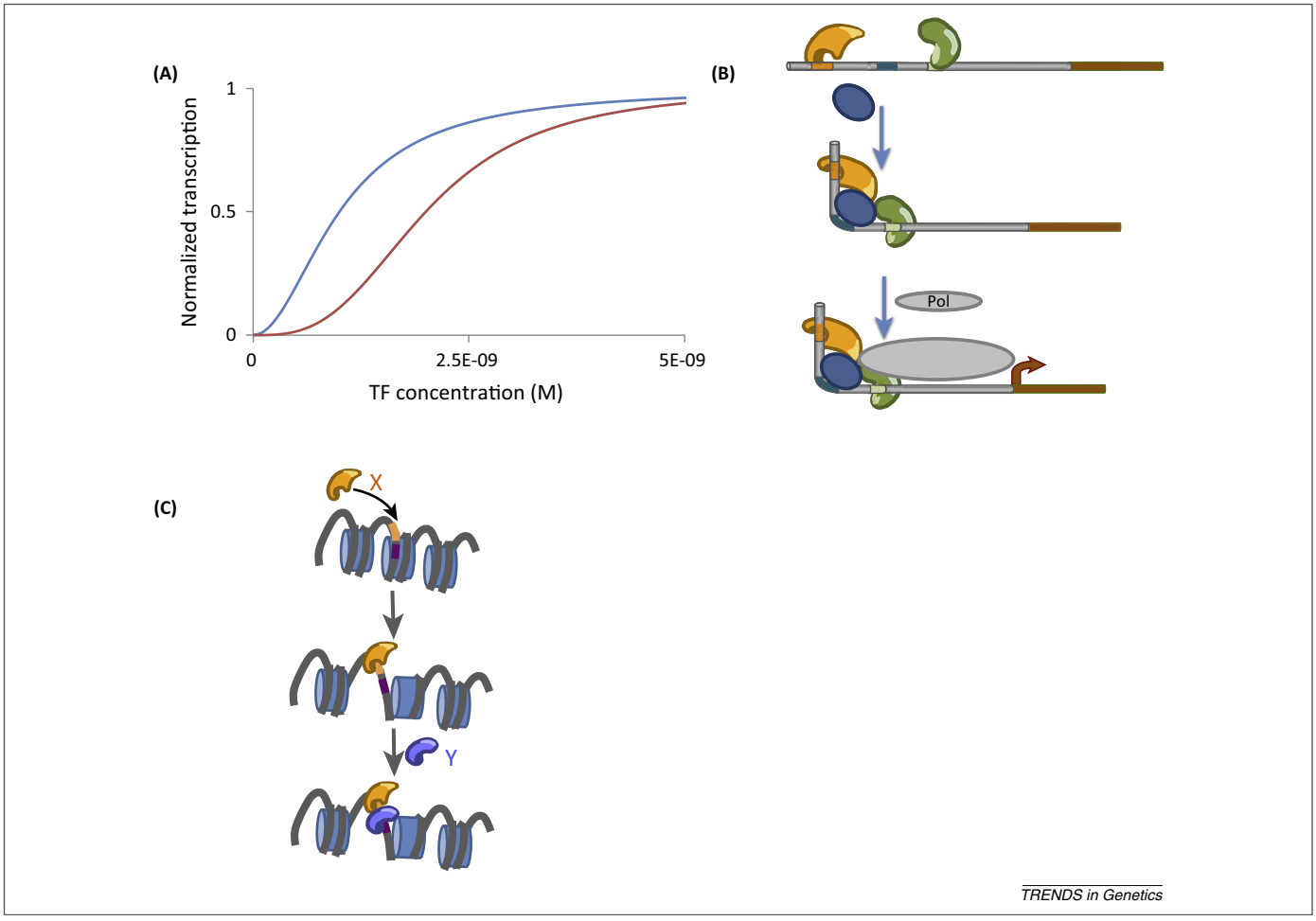

Figure 5. Cooperativities. **(A)** Normalized transcriptional response as a function of TF concentration (according to Equation 2, with three binding sites, blue curve). Note how doubling K′ (representing the interaction with the less specific paralog) shifts the curve to the right (red curve). This implies that a higher concentration is required to obtain the same response. Homotypic cooperativity is possible only when a CRR has the appropriate architecture that implies the existence of multiple binding sites for the TF in question (which can hardly appear randomly). **(B)** Cooperation fostered by partners that mediate interactions between two DNA-binding proteins by bending DNA [62]. For instance, the enhancer of the interferon β (*IFN-β*) gene contains sites recognized by AP-1 factors, NFκB, and IRF but cannot be activated without the intervention of the architectural factor HMG-I(Y), which bends DNA and facilitates the interaction between TFs bound or binding to the regulatory element [63]. Here, the blue protein mediates the interaction between two DNA-binding proteins by bending DNA. **(C)** At least *in vitro*, binding side-by-side of several TFs on DNA in interaction with nucleosomes is inherently cooperative. This is so because the interaction of the first-arriving factor with its site can destabilize the DNA–nucleosome interaction and increase the exposure of adjacent sites [64,65]. Abbreviations: TF, transcription factor; CRR, *cis*-regulatory region; AP-1, activator protein 1; NFκB, nuclear factor κB; IRF, interferon regulatory factor.

discrimination. For instance, for three binding sites, if $K'_{a/DNA}$ is just two times (weaker) that of $K'_{A/DNA}$, when A has elicited 50% of maximal transcription, the same concentration of a will elicit only 10% (Figure 5). This discrimination power increases in a multiplicative manner with the number of binding sites involved and will be further increased by differences in the cooperativity level of A on its cognate CRR (with respect to that of a on the same CRR). Thus, although the highest affinity sequence can be highly similar for paralogs, subtle mismatches found in real targets lead to differences in K′. For example, most of the reported p53-binding sites have affinity values of up to 1.5 log$K$ units weaker than the highest affinity sequence. This means that various targets can be triggered by different concentrations of p53 depending on the specific CRR architectures. In the context of our discussion, differential binding penalties due to sequence mismatches will result in the selection of nonoverlapping sets of binding sites by different p53 family members [49].

Coming back to Equation 1, it also helps understand how small differences in the interactions between partners (say, A and B) with each other and with their cognate CRRs translate into strong differences of occupancy with respect to what happens for the (weaker) competitor paralogs a and b. Thus, even when the members of a TF family recognize similar REs, the existence of cooperative and/or synergistic interactions can help discriminate between functional/productive binding from nonfunctional binding (specific or not). Heterotypic interactions underlie a combinatorial strategy that explains to a wide extent the specificity of TFs in target recognition, as well as allows cells to respond appropriately to internal or external signals, as discussed in the following section.

Combinatorial signal transduction and heterotypic cooperativity

Cells are continuously sensing cues from the external and internal environment. The signaling cascades involved in

these processes have been perceived for decades as linear pathways. However, there are so many TFs that seem to respond to the same signal transducer (say, the kinases AKT or PKA) that it is difficult to conceive how an unambiguous correspondence between signals and outputs can be achieved. This can be explained by the existence of combinatorial interactions between TFs that ensure recognition of specific target subsets in response to a particular signal. This implies that the signal in question must trigger a 'combination' of transduction pathways that will modulate the required TFs. As an idealized example, consider a signal S1 that is transduced by (at least) two parallel pathways that activate the TFs A and B (and probably others). TFs A and B recognize a subset of promoter sets having the appropriate REs. To allow discrimination from a signal S2, transduction of the latter must involve another (not necessarily exclusive) combination of pathways (Figure 6). The corollary of this example is that the right transcriptional output is elicited only when several signaling pathways (responding to a single input or multiple inputs) ensure binding of distinct TFs at specific promoter subsets. Important in this respect are post-translational modifications (PTMs) that can modulate protein–protein interactions, shift DNA-binding affinities, and/or site preferences. A complete discussion of this topic along with several examples is available elsewhere [50].

Differential partnership and chromatin-based direct or indirect effects explain why many signaling pathways have different effects depending on the cell types, despite their signaling through a limited number of common effectors. For instance, a STAT3 genome-wide location study in macrophages, embryonic stem, $CD4^+$ T, and AtT-20 cells has shown two different binding modes. The first one is cell type-independent and involves conserved STAT3-binding sites that regulate the function of STAT3 itself and cell growth. The other mode involves different transcriptional modules coordinated by STAT3 to drive cell type-specific transcriptional programs in association with cell type-specific TFs [51].

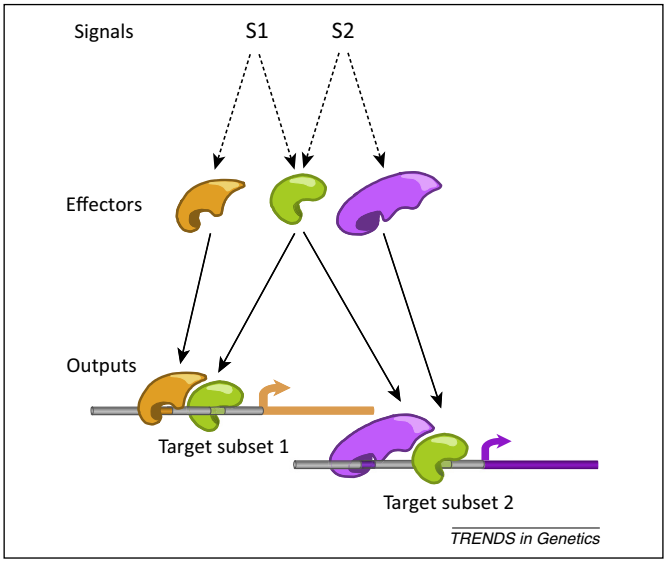

Figure 6. Encoding and decoding signals. S1 and S2 are signals that are transduced by different combinations of pathways (broken arrows). In this idealized example, the specificity of the output with respect to the signal is obtained by the combinatorial occupancy of target promoters. At the promoter level, this process takes advantage of heterotypic cooperativity and transcriptional synergy as discussed in the text. Signal transduction can involve the apposition of PTM marks on their target TFs. Examples are provided by the activation of *even-skipped* in *Drosophila* muscle precursor cells, which depends on the Wingless, TGFβ and RTK signaling cascades activating their downstream effectors Tcf, Mad, and Pnt [66], and the activation of the flocculin *FLO11* gene in *Saccharomyces cerevisiae*, which can only take place if TOR, MAPK, and PKA pathways are activated and activate the TFs Tec1 and Flo8 [67]. Abbreviations: TF, transcription factor; PTMs, post-translational modifications; TGFβ, transforming growth factor β; RTK, receptor tyrosine kinase; TOR, target of rapamycin; MAPK, mitogen-activated protein kinase; PKA, protein kinase A.

Concluding remarks

Specific binding by TFs is influenced by chromatin accessibility and cooperative and synergistic interactions between TFs and with the transcription machinery. This explains the differences observed between *in vitro* and *in vivo* binding. As previously outlined, many genome-wide location analyses have detected thousands of accessible CRRs in various genomes. However, these snapshots cannot tell whether TFs bind to these sites because they are accessible or if, on the contrary, TF binding renders chromatin accessible. This question is still under debate but some examples mentioned above suggest that both views might be valid depending on the TFs and cell types [52–54].

Another important question still to be answered is whether nonfunctional TF binding is truly 'nonfunctional'. In this context, it has been proposed that TF dosage changes, such as those induced by an allele deletion or duplication can be buffered, precisely because TFs interact with nonfunctional binding sites. Indeed, simple simulations show that depending on the balance between functional and nonfunctional binding, the concentration of a TF involved in productive transcription reactions changes by a factor less than two when total TF concentration is either halved or doubled [55]. This requires, as predicted above, the coevolution of TF concentrations, the number of functional and nonfunctional sites, and the relevant interaction constants. Consistent with these ideas, genome-wide analyses in plants show that gene duplicates can be retained following tetraploidy simply because their *cis*-regulatory sequences are the targets of dosage-sensitive TFs [56,57]. On the experimental side, a recent study using the LacI system in bacteria shows that the presence of competing binding sites produces a titration-like phenomenon in the transition from conditions where LacI is limiting compared to the number of binding sites to a regime in which the former is in excess [58]. Similar experiments are possible in eukaryotes to explore to what extent changes in the amount of a TF can be buffered by nonfunctional binding.

Another open question regarding the recruitment of TFs to regulatory regions centers around nuclear architecture. Many studies have provided evidence that transcription takes place within discrete nuclear factories [59]. Not surprisingly, coregulated genes are often localized in close proximity, but are not physically associated, which allows the rapid triggering of their coexpression in response to a signal [60]. This model suggests that TFs do not attract the transcription machinery (which is assumed to already be at the factories) to the CRRs as simplistically stated above, but the other way round. However, it is not completely clear if these views are mutually exclusive or if the reality is in-between these two models. This deserves further molecular exploration and mathematical modeling.

Finally, it is worth noting that the enormous amounts of genome-wide TF location data provided by initiatives such as the ENCODE consortium [61] is only the starting point for refining the details of TF binding. As previously stated, the dynamic dimension is missing in most of these 'en masse' approaches. There is still much work to be done in understanding the subtle dynamics that fine-tune gene regulation.

Acknowledgments

Work in the laboratory is supported by Ligue Nationale contre le Cancer, the University Paris Diderot-Paris, the CNRS (Centre National de la Recherche Scientifique), and the ANR (Agence Nationale de la Recherche). We thank the editor and two anonymous referees for their helpful comments.

References

1 Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763
2 Wunderlich, Z. and Mirny, L.A. (2009) Different gene regulation strategies revealed by analysis of binding motifs. *Trends Genet.* 25, 434–440
3 Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510
4 Jolma, A. *et al.* (2010) Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* 20, 861–873
5 Furey, T.S. (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat. Rev. Genet.* 13, 840–852
6 Brent, M.M. *et al.* (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure* 16, 1407–1416
7 Klug, A. (2010) The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q. Rev. Biophys.* 43, 1–21
8 Lin, S. and Riggs, A.D. (1975) The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell* 4, 107–111
9 Kao-Huang, Y. *et al.* (1977) Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: measurement of DNA-bound *Escherichia coli* lac repressor in vivo. *Proc. Natl. Acad. Sci. U.S.A.* 74, 4228–4232
10 Phair, R.D. *et al.* (2004) Global nature of dynamic protein–chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.* 24, 6393–6402
11 Elf, J. *et al.* (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316, 1191–1194
12 Fisher, W.W. *et al.* (2012) DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21330–21335
13 Biggin, M.D. (2011) Animal transcription networks as highly connected, quantitative continua. *Dev. Cell* 21, 611–626
14 Simicevic, J. *et al.* (2013) Absolute quantification of transcription factors during cellular differentiation using multiplexed targeted proteomics. *Nat. Methods* 10, 570–576
15 Badis, G. *et al.* (2009) Diversity and complexity in DNA recognition by transcription factors. *Science* 324, 1720–1723
16 Robertson, G. *et al.* (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* 4, 651–657
17 Valouev, A. *et al.* (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* 5, 829–834
18 Li, X. *et al.* (2008) Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* 6, e27
19 Rabinovich, A. *et al.* (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. *Genome Res.* 18, 1763–1777
20 Gehring, W.J. *et al.* (1994) Homeodomain–DNA recognition. *Cell* 78, 211–223
21 Zhao, Y. and Potter, S.S. (2001) Functional specificity of the Hoxa13 homeobox. *Development* 128, 3197–3207
22 Greer, J.M. *et al.* (2000) Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature* 403, 661–665
23 Kaplan, T. *et al.* (2011) Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early *Drosophila* development. *PLoS Genet.* 7, e1001290
24 MacArthur, S. *et al.* (2009) Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 10, R80
25 Zaret, K.S. and Carroll, J.S. (2011) Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* 25, 2227–2241
26 Bossard, P. and Zaret, K.S. (1998) GATA transcription factors as potentiators of gut endoderm differentiation. *Development* 125, 4909–4917
27 Gualdi, R. *et al.* (1996) Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev.* 10, 1670–1682
28 McPherson, C.E. *et al.* (1993) An active tissue-specific enhancer and bound transcription factors existing in a precisely positioned nucleosomal array. *Cell* 75, 387–398
29 Cirillo, L.A. *et al.* (1998) Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* 17, 244–254
30 Clark, K.L. *et al.* (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412–420
31 Lee, C.S. *et al.* (2005) The initiation of liver development is dependent on Foxa transcription factors. *Nature* 435, 944–947
32 Li, Z. *et al.* (2009) Foxa1 and Foxa2 regulate bile duct development in mice. *J. Clin. Invest.* 119, 1537–1545
33 Magnani, L. *et al.* (2011) Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet.* 27, 465–474
34 Charoensawan, V. *et al.* (2010) Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* 38, 7364–7377
35 Lammens, T. *et al.* (2009) Atypical E2Fs: new players in the E2F transcription factor family. *Trends Cell Biol.* 19, 111–118
36 Amoutzias, G.D. *et al.* (2008) Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* 33, 220–229
37 Georges, A.B. *et al.* (2010) Generic binding sites, generic DNA-binding domains: where does specific promoter recognition come from? *FASEB J.* 24, 346–356
38 Gronemeyer, H. (1991) Transcription activation by estrogen and progesterone receptors. *Annu. Rev. Genet.* 25, 89–123
39 Yu, V.C. *et al.* (1991) RXRβ: a coregulator that enhances binding of retinoic acid, thyroid hormone, and vitamin D receptors to their cognate response elements. *Cell* 67, 1251–1266
40 Umesono, K. *et al.* (1991) Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell* 65, 1255–1266
41 Liu, Y. *et al.* (2008) The genome landscape of ERα- and ERβ-binding DNA regions. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2604–2609
42 Charn, T.H. *et al.* (2010) Genome-wide dynamics of chromatin binding of estrogen receptors α and β: mutual restriction and competitive site selection. *Mol. Endocrinol.* 24, 47–59
43 Joseph, R. *et al.* (2010) Integrative model of genomic factors for determining binding site selection by estrogen receptor-α. *Mol. Syst. Biol.* 6, 456
44 Cook, D.L. *et al.* (1998) Modeling stochastic gene expression: implications for haploinsufficiency. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15641–15646
45 Crauk, O. and Dostatni, N. (2005) Bicoid determines sharp and precise target gene expression in the *Drosophila* embryo. *Curr. Biol.* 15, 1888–1898
46 Lebrecht, D. *et al.* (2005) Bicoid cooperative DNA binding is critical for embryonic patterning in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13176–13181
47 Carey, M. (1998) The enhanceosome and transcriptional synergy. *Cell* 92, 5–8
48 Veitia, R.A. (2003) A sigmoidal transcriptional response: cooperativity, synergy and dosage effects. *Biol. Rev. Camb. Philos. Soc.* 78, 149–170

49 Brandt, T. *et al.* (2009) Conservation of DNA-binding specificity and oligomerisation properties within the p53 family. *BMC Genomics* 10, 628
50 Benayoun, B.A. and Veitia, R.A. (2009) A post-translational modification code for transcription factors: sorting through a sea of signals. *Trends Cell Biol.* 19, 189–197
51 Hutchins, A.P. *et al.* (2013) Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. *Nucleic Acids Res.* 41, 2155–2170
52 Shivaswamy, S. *et al.* (2008) Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS Biol.* 6, e65
53 Zawadzki, K.A. *et al.* (2009) Chromatin-dependent transcription factor accessibility rather than nucleosome remodeling predominates during global transcriptional restructuring in Saccharomyces cerevisiae. *Mol. Biol. Cell* 20, 3503–3513
54 He, H.H. *et al.* (2010) Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* 42, 343–347
55 Veitia, R.A. *et al.* (2013) Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet.* 29, 385–393
56 Schnable, J.C. *et al.* (2011) Dose-sensitivity, conserved non-coding sequences, and duplicate gene retention through multiple tetraploidies in the grasses. *Front. Plant Sci.* 2, 2
57 Birchler, J.A. and Veitia, R.A. (2011) Protein–protein and protein–DNA dosage balance and differential paralog transcription factor retention in polyploids. *Front. Plant Sci.* 2, 64
58 Brewster, R.C. *et al.* (2014) The transcription factor titration effect dictates level of gene expression. *Cell* 156, 1312–1323
59 Edelman, L.B. and Fraser, P. (2012) Transcription factories: genetic programming in three dimensions. *Curr. Opin. Genet. Dev.* 22, 110–114
60 Banos, A. *et al.* (2013) Stochastic responses are not left to pure 'chance'. *Cell* 155, 499–502
61 Wang, J. *et al.* (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* 22, 1798–1812
62 Merika, M. and Thanos, D. (2001) Enhanceosomes. *Curr. Opin. Genet. Dev.* 11, 205–208
63 Yie, J. *et al.* (1999) The role of HMG I(Y) in the assembly and function of the IFN-β enhanceosome. *EMBO J.* 18, 3074–3089
64 Adams, C.C. and Workman, J.L. (1995) Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol. Cell. Biol.* 15, 1405–1421
65 Steger, D.J. and Workman, J.L. (1997) Stable co-occupancy of transcription factors and histones at the HIV-1 enhancer. *EMBO J.* 16, 2463–2472
66 Pires-daSilva, A. and Sommer, R.J. (2003) The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* 4, 39–49
67 Vinod, P.K. *et al.* (2008) Integration of global signaling pathways, cAMP-PKA, MAPK and TOR in the regulation of FLO11. *PLoS ONE* 3, e1663