

## REVIEW

# *In vitro* DNA-binding profile of transcription factors: methods and new insights

Jinke Wang, Jie Lu, Guangming Gu and Yingxun Liu

The State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, People's Republic of China

(Correspondence should be addressed to J Wang; Email: wangjinke@seu.edu.cn)

### Abstract

The DNA-binding specificity of transcription factors (TFs) has broad impacts on cell physiology, cell development and in evolution. However, the DNA-binding specificity of most known TFs still remains unknown. The specificity of a TF protein is determined by its relative affinity to all possible binding sites. In recent years, the development of several *in vitro* techniques permits high-throughput determination of relative binding affinity of a TF to all possible k bp-long DNA sequences, thus greatly promoting the characterization of DNA-binding specificity of many known TFs. All DNA

sequences that can be bound by a TF with various binding affinities form their DNA-binding profile (DBP). The DBP is important to generate an accurate DNA-binding model, identify all DNA-binding sites and target genes of TFs in the whole genome, and build transcription regulatory network. This study reviewed these techniques, especially two master techniques: double-stranded DNA microarray and systematic evolution of ligands by exponential enrichment in combination with parallel DNA sequencing techniques (SELEX-seq).

*Journal of Endocrinology* (2011) **210**, 15–27

### Introduction

Transcription factors (TFs) are central to almost every fundamental cellular process (Latchman 2008, Ladunga 2010) and account for ~5–10% of genes in eukaryotes (Reece-Hoyes *et al.* 2005, Adryan & Teichmann 2006, Ho *et al.* 2006, Lee *et al.* 2007). In mammalian TFs, approximately more than 700 were identified to be DNA-binding TFs (Messina *et al.* 2004, Lee *et al.* 2007); they bind with the TF binding sites (TFBSs) in the genome and regulate the expression of their target genes. Differential gene expression is achieved in part by the interaction of these DNA-binding regulatory TFs with various TFBSs.

DNA-binding specificity of TFs plays essential roles in cell physiology, cell development and organism evolution. However, the DNA-binding specificity of most known TFs still remains unknown. The specificity of a TF protein is determined by its relative affinity to all possible binding sites. In recent years, the development of several *in vitro* techniques permits high-throughput determination of relative binding affinity of a TF to all possible k bp-long DNA sequences. These high-throughput techniques greatly promoted the characterization of DNA-binding specificity of many known TFs.

All DNA sequences that can be bound *in vitro* by a TF with various binding affinities form its *in vitro* DNA-binding profile (DBP). The *in vitro* DBP provides an *in vitro*

sequence-recognition profile of a TF. The *in vitro* DBP is important in the generation of an accurate DNA-binding model (such as position weight matrix, PWM), identification of all DNA-binding sites and target genes of TFs in the whole genome, and construction of transcription regulatory network. The *in vitro* DBP also has promising applications in transcription therapy, a therapeutic strategy using TFs as targets for disease therapy (Redell & Tweardy 2005, Frank 2009, Li & Sethi 2010). Therefore, the studies on *in vitro* DBP of TFs attract increasing attention in this field.

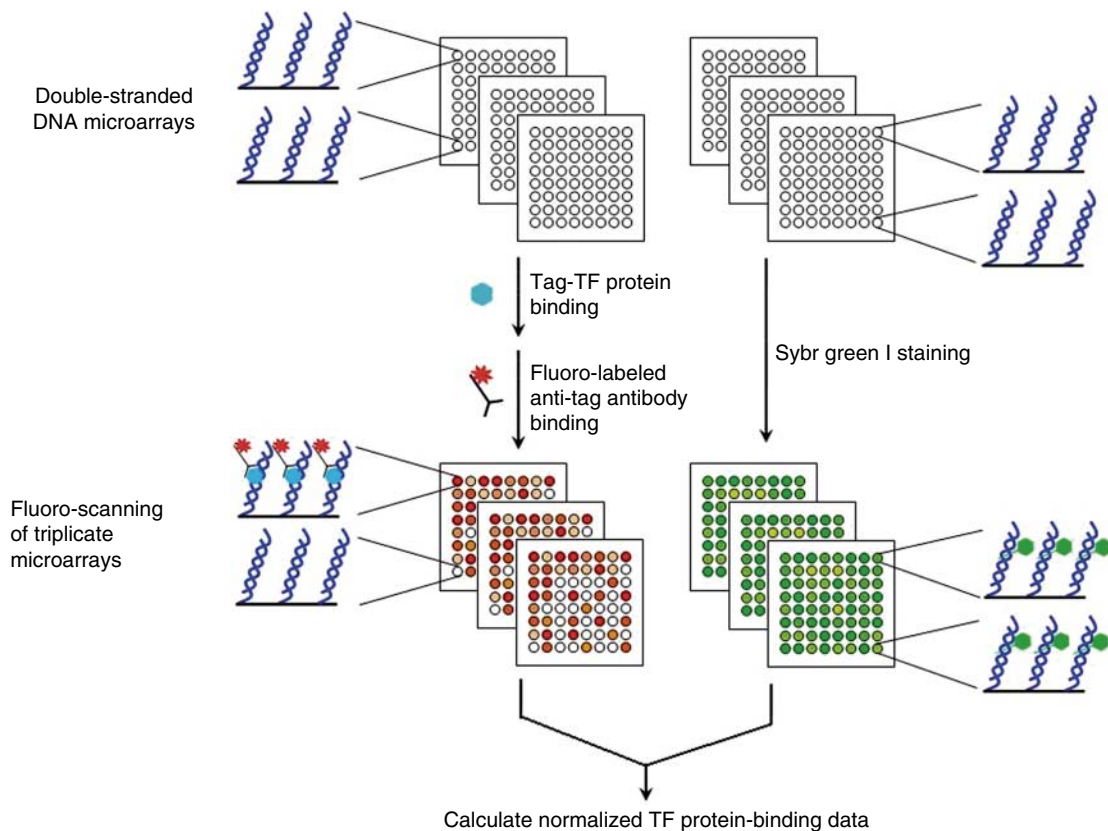
At present, the *in vitro* DBP is mainly generated by using two high-throughput methods, double-stranded DNA (dsDNA) microarray and systematic evolution of ligands by exponential enrichment in combination with parallel DNA sequencing techniques (SELEX-seq).

### dsDNA microarray

The double stranded DNA (dsDNA) microarray was first reported by Bulyk *et al.* (1999), which was also named protein-binding microarray (Berger *et al.* 2008, Badis *et al.* 2009). In principle, dsDNA microarray contains tens of thousands of dsDNA molecules in a small area on a glass slide, which can be used to detect the binding of a TF protein to these dsDNA molecules in a high-throughput format (Berger & Bulyk 2009).

The whole process of the dsDNA microarray experiment is schematically described in Fig. 1. The prerequisite of dsDNA microarray studies is to prepare the high-density dsDNA microarrays. However, no commercial dsDNA microarray chips can be purchased at present. Therefore, the high-density dsDNA microarray is manufactured in three steps. The first step

is to design DNA probes, especially complex probes (Berger *et al.* 2006, Mintseris & Eisen 2006, Philippakis *et al.* 2008). The second step is to authorize one of the biotechnical corporations, such as Affymetrix, Nimblegen, and Agilent, to manufacture high-density, single-stranded DNA (ssDNA) microarray. The third step is to convert ssDNA microarray



**Figure 1** Schematic description of dsDNA microarray approach. The first step of studying *in vitro* DBPs of TFs is to prepare the high-density dsDNA microarray as described in the text. The second step is to react the TF protein of interest to dsDNA microarray. For example, a purified TF protein tagged with an epitope (such as glutathione *S*-transferase, GST) is allowed to bind directly to a dsDNA microarray (Berger *et al.* 2006). The third step is to report the binding interactions of TF protein with all dsDNA probes on microarray. For example, a fluorophore-conjugated antibody specific to the epitope (such as Alexa488-conjugated antibody to GST) is allowed to bind to dsDNA microarray followed by TF protein binding reaction (Berger *et al.* 2006). One-step binding of a fluorescently labeled TF protein to dsDNA microarray can also be adopted (Kim *et al.* 2009). The purified, epitope-tagged TF protein can also be replaced by a cell nuclear extract containing TF protein of interest (Egener *et al.* 2005). In correspondence, the antibody specific to the epitope has to be replaced with an antibody specific to TF protein, and a fluorophore-conjugated second antibody is used to report the bound TF protein. In this case, the detection process increases one step. Finally, the dsDNA microarray is scanned with a genechip scanner, such as GSI Lumonics ScanArray 5000. To confirm the reproducibility of dsDNA microarray detection, the experiments are performed in triplicate. To eliminate the influence of the density of dsDNA probes in each of the features in microarray on the signal of TF protein binding, the density of dsDNA probes of each microarray used to detect TF protein binding must be detected by using a second fluorescence signal, such as Sybrgreen I (a dsDNA-specific fluorescent dye) stained with dsDNA microarray (Mukherjee *et al.* 2004), Cyanine 3 (Cy3) linked to dUTP, which is incorporated in dsDNA in primer extension of ssDNA microarray (Egener *et al.* 2005), and Cy5 coupled to ddATP, which is used to fluorescently label DNAs on the microarray using terminal transferase (Berger *et al.* 2006). The signal of these detections of the density of dsDNA probes is subtracted from the signal of the TF protein binding as background. The averaged background-subtracted, normalized signal intensities for all spots (features) are used as binding affinity data to perform subsequent bioinformatics studies, such as finding specificity and relative binding affinity of TF protein to various DNA sequences, DNA-binding motif, additivity, and interdependence of nucleotides in a binding site, predicting DNA-binding sites and target genes of TF protein in genome. Full colour version of this figure available via <http://dx.doi.org/10.1530/JME-11-0010>.

into dsDNA microarray by using several approaches, such as constant primer elongation (Bulyk *et al.* 1999), hairpin primer elongation (Wang *et al.* 2003c), or hairpin formation (Warren *et al.* 2006).

The greatest advantage of dsDNA microarray is that the binding interaction of a certain TF with all possible sequence variants of a given length can be simultaneously detected in a single assay. In recent years, dsDNA microarrays containing all possible 8–10 bp DNA duplexes have already been used to study *in vitro* DBPs of TFs and low-molecular weight ligands. For example, dsDNA microarray with all possible 8 bp sequences was used to determine the binding preferences of the majority of mouse homeodomains (168) (Berger *et al.* 2008) and the sequence-recognition properties of an engineered small molecule, PA1 (polyamide engineered to target a specific DNA sequence) (Warren *et al.* 2006). DsDNA microarray with all possible 9 bp sequences was used to profile the DNA-binding spectrum of yeast TFs CBF-1 and CBF1/DREB1B and rice TF OsNAC6 (Kim *et al.* 2009). DsDNA microarray containing all possible 10 bp sequences was used to characterize the binding specificities of five TFs, from yeast (CBF1 and RAP1), worm (CEH-22), mouse (Zif268), and human (OCT1) (Philippakis *et al.* 2008). Similarly, dsDNA microarray with every possible 10 bp sequences was used to study sequence-recognition profile of TF AP2 (De Silva *et al.* 2008). These dsDNA microarray studies yielded a comprehensive binding profile across the entire sequence space of a binding site and collected high-content data of ‘specificity landscape,’ which simultaneously displays the affinity and specificity of a million-plus DNA sequences to DNA-binding molecules (Carlson *et al.* 2010).

At present, dsDNA microarray has already been applied to profile *in vitro* DBPs of many DNA-binding molecules. For example, the high-density dsDNA microarray has been successfully used to profile *in vitro* DBPs of TFs (Berger *et al.* 2006, Alleyne *et al.* 2009, Zhu *et al.* 2009), and the studies on *in vitro* DBPs of TFs and other DNA-binding molecules have made tremendous progress (Warren *et al.* 2006, Maerkl & Quake 2007, Puckett *et al.* 2007, Keles *et al.* 2008, Bonham *et al.* 2009, Hauschild *et al.* 2009, Kim *et al.* 2009, Bolotin *et al.* 2010, Carlson *et al.* 2010). In 2009, the fine *in vitro* DBPs of 104 mouse TFs were successfully profiled by using dsDNA microarray technology (Badis *et al.* 2009). So far, *in vitro* DBPs of 406 TFs have already been profiled by using dsDNA microarray technology and stored in the UniPROBE database (Newburger & Bulyk 2009, Robasky & Bulyk 2011).

Many important parameters of DNA/TF interaction can be extracted from *in vitro* DBPs, among which the most important is DNA-binding specificity of TFs. DNA-binding specificity of TFs has broad impact on cell physiology, cell development and in evolution (Stormo 2000, Bulyk 2003). However, DNA binding specificity of most known TFs still remains unknown. For example, the DNA-binding specificity of only a small fraction of ~1400 human TFs is known. With the advent of dsDNA microarray technology, characterization of DNA-binding specificity of TFs rapidly

progressed (Bulyk *et al.* 2001, Berger & Bulyk 2006, Warren *et al.* 2006). For example, the DNA-binding specificity of 104 known and predicted mouse TFs from 22 different DNA-binding domain (DBD) structural classes found in metazoan TFs were determined by using the universal dsDNA microarray technology (Badis *et al.* 2009). For the vast majority of these TFs, this is the first time it was possible to obtain their high-resolution binding specificity data. A merit of characterizing DNA-binding specificity of TFs with dsDNA microarray is that the binding specificity of all TFs, regardless of structural class or species of origin, can be effectively solved by this method, even if no initial information about the binding site is available. In addition, by giving complete information about all possible sequences, dsDNA microarray can provide a true picture of the sequence specificity.

Another important information that can be derived from dsDNA microarray experiments is the relative binding affinity of a TF to all possible sequences of a given length (Philippakis *et al.* 2008, Berger & Bulyk 2009). Determination of the relative binding affinity of different TFs to their various DNA-binding sites is fundamentally important for a comprehensive understanding of gene regulation. The whole profile of DNA-binding affinity of a TF protein to all possible DNA sequence is useful for identifying all functional TFBSs of a TF, especially those TFBSs with low binding affinity. The *in vivo* studies revealed that both high- and low-affinity TFBSs had biological function, and the biologically important TFBSs are often not of maximal affinity (Jiang & Levine 1993, Tuupainen *et al.* 2009). A dsDNA microarray study demonstrated that the TFBSs in a wide affinity range were conserved and associated with regulatory function, and besides high-affinity TFBSs, numerous moderate- and low-affinity TFBSs were under negative selection in the mouse genome (Jaeger *et al.* 2010). The TFBSs with low and medium affinity are indispensable to the construction of the most accurate binding site models in bioinformatics (Roulet *et al.* 2002). The binding affinity data can also be used to evaluate the regulatory capability of a binding site to its target genes *in vivo*. In addition, a complete profile of DNA-binding affinity of a TF has promising biomedical applications. For example, DNA sequences with high affinity can be developed as drugs for transcription therapy, such as TF decoys (short duplex oligonucleotide containing DNA-binding site, which can be bound by a TF) (Mann & Dzau 2000, Tomita *et al.* 2007).

The dsDNA microarray also permits the discovery of subtle preferences of a TF to various DNA sequences, additivity between adjacent nucleotides or interdependencies among different positions in TFBSs, and functional polymorphism in TFBSs. A compact and universal dsDNA microarray can be used to rapidly determine the relative binding preferences of any TF from any organism (Berger *et al.* 2006). The complete reference tables of all possible binding sites on dsDNA microarray are important for comparing protein-binding preferences for various DNA sequences. For example, the universal dsDNA microarrays provided a complete reference

table of the relative binding preference of a TF for each gapped and ungapped 8 bp sequence variant (Badis *et al.* 2009). The dsDNA microarray can also be used to find the interdependence between nucleotides in TFBSs. For example, dsDNA microarray found 19 clear cases of 'position interdependence' TFs, which exhibited strong interdependence among the nucleotide positions of their binding sites (Badis *et al.* 2009). DsDNA microarray study revealed that position interdependence occurred on a broad scale and had important implications (Badis *et al.* 2009). It was also found that interdependent nucleotide positions were not always adjacent to each other (Badis *et al.* 2009). For example, Myb exhibited strong interdependence at positions separated by one nucleotide, with preference for binding either AACCGTCA or AACTGCCA (Badis *et al.* 2009). Position interdependencies frequently spanned more than just dinucleotides. For example, estrogen-related receptor  $\alpha$  had a strong preference for binding either CAAGGTCA or AGGGGTCA, but not CAGGGTCA or CGGGGTCA (Badis *et al.* 2009). DsDNA microarray study revealed that nucleotides of TFBSs exert interdependent effects on the binding affinities of TFs (Bulyk *et al.* 2002). The extensive existence of position interdependence in TFBSs suggests that it is important to consider the position interdependence in making accurate TFBS models because commonly used TFBS models assumed mononucleotide independence. The additivity between adjacent nucleotides in TFBSs was also found by dsDNA microarray (Benos *et al.* 2002, Bulyk *et al.* 2002); this suggests that the additive models are very useful for the identification of TFBSs in genomes (Benos *et al.* 2002). The *in vitro* DBPs can also be used together with computational models to identify the polymorphisms that affect TF binding and disease predisposition (Tuupanen *et al.* 2009).

The *in vitro* DBP data produced by dsDNA microarray can be used to identify TFBSs and target genes of TFs in genome. For example, the *in vitro* DBPs were used together with computational models to identify target genes of mammalian TFs, such as *Tcf4* (Hallikas & Taipale 2006, Hallikas *et al.* 2006). A dsDNA microarray was successfully used to identify the genome-wide binding sites and target genes of yeast TFs *Abf1*, *Rap1*, and *Mig1* (Mukherjee *et al.* 2004). Based on the *in vitro* DBPs obtained through dsDNA microarray and SELEX-seq described below, the most accurate binding site models, such as PWM, can be built. These models are very helpful for identifying binding sites and target genes of TFs in genomes. By searching the sequences corresponding to these models in genomes with TFBS identification search engines, binding sites and target genes of TFs in the whole genome can be identified. At present, numerous such TFBS identification search engines have been developed, such as position-specific scoring matrix (Stormo 2000), dictionary model (Sabatti *et al.* 2005), artificial neural network (Workman & Stormo 2000), hidden Markov model (Marinescu *et al.* 2005, Drawid *et al.* 2009), Bayesian network (Chen *et al.* 2010), and P-Match (Chekmenev *et al.* 2005).

Accumulation of TFBSs and target genes is indispensable for construction of transcription regulatory network. Although *in vivo* approaches, such as ChIP-chip (ChIP coupled with DNA microarray chip) (Ren *et al.* 2000) and ChIP-seq (ChIP coupled with parallel DNA sequencing) (Robertson *et al.* 2007), have generated *in vivo* DBPs for many TFs, however, because ChIP-based methods identify TFBSs in a particular cell at the time point of formaldehyde cross-linking, different cell types may need to be cultured in an indeterminate number of different conditions (such as stimulation) to determine all the biologically relevant DNA-binding sites of a given TF. In contrast, dsDNA microarray is an *in vitro* technology that does not depend on any certain cells and cultivating conditions; therefore, it can exhaustively identify all possible DNA targets that can be bound by a TF, and thus making comprehensive *in vitro* DBPs of TFs. These *in vitro* DBPs can be used to identify all potential binding sites of a particular TF in genome. In addition, dsDNA microarray has the capability of identifying DNA-binding sites of all TFs from any species, regardless of the level to which its genome has been characterized; however, ChIP-based methods can only be used to investigate TFs from species which genomes have already been characterized.

A limitation of *in vitro* techniques like dsDNA microarray in identification of TFBSs is that they cannot determine if or when the identified binding sites are utilized *in vivo*. Therefore, determination of *in vivo* relevance of *in vitro* identified TFBSs remains a great challenge at present. However, some pioneering investigations have already been performed in this field. For example, an *in vitro* study of DNA-binding specificity of yeast TFs *Abf1*, *Rap1*, and *Mig1* revealed that in addition to previously identified targets, *Abf1*, *Rap1*, and *Mig1* bound to 107, 90, and 75 putative new target intergenic regions, respectively, and many of them were upstream of previously uncharacterized open reading frames (Mukherjee *et al.* 2004). Comparative sequence analysis indicated that many of these newly identified sites are highly conserved across five sequenced *sensu stricto* yeast species. Therefore, these newly identified sites should be functional *in vivo* binding sites that may be used in a condition-specific manner. This study reveals that dsDNA microarray can find a large number of binding sites that cannot be found by ChIP-based methods. The DNA-binding specificities of the ETS (E-26) family determined with *in vitro* dsDNA microarray can be confirmed by *in vivo* ChIP-seq technology (Wei *et al.* 2010). It was also found that even relatively small differences in *in vitro* binding specificity of a TF contributed to site selectivity *in vivo* (Wei *et al.* 2010). Recently, dsDNA microarray was used in an integrated approach to identify target genes of human hepatocyte nuclear factor 4 $\alpha$  (Bolotin *et al.* 2010). Comparison of the dsDNA microarray data with ChIP-based data may provide insights into the usage of individual TF-binding sites *in vivo* (Mukherjee *et al.* 2004, Warren *et al.* 2006).



Our studies revealed that the predicted TFBSs and target genes with data from dsDNA microarray experiments provide a valuable blueprint to high-efficiency identification of functional DNA-binding sites and target genes of TFs. In recent years, our laboratory pursued studies of the *in vitro* DBPs based on the dsDNA microarray technique. We developed three methods for preparing unimolecular (hairpin) dsDNA microarray (Wang *et al.* 2003a,c, 2005) and used the prepared hairpin-dsDNA microarray to detect the binding of NF- $\kappa$ B to large numbers of DNA sequences (Wang *et al.* 2003b). We found that NF- $\kappa$ B bound to some mutated DNA sites with high affinity. We speculated that if these sites exist in human genome, they may be potential DNA-binding targets of NF- $\kappa$ B, and the genes neighboring these sites may be the target genes of NF- $\kappa$ B. Through a genome-wide search of the human genome, we found that these sites were indeed distributed in the human genome, we thus predicted these sites to be putative DNA targets of NF- $\kappa$ B and, correspondingly, predicted that the genes neighboring these sites were putative target genes of NF- $\kappa$ B. Through a literature search, we found that some predicted target genes have been identified as the functional NF- $\kappa$ B target genes by the previous experimental studies, such as *NFKB2*, *NFKB1A*, *BCL2*, and *VEGFC*. At the same time, we verified some selected typical disease-related genes, such as *STAT1*, *MIA-53*, *HFE-625*, and *LTBP-1*, which were not reported to be target genes of NF- $\kappa$ B, with ChIP-chip and gene expression profile.

It is necessary to point out that in identification of functional binding sites in the genome dependent on *in vitro* DBP data, the binding context of TFs, such as epigenetic modification of DNA, nucleosome position, chromosome structure, allosteric effects, and fluctuating cellular conditions, should be taken into consideration. A recent review pointed out that TFs' selection of specific DNA response element (REs) in the presence of degenerate sequences cannot be viewed only from the standpoint of DNA sequence variability and TF-binding affinity under steady-state conditions (Pan *et al.* 2009). It was proposed that the fluctuating cellular conditions should be a key factor in the TFs' selection of specific binding sites among the numerous similar binding sites present in the genome, because they lead to dynamic changes in the ensemble of protein (and DNA) conformational states via allosteric effects (Pan *et al.* 2009). This proposition is supported by the studies on regulatory diversity within the p53 transcriptional gene network selectivity in p53-dependent transcription (Espinosa 2008), which revealed that the p53-dependent transcriptional program is remarkably flexible, as it varies with the nature of p53-activating stimuli, the cell type, and the duration of the activation signal. These studies demonstrate that although the differential affinity of TF to various DNA-binding sites is a major factor in functional control, other factors are also important. Another recent review about the mechanisms of TF selectivity to TFBSs outlined that the recognition of selective binding site sequence and TF activation involve three major factors: the cellular network, protein and DNA as dynamic

conformational ensembles, and the tight packing of multiple TFs and coregulators on stretches of regulatory DNA (Pan *et al.* 2010). It was also revealed that the selective binding of p53 is achieved via a chromatin-dependent mechanism, but not through modulation of its binding affinity to certain REs (Millau *et al.* 2010). It was proposed that the formation of stress-specific p53 binding patterns is due to chromatin and chromatin remodeling, rather than the modulation of sequence-specific p53 binding affinity. It was revealed that several features, including but not limited to, the epigenetic landscape of the locus, p53 posttranslational modifications, the nature of the p53 RE, and p53-interacting partners, function in concert to determine the target promoter selectivity and the specificity of the p53 transcriptional response (Beckerman & Prives 2010).

It was demonstrated that DNA-induced allosteric effects on TFs play a critical role in TFs' selection of specific binding sites. It was revealed that the TFBS can act as allosteric effectors to determine the TFs' conformation. The selective gene transcription is not only mediated by TFs binding to TFBSs but TFs may also be modified in an allosteric manner by TFBSs themselves to generate the pattern of regulation that is appropriate to an individual gene (Lefstin & Yamamoto 1998). For example, the differential interaction of the DBD of estrogen receptor (ER) with the A2 and pS2 estrogen-responsive elements (EREs) brings about global changes in ER conformation (Wood *et al.* 1998). The conformational changes in ER induced by individual ERE sequences lead to the association of the receptor with different TFs and assist in the differential modulation of estrogen-responsive genes in target cells. The allosteric effects of DNA sites on the configuration of TF Pit-1 played essential role in control of differential expression of GH in different cells, somatotrope and lactotrope (Scully *et al.* 2000). It was also revealed that DNA-binding sites could allosterically modulate the transcriptional regulatory activity of glucocorticoid receptor (GR; Gronemeyer & Bourguet 2009). The transcriptional regulatory activity of the GR does not correlate with the affinity with which it binds to different GR-binding sites (GBSs), but rather with the sequence of the GBS, because the conformation of the GR's domain relevant to transcriptional regulatory activity was determined by the sequence of the GBS to which GR was bound (Meijsing *et al.* 2009). GR-binding sequences, differing by as little as a single bp, differentially affect GR conformation and regulatory activity. Therefore, it was proposed that DNA is a sequence-specific allosteric ligand of GR that tailors the activity of the receptor toward specific target genes. The ability of specific DNA sequences to allosterically regulate the transcriptional regulatory activity of GR provides a mechanism to achieve gene-specific regulatory activity, by which GR finely tunes its target gene network (Meijsing *et al.* 2009).

In addition to its great values in basic biological research described above, the *in vitro* DBPs of TFs also have promising biomedical applications. For example, *in vitro* DBPs of TFs can be used to guide the design and selection of artificial TFs

(Gommans *et al.* 2005, 2007, Klug 2005), small molecules of TF mimics (Kwon *et al.* 2004, Xiao *et al.* 2007, Block *et al.* 2009, Rodriguez-Martinez *et al.* 2010, Kushal *et al.* 2011), and TF decoys (Penolazzi *et al.* 2007, Tomita *et al.* 2007), which can be developed as drugs for transcription therapy. The *in vitro* DBPs can also accelerate the creation of precision-tailored DNA therapeutics (Carlson *et al.* 2010).

### SELEX-seq

SELEX, also referred to as *in vitro* selection or *in vitro* evolution, is an evolutionary process that allows the extraction, from an initially random pool of aptamers, of those molecules capable of binding to the target of interest (Stoltenburg *et al.* 2007). It was originally developed to screen oligonucleotides of either ssDNA or RNA that can specifically bind to DNA or RNA-binding proteins (Oliphant *et al.* 1989, Ellington & Szostak 1990, Tuerk & Gold 1990). At present, SELEX is used to develop high-affinity nucleic acid aptamers not only for a wide variety of pure molecules (such as protein) (Park *et al.* 2009) but also for complex systems such as live cells (Cell-SELEX) (Paul *et al.* 2009, Avci-Adali *et al.* 2010, Sefah *et al.* 2010). The screened DNA or RNA aptamers can be applied to basic research and disease diagnosis and treatment (Djordjevic 2007, Marton *et al.* 2010).

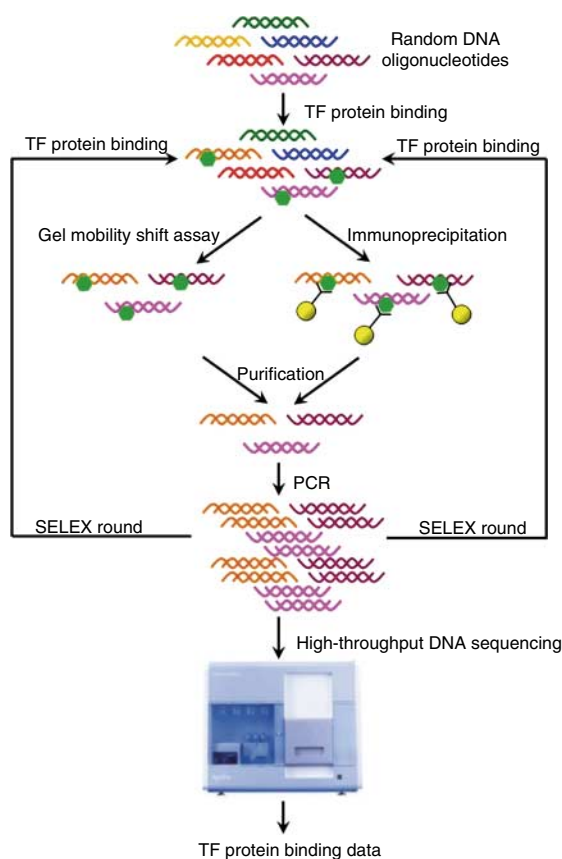
The procedures of the SELEX experiment include *in vitro* chemical synthesis of a single chain oligonucleotide library, mixing the oligonucleotide library with the target molecules such as RNA-binding protein to form complexes of a target molecule and oligonucleotide, isolation of bound oligonucleotides, and PCR amplification of enriched oligonucleotides to prepare a new library for the next round of the selection process. Through several rounds of repeated screening, the aptamers with high affinity and specificity can be obtained. As for the studies of TFs, the most critical step of SELEX is the isolation of DNA-protein complexes from free DNA. The methods used in this step include gel retardation assay (Tsai & Reed 1998, Tantin *et al.* 2008), affinity chromatography (Liu & Stormo 2005), filter-binding assay (Alex *et al.* 1992, Ferraris *et al.* 2010), and other approaches (Xue 2005, Gopinath 2007, Kim *et al.* 2010).

In previous studies, SELEX was frequently used for the purpose of characterizing the binding specificity of TFs. In such an experiment, SELEX yielded a library of dsDNA molecules binding to TF proteins, which was then used to generate a computational model, e.g. a position-specific scoring or weight matrix that served to predict binding sites of TFs in regulatory DNA sequences. SELEX has already been used to determine the binding specificity of many TFs, such as Sox2 (Maruyama *et al.* 2005), Oct4 (Tantin *et al.* 2008), Nanog (Mitsui *et al.* 2003), c-Myc (Papoulas *et al.* 1992), AP2, bHLH, NAC, MYB (Xue 2005), NF- $\kappa$ B (Kunsch *et al.* 1992), and GSKF (Shields & Yang 1998). However, these studies used low-throughput cloning DNA sequencing

technology; therefore, only limited numbers of DNA molecules (rarely exceeding 100 sequences) were sequenced. These limited DNA sequences produce low-resolution *in vitro* DBPs of TFs. To increase the sequencing throughput of the traditional SELEX method, the concatemerization step of serial analysis of gene expression (SAGE) was incorporated with SELEX (Roulet *et al.* 2002). In SELEX-SAGE, the SELEX-screened dsDNA fragments were first digested with a restriction endoenzyme BglII, then the digested dsDNA fragments with stick ends were concatemerized and cloned, finally the cloned DNA was sequenced by cloning DNA sequencing technology. SELEX-SAGE can generate large numbers (> 1000) of ligands in a single assay; therefore, this method was called high-throughput SELEX (HTPSELEX), which can originate large volumes of data (Jagannathan *et al.* 2006). However, SELEX-SAGE was still a method dependent on cloning DNA sequencing technology. Its DNA sequencing throughput is not high enough for generating comprehensive *in vitro* DBPs of TFs.

In the last two years, a new high-throughput technique named SELEX-seq was developed, which combined the conventional SELEX technique with massively parallel DNA sequencing techniques, such as Illumina SOLEXA. Zykovich *et al.* (2009) first reported SELEX-seq technique and used it to investigate DNA-binding motif and relative DNA-binding affinity of the TFs Zif268 and Aart; they named the technique bind-n-seq. Jolma *et al.* (2010) developed an improved SELEX-seq technique and applied it to building *in vitro* DBPs of 19 TFs. They validated the method by determining binding specificities of TFs belonging to 14 different classes and confirming the specificities for NFATC1 and RFX3 by using ChIP-seq (Jolma *et al.* 2010). These successful proof-in-principle studies demonstrated the great value of this technique in profiling *in vitro* DNA-binding spectrums of TFs. This technique is powerful but cost-effective: it can generate over 5 million sequences in a single assay with a cost as low as 1500 dollars. This technique thus provides a true high-throughput method for building comprehensive *in vitro* DBPs of TFs or other DNA-binding molecules. The whole process of this technique is schematically described in Fig. 2.

Most DNA/TF-binding information produced by dsDNA microarray can also be obtained by SELEX-seq. For example, SELEX-seq can be used to characterize DNA-binding specificity of TFs in ultrahigh resolution and determine the relative binding affinities of TFs to millions of DNA sequences. A recent SELEX-seq study revealed that the enrichment of a sequence in SELEX is proportional to the relative affinity of a TF protein to it (Zykovich *et al.* 2009). SELEX-seq data can also be used to determine DNA-binding motifs of TFs (Zykovich *et al.* 2009). For example, the binding motifs of two well-characterized zinc-finger proteins (Zif268 and Aart) were found with SELEX-seq data using the motif-finding program MEME, and the found motifs were similar to those previously derived from the cyclic amplification and selection of targets (Zykovich *et al.* 2009). The SELEX-seq-generated binding profile of mouse TF



**Figure 2** Schematic description of the SELEX-seq approach. The basic procedures of SELEX-seq include three steps. The first step is to design and synthesize a ssDNA library and then convert ssDNA into dsDNA using primer extension. The ssDNA library contained all possible oligonucleotides in random sequence of length  $k$ . The length of  $k$  can be determined according to the nucleotide numbers of DNA-binding consensus bound by a TF protein of interest. The second step is to bind TF proteins to randomized dsDNA library and isolate the protein-bound dsDNAs from binding reaction using gel mobility shift assay (also called electrophoresis mobility shift assay, EMSA) (Zykovich *et al.* 2009), affinity chromatography (Zykovich *et al.* 2009), or TF protein-coupled microwell plate (Jolma *et al.* 2010). The isolated dsDNAs are amplified by PCR to prepare a new library for the next round of the selection process. Through several rounds of repeated screening, the TF protein-bound dsDNAs were enriched. To find all sequences that can be bound by a TF protein with various binding affinities, especially sequences with low affinity, the isolated dsDNAs from each round of selection can be separately collected for sequencing (Zykovich *et al.* 2009). The third step of SELEX-seq is to sequence the bound dsDNAs with massively parallel DNA sequencing techniques, such as Illumina SOLEXA. The sequencing reads are filtered with filters including only A, C, G, and T letters allowed, valid bar code, and constant regions and unique random regions. If multiple DNA samples are simultaneously sequenced, the filtered reads are sorted according to bar code sequence. The qualified reads data are then used to perform subsequent bioinformatics analysis, including finding motifs or position weight matrix (PWM) models with some typical algorithms for this purpose, such as multiple EM for motif elicitation (MEME) (Bailey & Elkan 1994). A more detailed experimental and computational procedure to infer parameters of TF-DNA interaction from SELEX experiments was described by some studies (Djordjevic & Sengupta 2006). Full colour version of this figure available via <http://dx.doi.org/10.1530/JME-11-0010>.

eomesodermin (EOMES) was very similar to the dsDNA microarray-derived profile (Jolma *et al.* 2010). The SELEX-seq-generated binding profiles of 18 TFs were generally in good agreement with the existing data; however, some notable differences were seen, including TF POU2F2 and RFX3 (Jolma *et al.* 2010). It was also validated that the binding profiles generated using SELEX-seq method were relevant for the *in vivo* situation. For example, the enriched sequence motifs of TF RFX3 and NFATC1 in K562 and Jurkat cells from the ChIP-seq peaks using the MEME algorithm revealed a profile that was very similar to that generated using SELEX-seq method (Jolma *et al.* 2010). The broad utility of the SELEX-seq method was highlighted by the profiles for 14 TFs, which belong to 23 major DBD families occupying most major branches of TF-binding specificities (Jolma *et al.* 2010).

SELEX-seq has several significant advantages over dsDNA microarray (Table 1). SELEX-seq combines SELEX with massive parallel DNA sequencing technique, thus possessing advantages of both techniques. SELEX-seq needs no complex design and on-chip synthesis of oligonucleotides and provides an easy cost-effective alternative approach to research *in vitro* DBPs of TFs beside dsDNA microarray. Along with rapid development of DNA sequencing techniques, most advanced DNA sequencing techniques become affordable and reachable to general researchers, including equipment and commercialized DNA sequencing services. Moreover, SELEX-seq has ultrahigh throughput over dsDNA microarray by using bar code technique (Fig. 3). For example, up to 28 samples were simultaneously analyzed in one sequencing by using 3 nt bar-coded oligonucleotides (Zykovich *et al.* 2009). Due to its significant advantages over dsDNA microarray technology, SELEX-seq may replace dsDNA microarray technology as the master technique in future *in vitro* DBP studies.

It was noted that in current dsDNA microarray and SELEX-seq studies, many experiments targeted to mammalian TFs employed the purified recombinant TF proteins expressed in bacteria, such as *Escherichia coli* (Mukherjee *et al.* 2004, Berger *et al.* 2006, Kim *et al.* 2009, Zykovich *et al.* 2009). These recombinant proteins prepared from prokaryote expression may exclude possible post-translational modifications of mammalian TF proteins and omit the potential effects of these modifications on the DNA-binding activity of TFs. However, like many other proteins, TFs are post-translationally modified under different conditions and by different modifiers, such as phosphorylation, hydroxylation, acetylation, ubiquitination, and sumoylation (Grove & Walhout 2008). The post-translational modifications can affect the regulatory activity of a TF, as well as its localization or stability. For instance, the post-translational modifications play essential roles in regulating the activity of EST TF superfamily (Tootle & Rebay 2005). The phosphorylation was reported to inhibit the DNA binding of Ets1, Er81, and Erm of ETS TF superfamily, but enhance the DNA binding of Sap1, Elk-1, and Elf-1. The acetylation of ER $\alpha$  at

**Table 1** Advantages and limitations of dsDNA microarray and SELEX-seq techniques

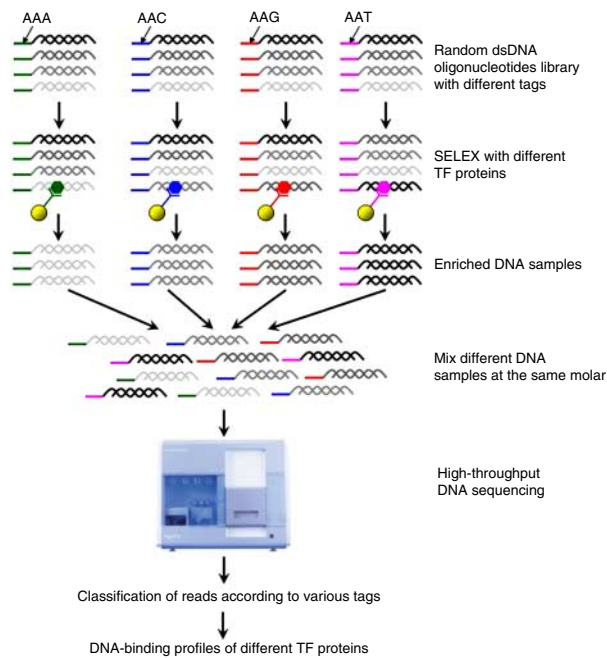
	dsDNA microarray	SELEX-seq
Throughput	Low throughput: one chip, one protein. Chip cannot be repeatedly used to detect several proteins	High-throughput: the use of bar coding techniques in MPS (Fig. 3) allows multiple protein samples to be analyzed simultaneously
Capacity	Limited by the number of features that can be placed on the array. The $10^6$ feature (all possible 10 bp sequences) is the limit for array technology today, but many TFs have binding sites longer than 10 bp. About $10^9$ features (all possible 15 bp sequences) are far beyond the current capacity of DNA microarray technique	Not limited to 10 bp binding sites. Proteins with long DNA-binding sites can be investigated. For example, the <i>in vitro</i> interactions of two TFs (Zif268 and Aart) to all possible 21 bp sequences were detected with this technique. In traditional SELEX, the length of random sequences can be up to 35 bp
Technique	Techniques for manufacturing high-density DNA microarray are unreachable for many biotechnical corporations and general researchers. The custom-ordered DNA microarray can only be provided by a few corporations, such as Affymetrix, Nimblegen, and Agilent	Techniques for parallel DNA sequencing are reachable for many biotechnical corporations and some researchers. The general researchers can get cost-effective and commercial service of DNA sequencing from many biotechnical corporations
Equipment	No commercialized machines. Complex machines for manufacturing high-density DNA microarray are possessed by a few corporations, such as Affymetrix, Nimblegen, and Agilent. In addition, detection process with dsDNA microarray need costly genechip scanners	Machines for parallel DNA sequencing can be purchased and owned by many biotechnical corporations and some researchers. The SELEX processes do not need any costly and complex machines
Probe design	Complex and tedious design of millions of probes synthesized on-chip. With high cost of probe design	Easy design and synthesis of random sequence DNA library. No cost of probe design
Cost	Cost of manufacturing high-density DNA microarray is still prohibitive to general researchers. To customize a DNA microarray with 1 million features can cost over 6000 dollars in China	Cost of parallel DNA sequencing is affordable to general researchers. This technique is cost effective, which can generate over 5 millions of sequences in a single assay with the cost as low as 1500 dollars
Determination of relative affinity and specificity	The relative binding affinity is directly reported by fluorescence signal. The affinity to all sequences, including low-affinity sequences, can be found. No PCR is needed in detecting process. The high-density dsDNA microarray may be helpful to find the fine specificity difference between sequences with single nucleotide differences	The relative binding affinity is calculated with an abundance of sequences in total sequenced reads. Some low-affinity sequences may be omitted. PCR is used in SELEX and sequencing which may introduce bias to some sequences or artifacts. SELEX-seq may have difficulty finding the fine specificity difference between sequences with single nucleotide differences

conserved lysine residues resulted in enhanced DNA-binding activity (Kim *et al.* 2006). The post-translational modification of the DNA-binding subunits by phosphorylation, acetylation, and ubiquitination regulated the activity of NF- $\kappa$ B (Mattioli *et al.* 2006, Geng *et al.* 2009, Moreno *et al.* 2010). To overcome the limitations of these TF proteins expressed in prokaryotes, the newest *in vitro* DBP study with SELEX-seq used purified TF proteins expressed in eukaryotes, such as mammalian cells (Jolma *et al.* 2010). The TF proteins expressed in mammalian cells can thus be used to characterize DNA-binding preferences of proteins requiring post-translational modifications. Another important problem regarding TF protein is that some experiments were performed only with the purified recombinant DBDs, not full length, of mammalian TF proteins expressed in bacteria (Zykovich *et al.* 2009) or mammalian cells (Jolma *et al.* 2010). These experiments with DBDs may also have serious limitations, because some regions of a full-length TF protein contribute to the dimerization of members of some TF family or superfamily, and this kind of dimerization to form homodimers or heterodimers is critical to the DNA-binding activity of TFs. For example, the TF E2F often binds its DNA-binding sites with low affinity, however, when dimerized with DP, its DNA-binding activity is greatly

enhanced (Tao *et al.* 1997). Therefore, the full-length TF proteins expressed in mammalian cells should be used in future *in vitro* DBP studies. In addition, it is important to use TF protein samples combining two different members of a TF family or superfamily in future studies, in order to find a more complete and accurate DNA-binding spectrum of TFs. However, in most current *in vitro* DBP studies with dsDNA microarray and SELEX-seq, only a single TF protein was used. The binding affinity experiments should be carried out in the presence of other protein factors.

It is worthy to note that besides methods described above, many other high-throughput methods were also developed for *in vitro* quantifying DNA-binding specificities of TFs in recent years. These methods include oligonucleotide mass tags and mass spectroscopy (Zhang *et al.* 2007), DIP-chip (Liu *et al.* 2005), microarray evaluation of genomic aptamers by shift (MEGAsift; Tantin *et al.* 2008, Ferraris *et al.* 2010), mechanically induced trapping of molecular interactions (MITOMI; Fordyce *et al.* 2010), microwell-based competition assay (Wei *et al.* 2010), transcriptional regulatory sequences (TRSs) interrogating with a flow cytometry and deep sequencing (TRS-FY-DS; Kinney *et al.* 2010), synthetic saturation mutagenesis (Patwardhan *et al.* 2009), and bacterial one-hybrid selections (Meng *et al.* 2005,





**Figure 3** Schematic description of parallel SELEX-seq analysis of DNA-binding of multiple TFs. The bar code technique was employed in SELEX-seq when it was developed (Zykovich *et al.* 2009). In this case, a bar code sequence consisting of a few nucleotides is added to each SELEX-seq oligo substrate. The different bar-coded oligonucleotides are applied to different TF protein samples (Jolma *et al.* 2010), or different rounds of selection of a same TF protein (Zykovich *et al.* 2009). After SELEX selection of each protein or round, the enriched DNA samples are mixed in the same molar and sequenced as a single DNA sample. After sequencing and reads quantification, the qualified reads are sorted according to bar code sequence and then applied to independent bioinformatics analysis. The bar code technique can greatly improve the detection throughput and lower the experimental cost of SELEX-seq. For example, up to 28 samples were simultaneously analyzed in one sequencing by using 3 nt bar-coded oligonucleotides (Zykovich *et al.* 2009). Using 256 oligonucleotide libraries with different barcode sequences, the binding specificities of 256 different TFs can be analyzed in a single sequencing run (Jolma *et al.* 2010). Full colour version of this figure available via <http://dx.doi.org/10.1530/JME-11-0010>.

Noyes *et al.* 2008). Some of these methods have certain special functions and can provide some additional information on DNA–TF binding specificity, which cannot be produced by dsDNA microarray and SELEX-seq. For example, MITOMI is a microfluidics-based approach that can simultaneously discover both high- and low-affinity target sequences and measure their relative and absolute affinities (Maerkl & Quake 2007). The significant advantage of this method is its capability of detecting low-affinity transient binding events. This method was successfully used to measure the relative binding affinities to oligonucleotides covering all possible 8 bp DNA sequences and created comprehensive maps of sequence preferences of 28 TFs with a variety of DBDs of *Saccharomyces cerevisiae*. Furthermore, some of these maps were

proven difficult to be studied by other techniques (Fordyce *et al.* 2010). However, this method needs to independently prepare large numbers of different dsDNA, which were noncovalently spotted on substrate to fabricate DNA microarray. Moreover, this method needs complex devices of microfluidics. The microwell-based competition assay can be used to directly quantitatively measure the affinity of DNA–protein binding interactions and determine the sequence specificities of DNA-binding proteins (Hallikas & Taipale 2006). This method is suitable for high-throughput screening to identify proteins or small molecules that modulate protein–DNA binding interactions. However, this method requires prior knowledge of one high-affinity binding site for the protein of interest. The methods of TRS-FY-DS and synthetic saturation mutagenesis combined the traditional reporter construct technique that is used to detect the transcription activity of a DNA sequence in cells with new parallel DNA sequencing techniques (Patwardhan *et al.* 2009, Kinney *et al.* 2010). The advantage of these two methods is to detect the DNA-binding specificity of a TF to various DNA sequences at the level of transcription activation in true intracellular environment. Therefore, they can reveal the transcription activation functions of different DNA sequences under the complex interaction with TFs and their cofactors.

## Summary

Regulatory TFs are a class of sequence-specific DNA-binding proteins that play essential roles in the regulation of gene expression. As more and more TFs of many organisms are identified, identification of all their functional TFBSs and target genes in the whole genome and construction of transcription regulatory network controlled by them are increasingly of importance. Therefore, two kinds of research become more and more intense in the field of TF-related studies. One is the identification of *in vivo* TFBSs and target genes using ChIP-base techniques, such as ChIP-chip and ChIP-seq, and global gene expression profiling DNA microarray. The other is the characterization of *in vitro* DNA-binding spectrum of TFs via dsDNA microarray and SELEX-seq techniques. The latter attracts increasing attentions due to its comprehensiveness in characterizing DNA-binding specificity and quantifying relative or absolute DNA-binding affinity of TFs. The exhaustive data obtained through these *in vitro* studies play critical roles in decoding gene regulatory codes and deep understanding complex transcriptional regulatory networks and the mechanism through which TFs control the fine temporal and special expressions of genes. At present, many *in vitro* methods chaired by dsDNA microarray and SELEX-seq have been developed, and large amounts of DNA-binding data are rapidly accumulated along with practical applications of these methods. At the same time, the corresponding bioinformatics are also promptly developed. Thus, it can be proposed that developments and applications

of these experimental and computational approaches will greatly improve future studies on TFs, which have already become the promising leads of genomics, system biology, regulatory biology, and transcription therapy biomedicine.

### Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

### Funding

This study was funded by the National Natural Science Foundation of China (60871014) and Supporting Program of New Century Excellent Talents of Ministry of Education (NCET-08-0110).

### References

- Adryan B & Teichmann SA 2006 FlyTF: a systematic review of site-specific transcription factors in the fruit fly *Drosophila melanogaster*. *Bioinformatics* **22** 1532–1533. (doi:10.1093/bioinformatics/bdl143)
- Alex R, Sozeri O, Meyer S & Dildrop R 1992 Determination of the DNA-sequence recognized by the Bhlh-Zip domain of the N-Myc protein. *Nucleic Acids Research* **20** 2257–2263. (doi:10.1093/nar/20.9.2257)
- Alleyne TM, Pena-Castillo L, Badis G, Talukder S, Berger MF, Gehrke AR, Philippakis AA, Bulyk ML, Morris QD & Hughes TR 2009 Predicting the binding preference of transcription factors to individual DNA k-mers. *Bioinformatics* **25** 1012–1018. (doi:10.1093/bioinformatics/btn645)
- Avci-Adali M, Metzger M, Perle N, Ziemer G & Wendel HP 2010 Pitfalls of cell-systematic evolution of ligands by exponential enrichment (SELEX): existing dead cells during *in vitro* selection anticipate the enrichment of specific aptamers. *Oligonucleotides* **20** 317–323. (doi:10.1089/oli.2010.0253)
- Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, Jaeger SA, Chan ET, Metzler G, Vedenko A, Chen XY *et al.* 2009 Diversity and complexity in DNA recognition by transcription factors. *Science* **324** 1720–1723. (doi:10.1126/science.1162327)
- Bailey TL & Elkan C 1994 Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings/International Conference on Intelligent Systems for Molecular Biology* **2** 28–36.
- Beckerman R & Prives C 2010 Transcriptional regulation by p53. *Cold Spring Harbor Perspectives in Biology* **2** a000935. (doi:10.1101/cshperspect.a000935)
- Benos PV, Bulyk ML & Stormo GD 2002 Additivity in protein–DNA interactions: how good an approximation is it? *Nucleic Acids Research* **30** 4442–4451. (doi:10.1093/nar/gkf578)
- Berger MF & Bulyk ML 2006 Protein binding microarrays (PBMs) for rapid, high-throughput characterization of the sequence specificities of DNA binding proteins. *Methods in Molecular Biology* **338** 245–260. (doi:10.1385/1-59745-097-9:245)
- Berger MF & Bulyk ML 2009 Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. *Nature Protocols* **4** 393–411. (doi:10.1038/nprot.2008.195)
- Berger MF, Philippakis AA, Qureshi AM, He FS, Estep PW III & Bulyk ML 2006 Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnology* **24** 1429–1435. (doi:10.1038/nbt1246)
- Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET *et al.* 2008 Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133** 1266–1276. (doi:10.1016/j.cell.2008.05.024)
- Block KM, Wang H, Szabo LZ, Polaske NW, Henchey LK, Dubey R, Kushal S, Laszlo CF, Makhoul J, Song Z *et al.* 2009 Direct inhibition of hypoxia-inducible transcription factor complex with designed dimeric epidthiodiketopiperazine. *Journal of the American Chemical Society* **131** 18078–18088. (doi:10.1021/ja807601b)
- Bolotin E, Liao H, Ta TC, Yang C, Hwang-Verslues W, Evans JR, Jiang T & Sladek FM 2010 Integrated approach for the identification of human hepatocyte nuclear factor 4alpha target genes using protein binding microarrays. *Hepatology* **51** 642–653. (doi:10.1002/hep.23357)
- Bonham AJ, Neumann T, Tirrell M & Reich NO 2009 Tracking transcription factor complexes on DNA using total internal reflectance fluorescence protein binding microarrays. *Nucleic Acids Research* **37** e94. (doi:10.1093/nar/gkp424)
- Bulyk ML 2003 Computational prediction of transcription-factor binding site locations. *Genome Biology* **5** 201. (doi:10.1186/gb-2003-5-1-201)
- Bulyk ML, Gentale E, Lockhart DJ & Church GM 1999 Quantifying DNA–protein interactions by double-stranded DNA arrays. *Nature Biotechnology* **17** 573–577. (doi:10.1038/9878)
- Bulyk ML, Huang X, Choo Y & Church GM 2001 Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. *PNAS* **98** 7158–7163. (doi:10.1073/pnas.111163698)
- Bulyk ML, Johnson PL & Church GM 2002 Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. *Nucleic Acids Research* **30** 1255–1261. (doi:10.1093/nar/30.5.1255)
- Carlson CD, Warren CL, Hauschild KE, Ozers MS, Qadir N, Bhimsaria D, Lee Y, Cerrina F & Ansari AZ 2010 Specificity landscapes of DNA binding molecules elucidate biological function. *PNAS* **107** 4544–4549. (doi:10.1073/pnas.0914023107)
- Chekmenov DS, Haid C & Kel AE 2005 P-Match: transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Research* **33** W432–W437. (doi:10.1093/nar/gki441)
- Chen X, Hoffman MM, Bilmes JA, Hesselberth JR & Noble WS 2010 A dynamic Bayesian network for identifying protein-binding footprints from single molecule-based sequencing data. *Bioinformatics* **26** i334–i342. (doi:10.1093/bioinformatics/btq175)
- De Silva EK, Gehrke AR, Olszewski K, Leon I, Chahal JS, Bulyk ML & Llinas M 2008 Specific DNA-binding by apicomplexan AP2 transcription factors. *PNAS* **105** 8393–8398. (doi:10.1073/pnas.0801993105)
- Djordjevic M 2007 SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomolecular Engineering* **24** 179–189. (doi:10.1016/j.bioeng.2007.03.001)
- Djordjevic M & Sengupta AM 2006 Quantitative modeling and data analysis of SELEX experiments. *Physical Biology* **3** 13–28. (doi:10.1088/1478-3975/3/1/002)
- Drawid A, Gupta N, Nagaraj VH, Gelinas C & Sengupta AM 2009 OHMM: a hidden Markov model accurately predicting the occupancy of a transcription factor with a self-overlapping binding motif. *BMC Bioinformatics* **10** 208. (doi:10.1186/1471-2105-10-208)
- Egener T, Roulet E, Zehnder M, Bucher P & Mermod N 2005 Proof of concept for microarray-based detection of DNA-binding oncogenes in cell extracts. *Nucleic Acids Research* **33** e79. (doi:10.1093/nar/gni079)
- Ellington AD & Szostak JW 1990 *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346** 818–822. (doi:10.1038/346818a0)
- Espinosa JM 2008 Mechanisms of regulatory diversity within the p53 transcriptional network. *Oncogene* **27** 4013–4023. (doi:10.1038/onc.2008.37)
- Ferraris L, Stewart AP, Gemberling MP, Reid DC, Lapadula MJ, Thompson WA & Fairbrother WG 2010 High-throughput mapping of protein occupancy identifies functional elements without the restriction of a candidate factor approach. *Nucleic Acids Research* **39** e33. (doi:10.1093/nar/gkq1213)
- Fordyce PM, Gerber D, Tran D, Zheng J, Li H, DeRisi JL & Quake SR 2010 *De novo* identification and biophysical characterization of transcription-factor binding sites with microfluidic affinity analysis. *Nature Biotechnology* **28** 970–975. (doi:10.1038/nbt.1675)
- Frank DA 2009 Targeting transcription factors for cancer therapy. *IDrugs: the Investigational Drugs Journal* **12** 29–33.

- Geng H, Wittwer T, Dittrich-Breiholz O, Kracht M & Schmitz ML 2009 Phosphorylation of NF-kappaB p65 at Ser468 controls its COMMD1-dependent ubiquitination and target gene-specific proteasomal elimination. *EMBO Reports* **10** 381–386. (doi:10.1038/embor.2009.10)
- Gommans WM, Haisma HJ & Rots MG 2005 Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command. *Journal of Molecular Biology* **354** 507–519. (doi:10.1016/j.jmb.2005.06.082)
- Gommans WM, McLaughlin PM, Lindhout BI, Segal DJ, Wiegman DJ, Haisma HJ, van der Zaal BJ & Rots MG 2007 Engineering zinc finger protein transcription factors to downregulate the epithelial glycoprotein-2 promoter as a novel anti-cancer treatment. *Molecular Carcinogenesis* **46** 391–401. (doi:10.1002/mc.20289)
- Gopinath SC 2007 Methods developed for SELEX. *Analytical and Bioanalytical Chemistry* **387** 171–182. (doi:10.1007/s00216-006-0826-2)
- Gronemeyer H & Bourguet W 2009 Allosteric effects govern nuclear receptor action: DNA appears as a player. *Science Signaling* **2** pe34. (doi:10.1126/scisignal.273pe34)
- Grove CA & Walhout AJ 2008 Transcription factor functionality and transcription regulatory networks. *Molecular BioSystems* **4** 309–314. (doi:10.1039/b715909a)
- Hallikas O & Taipale J 2006 High-throughput assay for determining specificity and affinity of protein–DNA binding interactions. *Nature Protocols* **1** 215–222. (doi:10.1038/nprot.2006.33)
- Hallikas O, Palin K, Sinjushina N, Rautiainen R, Partanen J, Ukkonen E & Taipale J 2006 Genome-wide prediction of mammalian enhancers based on analysis of transcription-factor binding affinity. *Cell* **124** 47–59. (doi:10.1016/j.cell.2005.10.042)
- Hauschild KE, Stover JS, Boger DL & Ansari AZ 2009 CSI-FID: high throughput label-free detection of DNA binding molecules. *Bioorganic and Medicinal Chemistry Letters* **19** 3779–3782. (doi:10.1016/j.bmcl.2009.04.097)
- Ho SW, Jona G, Chen CT, Johnston M & Snyder M 2006 Linking DNA-binding proteins to their recognition sequences by using protein microarrays. *PNAS* **103** 9940–9945. (doi:10.1073/pnas.0509185103)
- Jaeger SA, Chan ET, Berger MF, Stottmann R, Hughes TR & Bulky ML 2010 Conservation and regulatory associations of a wide affinity range of mouse transcription factor binding sites. *Genomics* **95** 185–195. (doi:10.1016/j.ygeno.2010.01.002)
- Jagannathan V, Roulet E, Delorenzi M & Bucher P 2006 HTPSELEX – a database of high-throughput SELEX libraries for transcription factor binding sites. *Nucleic Acids Research* **34** D90–D94. (doi:10.1093/nar/gkj049)
- Jiang J & Levine M 1993 Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72** 741–752. (doi:10.1016/0092-8674(93)90402-C)
- Jolma A, Kivioja T, Toivonen J, Cheng L, Wei G, Enge M, Taipale M, Vaquerizas JM, Yan J, Sillanpaa MJ *et al.* 2010 Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Research* **20** 861–873. (doi:10.1101/gr.100552.109)
- Keles S, Warren CL, Carlson CD & Ansari AZ 2008 CSI-Tree: a regression tree approach for modeling binding properties of DNA-binding molecules based on cognate site identification (CSI) data. *Nucleic Acids Research* **36** 3171–3184. (doi:10.1093/nar/gkn057)
- Kim MY, Woo EM, Chong YT, Homenko DR & Kraus WL 2006 Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor. *Molecular Endocrinology* **20** 1479–1493. (doi:10.1210/me.2005-0531)
- Kim MJ, Lee TH, Pakh YM, Kim YH, Park HM, Choi YD, Nahm BH & Kim YK 2009 Quadruple 9-mer-based protein binding microarray with DsRed fusion protein. *BMC Molecular Biology* **10** 91. (doi:10.1186/1471-2199-10-91)
- Kim SG, Lee S, Seo PJ, Kim SK, Kim JK & Park CM 2010 Genome-scale screening and molecular characterization of membrane-bound transcription factors in Arabidopsis and rice. *Genomics* **95** 56–65. (doi:10.1016/j.ygeno.2009.09.003)
- Kinney JB, Murugan A, Callan CG Jr & Cox EC 2010 Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *PNAS* **107** 9158–9163. (doi:10.1073/pnas.1004290107)
- Klug A 2005 Towards therapeutic applications of engineered zinc finger proteins. *FEBS Letters* **579** 892–894. (doi:10.1016/j.febslet.2004.10.104)
- Kunsch C, Ruben SM & Rosen CA 1992 Selection of optimal kappa B/Rel DNA-binding motifs: interaction of both subunits of NF-kappa B with DNA is required for transcriptional activation. *Molecular and Cellular Biology* **12** 4412–4421.
- Kushal S, Wang H, Laszlo CF, Szabo LZ & Olenyuk BZ 2011 Inhibition of hypoxia-inducible transcription factor complex with designed epoxy-lythiodiketopiperazine. *Biopolymers* **95** 8–16. (doi:10.1002/bip.21550)
- Kwon Y, Arndt HD, Mao Q, Choi Y, Kawazoe Y, Dervan PB & Uesugi M 2004 Small molecule transcription factor mimic. *Journal of the American Chemical Society* **126** 15940–15941. (doi:10.1021/ja0445140)
- Ladunga I 2010 *Computational Biology of Transcription Factor Binding*. New York: Springer.
- Latchman DS 2008 *Eukaryotic Transcription Factors*. Amsterdam; Boston: Elsevier/Academic Press.
- Lee AP, Yang Y, Brenner S & Venkatesh B 2007 TFCONES: a database of vertebrate transcription factor-encoding genes and their associated conserved noncoding elements. *BMC Genomics* **8** 441. (doi:10.1186/1471-2164-8-441)
- Lefstin JA & Yamamoto KR 1998 Allosteric effects of DNA on transcriptional regulators. *Nature* **392** 885–888. (doi:10.1038/31860)
- Li F & Sethi G 2010 Targeting transcription factor NF-kappa B to overcome chemoresistance and radioresistance in cancer therapy. *Biochimica et Biophysica Acta* **1805** 167–180. (doi:10.1016/j.bbcan.2010.01.002)
- Liu J & Stormo GD 2005 Combining SELEX with quantitative assays to rapidly obtain accurate models of protein–DNA interactions. *Nucleic Acids Research* **33** e141. (doi:10.1093/nar/gni139)
- Liu X, Noll DM, Lieb JD & Clarke ND 2005 DIP-chip: rapid and accurate determination of DNA-binding specificity. *Genome Research* **15** 421–427. (doi:10.1101/gr.3256505)
- Maerkl SJ & Quake SR 2007 A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315** 233–237. (doi:10.1126/science.1131007)
- Mann MJ & Dzau VJ 2000 Therapeutic applications of transcription factor decoy oligonucleotides. *Journal of Clinical Investigation* **106** 1071–1075. (doi:10.1172/JCI11459)
- Marinescu VD, Kohane IS & Riva A 2005 MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes. *BMC Bioinformatics* **6** 79. (doi:10.1186/1471-2105-6-79)
- Marton S, Reyes-Darias JA, Sanchez-Luque FJ, Romero-Lopez C & Berzal-Herranz A 2010 *In vitro* and *ex vivo* selection procedures for identifying potentially therapeutic DNA and RNA molecules. *Molecules* **15** 4610–4638. (doi:10.3390/molecules15074610)
- Maruyama M, Ichisaka T, Nakagawa M & Yamanaka S 2005 Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *Journal of Biological Chemistry* **280** 24371–24379. (doi:10.1074/jbc.M501423200)
- Mattioli I, Geng H, Sebald A, Hodel M, Bucher C, Kracht M & Schmitz ML 2006 Inducible phosphorylation of NF-kappa B p65 at serine 468 by T cell costimulation is mediated by IKK epsilon. *Journal of Biological Chemistry* **281** 6175–6183. (doi:10.1074/jbc.M508045200)
- Meijssing SH, Pufall MA, So AY, Bates DL, Chen L & Yamamoto KR 2009 DNA binding site sequence directs glucocorticoid receptor structure and activity. *Science* **324** 407–410. (doi:10.1126/science.1164265)
- Meng X, Brodsky MH & Wolfe SA 2005 A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nature Biotechnology* **23** 988–994. (doi:10.1038/nbt1120)
- Messina DN, Glasscock J, Gish W & Lovett M 2004 An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Research* **14** 2041–2047. (doi:10.1101/gr.2584104)
- Millau JF, Bandle OJ, Perron J, Bastien N, Bouchard EF, Gaudreau L, Bell DA & Drouin R 2010 Formation of stress-specific p53 binding



- patterns is influenced by chromatin but not by modulation of p53 binding affinity to response elements. *Nucleic Acids Research* (In Press). (doi:10.1093/nar/gkq1209)
- Mintseris J & Eisen MB 2006 Design of a combinatorial DNA microarray for protein–DNA interaction studies. *BMC Bioinformatics* **7** 429. (doi:10.1186/1471-2105-7-429)
- Mitsui K, Tokuzawa Y, Itoh H, Segawa K, Murakami M, Takahashi K, Maruyama M, Maeda M & Yamanaka S 2003 The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113** 631–642. (doi:10.1016/S0092-8674(03)00393-3)
- Moreno R, Sobotzik JM, Schultz C & Schmitz ML 2010 Specification of the NF-kappaB transcriptional response by p65 phosphorylation and TNF-induced nuclear translocation of IKK epsilon. *Nucleic Acids Research* **38** 6029–6044. (doi:10.1093/nar/gkq439)
- Mukherjee S, Berger MF, Jona G, Wang XS, Muzzey D, Snyder M, Young RA & Bulyk ML 2004 Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nature Genetics* **36** 1331–1339. (doi:10.1038/ng1473)
- Newburger DE & Bulyk ML 2009 UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. *Nucleic Acids Research* **37** D77–D82. (doi:10.1093/nar/gkn660)
- Noyes MB, Meng X, Wakabayashi A, Sinha S, Brodsky MH & Wolfe SA 2008 A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one–hybrid system. *Nucleic Acids Research* **36** 2547–2560. (doi:10.1093/nar/gkn048)
- Oliphant AR, Brandl CJ & Struhl K 1989 Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. *Molecular and Cellular Biology* **9** 2944–2949.
- Pan YP, Tsai CJ, Ma BY & Nussinov R 2009 How do transcription factors select specific binding sites in the genome? *Nature Structural and Molecular Biology* **16** 1118–1120. (doi:10.1038/nsmb1109-1118)
- Pan Y, Tsai CJ, Ma B & Nussinov R 2010 Mechanisms of transcription factor selectivity. *Trends in Genetics* **26** 75–83. (doi:10.1016/j.tig.2009.12.003)
- Papoulas O, Williams NG & Kingston RE 1992 DNA binding activities of c-Myc purified from eukaryotic cells. *Journal of Biological Chemistry* **267** 10470–10480.
- Park SM, Ahn JY, Jo M, Lee DK, Lis JT, Craighead HG & Kim S 2009 Selection and elution of aptamers using nanoporous sol–gel arrays with integrated microheaters. *Lab on a Chip* **9** 1206–1212. (doi:10.1039/b814993c)
- Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D & Shendure J 2009 High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nature Biotechnology* **27** 1173–1175. (doi:10.1038/nbt.1589)
- Paul A, Avci-Adali M, Ziemer G & Wendel HP 2009 Streptavidin-coated magnetic beads for DNA strand separation implicate a multitude of problems during cell–SELEX. *Oligonucleotides* **19** 243–254. (doi:10.1089/oli.2009.0194)
- Penolazzi L, Zennaro M, Lambertini E, Tavanti E, Torreggiani E, Gambari R & Piva R 2007 Induction of estrogen receptor alpha expression with decoy oligonucleotide targeted to NFATc1 binding sites in osteoblasts. *Molecular Pharmacology* **71** 1457–1462. (doi:10.1124/mol.107.034561)
- Philippakis AA, Qureshi AM, Berger MF & Bulyk ML 2008 Design of compact, universal DNA microarrays for protein binding microarray experiments. *Journal of Computational Biology* **15** 655–665. (doi:10.1089/cmb.2007.0114)
- Puckett JW, Muzikar KA, Tietjen J, Warren CL, Ansari AZ & Dervan PB 2007 Quantitative microarray profiling of DNA-binding molecules. *Journal of the American Chemical Society* **129** 12310–12319. (doi:10.1021/ja0744899)
- Redell MS & Tweardy DJ 2005 Targeting transcription factors for cancer therapy. *Current Pharmaceutical Design* **11** 2873–2887. (doi:10.2174/1381612054546699)
- Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA & Walhout AJ 2005 A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biology* **6** R110. (doi:10.1186/gb-2005-6-13-r110)
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E *et al.* 2000 Genome-wide location and function of DNA binding proteins. *Science* **290** 2306–2309. (doi:10.1126/science.290.5500.2306)
- Robasky K & Bulyk ML 2011 UniPROBE, update 2011: expanded content and search tools in the online database of protein–binding microarray data on protein–DNA interactions. *Nucleic Acids Research* **39** D124–D128. (doi:10.1093/nar/gkq1992)
- Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A *et al.* 2007 Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4** 651–657. (doi:10.1038/nmeth1068)
- Rodriguez-Martinez JA, Peterson-Kaufman KJ & Ansari AZ 2010 Small-molecule regulators that mimic transcription factors. *Biochimica et Biophysica Acta* **1799** 768–774.
- Roulet E, Busso S, Camargo AA, Simpson AJ, Mermod N & Bucher P 2002 High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nature Biotechnology* **20** 831–835. (doi:10.1038/nbt718)
- Sabatti C, Rohlin L, Lange K & Liao JC 2005 Vocabulon: a dictionary model approach for reconstruction and localization of transcription factor binding sites. *Bioinformatics* **21** 922–931. (doi:10.1093/bioinformatics/bti083)
- Scully KM, Jacobson EM, Jepsen K, Lunyak V, Viadiu H, Carriere C, Rose DW, Hooshmand F, Aggarwal AK & Rosenfeld MG 2000 Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science* **290** 1127–1131. (doi:10.1126/science.290.5494.1127)
- Sefah K, Shanguan D, Xiong XL, O'Donoghue MB & Tan WH 2010 Development of DNA aptamers using Cell–SELEX. *Nature Protocols* **5** 1169–1185. (doi:10.1038/nprot.2010.66)
- Shields JM & Yang VW 1998 Identification of the DNA sequence that interacts with the gut-enriched Kruppel-like factor. *Nucleic Acids Research* **26** 796–802. (doi:10.1093/nar/26.3.796)
- Stoltenburg R, Reinemann C & Strehlitz B 2007 SELEX-A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24** 381–403. (doi:10.1016/j.bioeng.2007.06.001)
- Stormo GD 2000 DNA binding sites: representation and discovery. *Bioinformatics* **16** 16–23. (doi:10.1093/bioinformatics/16.1.16)
- Tantin D, Gemberling M, Callister C & Fairbrother W 2008 High-throughput biochemical analysis of *in vivo* location data reveals novel distinct classes of POU5F1(Oct4)/DNA complexes. *Genome Research* **18** 631–639. (doi:10.1101/gr.072942.107)
- Tao Y, Kassatly RF, Cress WD & Horowitz JM 1997 Subunit composition determines E2F DNA-binding site specificity. *Molecular and Cellular Biology* **17** 6994–7007.
- Tomita N, Kashihara N & Morishita R 2007 Transcription factor decoy oligonucleotide-based therapeutic strategy for renal disease. *Clinical and Experimental Nephrology* **11** 7–17. (doi:10.1007/s10157-007-0459-6)
- Tootle TL & Rebay I 2005 Post-translational modifications influence transcription factor activity: a view from the ETS superfamily. *Bioessays* **27** 285–298. (doi:10.1002/bies.20198)
- Tsai RY & Reed RR 1998 Identification of DNA recognition sequences and protein interaction domains of the multiple–Zn–finger protein Roaz. *Molecular and Cellular Biology* **18** 6447–6456.
- Tuerk C & Gold L 1990 Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249** 505–510. (doi:10.1126/science.2200121)
- Tuupainen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei GH, Yan J, Niittymaki I *et al.* 2009 The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nature Genetics* **41** U885–U837. (doi:10.1038/ng.406)
- Wang J, Bai Y, Li T & Lu Z 2003a DNA microarrays with unimolecular hairpin double-stranded DNA probes: fabrication and exploration of sequence-specific DNA/protein interactions. *Journal of Biochemical and Biophysical Methods* **55** 215–232. (doi:10.1016/S0165-022X(03)00048-4)



- Wang JK, Li TX, Bai YF & Lu ZH 2003b Evaluating the binding affinities of NF- $\kappa$ B p50 homodimer to the wild-type and single-nucleotide mutant Ig- $\kappa$ B sites by the unimolecular dsDNA microarray. *Analytical Biochemistry* **316** 192–201. (doi:10.1016/S0003-2697(03)00049-6)
- Wang JK, Li TX, Bai YF, Zhu Y & Lu ZH 2003c Fabrication of unimolecular double-stranded DNA microarrays on solid surfaces for probing DNA–protein/drug interactions. *Molecules* **8** 153–168. (doi:10.3390/80100153)
- Wang JK, Li TX & Lu ZH 2005 A method for fabricating uni-dsDNA microarray chip for analyzing DNA-binding proteins. *Journal of Biochemical and Biophysical Methods* **63** 100–110. (doi:10.1016/j.jbbm.2005.03.006)
- Warren CL, Kratochvil NC, Hauschild KE, Foister S, Brezinski ML, Dervan PB, Phillips GN Jr & Ansari AZ 2006 Defining the sequence-recognition profile of DNA-binding molecules. *PNAS* **103** 867–872. (doi:10.1073/pnas.0509843102)
- Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M, Bonke M, Jolma A, Varjosalo M, Gehrke AR *et al.* 2010 Genome-wide analysis of ETS-family DNA-binding *in vitro* and *in vivo*. *EMBO Journal* **29** 2147–2160. (doi:10.1038/emboj.2010.106)
- Wood JR, Greene GL & Nardulli AM 1998 Estrogen response elements function as allosteric modulators of estrogen receptor conformation. *Molecular and Cellular Biology* **18** 1927–1934.
- Workman CT & Stormo GD 2000 ANN-Spec: a method for discovering transcription factor binding sites with improved specificity. *Pacific Symposium on Biocomputing* 467–478.
- Xiao X, Yu P, Lim HS, Sikder D & Kodadek T 2007 A cell-permeable synthetic transcription factor mimic. *Angewandte Chemie (International ed. in English)* **46** 2865–2868. (doi:10.1002/anie.200604485)
- Xue GP 2005 A CELD-fusion method for rapid determination of the DNA-binding sequence specificity of novel plant DNA-binding proteins. *Plant Journal* **41** 638–649. (doi:10.1111/j.1365-313X.2004.02323.x)
- Zhang L, Kasif S & Cantor AC 2007 Quantifying DNA–protein binding specificities by using oligonucleotide mass tags and mass spectroscopy. *PNAS* **104** 3061–3066. (doi:10.1073/pnas.0611075104)
- Zhu C, Byers KJ, McCord RP, Shi Z, Berger MF, Newburger DE, Saulrieta K, Smith Z, Shah MV, Radhakrishnan M *et al.* 2009 High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Research* **19** 556–566. (doi:10.1101/gr.090233.108)
- Zykovich A, Korfi I & Segal DJ 2009 Bind-n-Seq: high-throughput analysis of *in vitro* protein–DNA interactions using massively parallel sequencing. *Nucleic Acids Research* **37** e151. (doi:10.1093/nar/gkp802)

Received in final form 28 February 2011

Accepted 9 March 2011

Made available online as an Accepted Preprint  
9 March 2011