Anal Bioanal Chem (2010) 398:2551–2561
DOI 10.1007/s00216-010-4096-7

REVIEW

Analysis of DNA–protein interactions: from nitrocellulose filter binding assays to microarray studies

Reham Helwa · Jörg D. Hoheisel

Received: 9 July 2010 / Accepted: 3 August 2010 / Published online: 21 August 2010

Abstract Determination of the sequence of the human genome and knowledge of the genetic code have allowed rapid progress in the identification of mammalian proteins. However, far less is known about the molecular mechanisms that control expression of human genes and about the variations in gene expression that underlie many pathological states, including cancer. This is caused in part by lack of information about the binding specificities of DNA-binding proteins and particularly regulative important molecules such as transcription factors. It is consequently crucial to develop new technologies or improve existing ones for the analysis of DNA–protein interaction in order to identify and characterise DNA response elements and the related transcription factors or other DNA-binding proteins. The techniques that are currently available vary with respect to the type of result that can be expected from the assay: a mere qualitative demonstration of binding; the identification of response element sequences at high throughput; or a quantitative characterisation of affinities. This article gives an overview of early and recent methodologies applied to such ends.

Published in the special issue *Focus on Bioanalysis* with guest editors Antje J. Baeumner, Günter Gauglitz, and Frieder W. Scheller.

R. Helwa (✉) · J. D. Hoheisel
Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ),
Im Neuenheimer Feld 580,
69120 Heidelberg, Germany
e-mail: r.helwa@dkfz.de

R. Helwa
Molecular Cell Biology Laboratory, Zoology Department,
Faculty of Science, Ain Shams University,
Cairo 11566, Egypt

Keywords DNA structure · DNA-binding proteins · Transcription factors · Regulation

Introduction

DNA–protein interactions mediate a wide range of cellular processes ranging from cell cycle, via cellular differentiation and developmental time courses, to the maintenance of an intracellular metabolic and physiological balance [1–3]. In eukaryotes, a large number of diverse proteins are crucially involved in the regulation of transcription, which is a primary level of cellular regulation. Apart from the RNA polymerases themselves, there are histones, chromatin remodelling proteins, general transcription factors, several co-factors, and a host of sequence-specific transcription factors that direct transcription initiation to specific promoters [4]. Numerous diseases arise as a result of a breakdown of this part of the regulatory system. Transcription factors, for example, are over-represented among oncogenes [5]. For this reason, transcription factors have been studied extensively for decades. At the level of DNA analysis, many sequence elements were revealed that may be involved in regulating gene expression. Comparison of genome sequences of related organisms has identified thousands of evolutionarily conserved sequence motifs [6–8]. Furthermore, comparison of the sequences adjacent to co-regulated sets of genes frequently revealed shared sequence motifs [9–11]. Despite these studies, however, the exact nature of DNA–protein interaction still remains uncharacterised for the majority of human transcription factors [12].

For many years, biochemical assays were used to characterise DNA–protein interactions. However, such approaches are generally laborious and slow. Recent decades have witnessed the development of technologies that permit

analyses on a larger scale and in a more unbiased manner. These approaches are often either gene-centred or protein-centred [13]. In gene-centred approaches, an individual protein is used to identify DNA target sequences. Inversely, an individual DNA sequence is used to identify and study the relevant DNA-binding proteins in a protein-centred assay. Recently, microarray and sequencing technology were employed toward genome-wide analyses in both formats. As a result of these developments, more and more regulatory elements and factors are uncovered. In consequence, there is a continuously growing understanding of many basic regulatory processes that involve DNA–protein interaction. In this review, we describe the different technologies used for the acquisition and elucidation of this information and discuss their merits and disadvantages. For a quick overview, the technical basis, advantages and disadvantages as well as appropriate literature are summarised in Table 1.

Nitrocellulose filter binding assay

This assay was developed in the early stages of molecular biology in the 1970s. The manipulations are rapid enough to allow kinetic studies as well as equilibrium measurements [14–16]. The process is based on the fact that proteins bind to nitrocellulose without losing their DNA-binding capacity, while double-stranded DNA alone is not retained. For analysis, DNA and proteins are mixed and incubated under appropriate conditions. The mixture is then separated by electrophoresis and subsequently blotted onto nitrocellulose. Since only proteins bind, DNA remains on the membrane only if in complex with a protein. The exact amount of DNA that is retained on the nitrocellulose can be quantified by measuring a label that is introduced to the DNA prior to incubation with the protein.

The amount of information obtained from this kind of assay is limited, however. Only the mere retention of labelled nucleic acid is detected and not the identity of the proteins involved, or the proportion of binding activity attributable to an individual protein if more than one protein in the mixture exhibits DNA-binding capacity [17]. Also, the actual DNA binding site cannot be localised with this assay unless defined DNA fragments are used in the analysis. Moreover, as a technical complication, single-stranded nucleic acids are retained at nitrocellulose filters under particular conditions, resulting in background that can obscure the measurement [18].

Footprinting assays

Footprinting assays exploit the fact that a protein, which is bound to a specific nucleic acid, will interfere with a chemical or enzymatic modification of that DNA fragment. Thus, the modification can be used to localise the contact area between protein and DNA [19].

For the DNase I footprinting assay [20, 21], a particular DNA fragment is labelled at one end and mixed with the protein of interest. Following binding, the DNA is treated with the enzyme deoxyribonuclease I (DNase I), which digests DNA that is not in close contact with a protein and thus not protected from digestion (Fig. 1). Performing a partial cleavage without protein produces labelled DNA fragments that—because of the random nature of cleavage—cover the entire size range of the original DNA. In the presence of a binding protein, however, protection occurs in a particular region and labelled DNA fragments of the respective length are not produced, while all longer or shorter ones are still present. Resolving the two samples on a polyacrylamide gel side by side, the differences in the resulting ladders of DNA bands is visualised via the incorporated label. Gaps in the band ladder of the sample, in which protein had been present, indicate binding sites. Comparison of the patterns with sequencing reactions allows the identification of protected sequences with single-nucleotide resolution.

Dimethyl sulphate protection footprinting is a chemical variant of the enzymatic DNase I footprinting and relies on the ability of dimethyl sulphate to methylate specifically guanine residues in DNA. The methylated G residues are cleaved by exposure to piperidine, whereas no cleavage occurs at unmethylated bases [22]. A protein bound to a DNA will protect the G residues from methylation and hence from cleavage by piperidine.

DNase I footprinting is still a standard method to identify the nucleic acid sequences within or near the binding site, although ChIP-on-chip (see below) and other approaches provide higher throughput. The footprint signal can be obtained under equilibrium conditions. On the other hand, incomplete or unspecific binding, which occurs relatively frequently, could result in an unclear footprint [21].

Methylation interference assay

This procedure is also based on the ability of dimethyl sulphate (DMS) to methylate G residues, which in turn are cleaved by piperidine [22], and the fact that methylation of purine residues in a DNA sequence inhibits formation of a DNA–protein complex [23, 24]. For the assay, an end-labelled DNA probe is partially methylated with DMS and incubated with a nuclear protein extract. The protein–DNA complexes formed during the incubation are then separated from the free DNA using electrophoretic mobility shift assay (EMSA). Both the protein-bound DNA and the free DNA are eluted from the gel, cleaved with piperidine and again resolved by denaturing polyacrylamide gel electro-

Table 1 Summary of the most commonly used techniques for analysis of DNA–protein interactions

Assay	Advantage(s)	Limitation(s)	References
Nitrocellulose filter binding assay	Relatively simple handling	No localisation of binding site; unspecific binding of single-stranded DNA obscures analysis	[14–18]
Footprinting assays	Technical simplicity; single-base resolution	Incomplete binding frequently results in unclear footprint	[19–22]
Methylation interference	Combined analysis of binding site and effect of epigenetic variations	Very complex workflow	[22–24]
Electrophoretic mobility shift assay (EMSA)	Technically simple assay that permits semi-quantitative studies	In complex analyses, no immediate information on binding sites or proteins involved	[25–29]
Chromatin immunoprecipitation (ChIP)	Applicable also for in vivo analyses; permits high throughput	Relies very strongly on antibody specificity	[30–34]
DNA adenine methyltransferase identification (DamID)	In vivo detection; no dependence on antibodies; no cross-linking thus lower risk of artefacts; higher sensitivity	Requirement of exogenous fusion proteins; not suitable for detecting rapid changes	[35, 36]
Surface plasmon resonance (SPR)	Real-time recording of association and dissociation; no label required; currently the gold standard	No high throughput	[37–39]
Systematic evolution of ligands by exponential enrichment (SELEX)	Enables in vitro selection of optimal binding partners	Only selection of best binding events	[40, 41]
Yeast one-hybrid system	In vivo assay	Very complex system	[42–44]
DNA microarrays	High throughput; detection of binding sites; semi-quantitative analysis	Analysis process for individual proteins	[45–55]
Protein microarrays	High throughput; detection of binding sites; semi-quantitative analysis	Monomer-specificity; very complex system	[56–60]
Proximity ligation	Highly specific and sensitive down to single-molecule detection	Complex sample preparation; dependence on good binders; analysis only of known combinations	[61]
Atomic force microscopy, X-ray crystallography, nuclear magnetic resonance	High-resolution structural information	No use for definition of interaction pairs or identification of genomic locations	[63–66]

phoresis. If methylation had occurred at a particular guanine residue that is critical for the DNA–protein interaction, the binding of the protein to that DNA was inhibited, resulting in the recovery of the DNA only from the free DNA fraction. The presence of particular DNA fragments in the free DNA fraction and their concomitant absence from the DNA–protein fraction indicates that those nucleotides are contact points of proteins.

While this approach also permits the study of the influence of methylation on protein binding, which is of particular relevance with respect to epigenetic regulation, the overall workflow is far too complex. In consequence, this method has rarely been applied and disappeared from recent literature.

Electrophoretic mobility shift assay (EMSA)

In comparison to the assays described above, EMSA is a relatively rapid and technically simple method to detect particular DNA–protein interactions [25, 26]. It relies on the fact that the electrophoretic mobility of a complex of nucleic acid and protein is less than that of the free nucleic acid (Fig. 2). Mobility shift assays are often used for qualitative purposes. However, under appropriate conditions, they can even provide quantitative data for the determination of binding stoichiometries, affinities and kinetics. The exact methodological details differ for each purpose as reviewed elsewhere [27, 28].

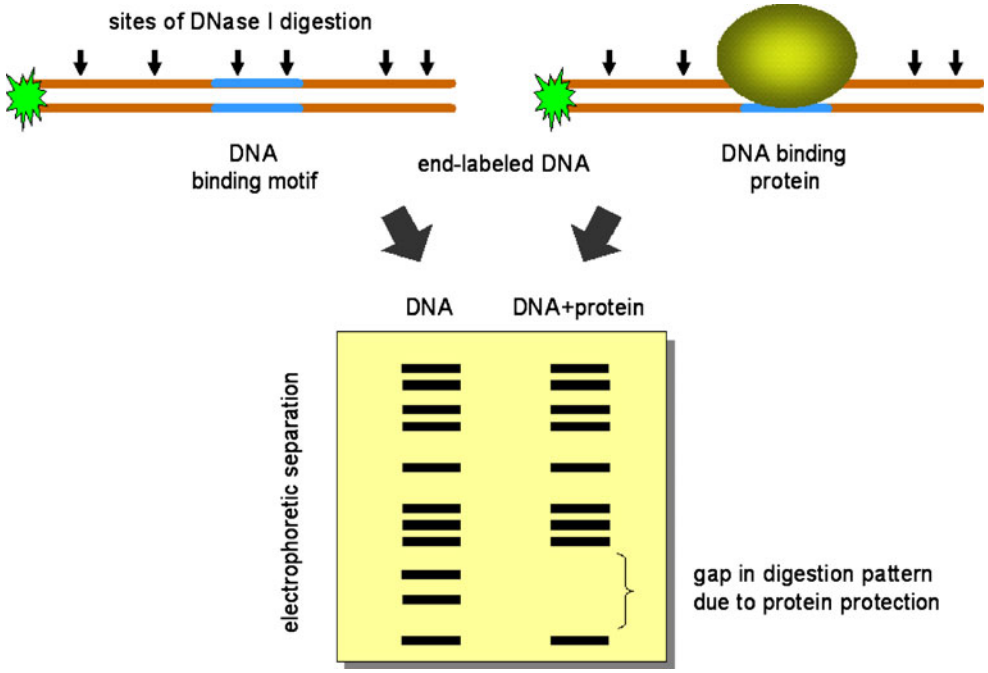

Fig. 1 Schematic representation of DNase I footprinting. Further details on the process are given in the text

Despite the fact that the technique is widely used to detect DNA–protein interactions, it has several limitations. Rapid dissociation during electrophoresis can prevent detection of complexes. Also, many complexes are significantly more stable in a gel than they are in free solutions [29]. In addition, EMSA is not informative about the sequences of the nucleic acids that are bound by the proteins. Another limitation is the fact that the observed mobility shift does not provide a direct measure of the molecular weight or identity of the proteins responsible. However, modifications such as an electrophoretic supershift assay and procedures that combine EMSA with Western blotting or mass spectrometry have been designed to identify the DNA-binding proteins [28].

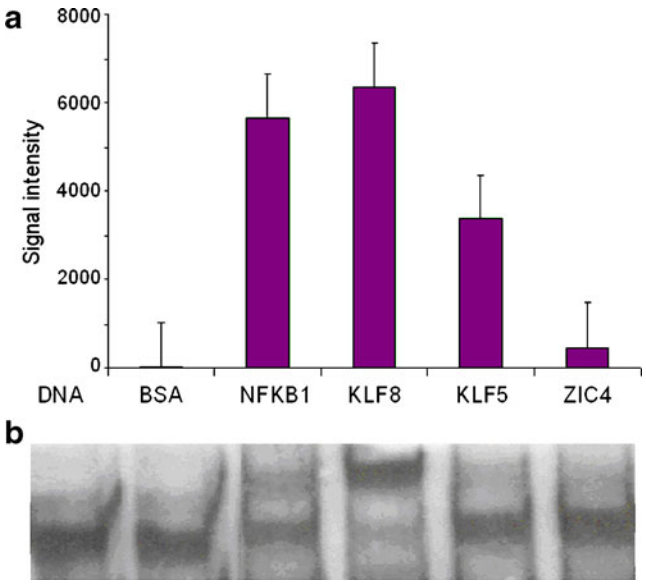

Fig. 2 Comparison of results from a protein array measurement and a related EMSA experiment (unpublished data from our own studies). Transcription factors were used to produce a protein microarray. **a** Signal intensities obtained upon an incubation with a 30-mer double-stranded oligonucleotide, whose sequence had been selected from the promoter region of *NFKB1*. Results at four transcription factors and a BSA negative control are shown, clearly illustrating the fact that particular sequences may be shared between transcription factors, especially if members of a family. **b** Corresponding results of an EMSA experiment. In the lane labelled with DNA, no protein had been added. The extent of the band shifts corresponds with the array data

Chromatin immunoprecipitation (ChIP)

ChIP is a method to identify DNA–protein complexes that occur in vivo and currently the standard method for the identification of histone modification locations and transcription factor binding sites [30]. Cells are initially treated with a cross-linking agent to link covalently any DNA-binding protein to the chromatin. Then, the cells are lysed and the genomic DNA is isolated and sonicated to produce sheared chromatin. An antibody specific to the protein of interest is added to the sonicated material and used to isolate the protein with all attached DNA via immunoprecipitation. The DNA is released by reversing the cross-linking and is subsequently purified. Classically, the DNA obtained from ChIP reactions was assessed by polymerase chain reaction (PCR), details of which are reviewed elsewhere [31]. In combination with microarrays that represent the genome (ChIP-on-chip), the assay became much more powerful, identifying the binding sequence and its location in the genome by hybridisation to a particular

microarray feature [32]. ChIP-seq is a more recent alternative, using second-generation DNA sequencing instead of microarray technology for sequence identification and thereby genomic localisation of the isolated DNA fragments [33].

The method is currently the state-of-the-art procedure for identification of interactions. ChIP has inherent features, however, that can make the identification of a DNA binding sites difficult. In particular, lacking specificity for the DNA-binding protein or the antibody used for precipitation can result in an experiment with insufficient enrichment [34]. In addition, the process does not allow for quantitative analysis of the affinity of the interaction partners.

DNA adenine methyltransferase identification (DamID)

Like ChIP, DamID is an assay to detect DNA–protein interactions in vivo. The target protein is expressed as a fusion molecule with DNA adenine methyltransferase (Dam) originally expressed in *Escherichia coli*. Dam methylates the adenine of a GATC site. Therefore, when the fusion protein binds to DNA, the Dam portion will methylate GATC sites that are located in the vicinity of the binding sites. The methylated sites in the target DNA and a control sample (expression of Dam alone) are detected by digestion with methyl-specific restriction enzymes, followed by amplification, labelling and hybridisation to a microarray as for a ChIP-on-chip assay [35]. DamID has been used to identify binding sites of proteins in different organisms and targeted the binding sites for sequence-specific transcripton factors, DNA methyltransferase and chromatin-associated proteins using PCR amplicon arrays or 60-mer oligonucleotide tiling arrays.

DamID has some advantages compared to ChIP. For example, DamID is not dependent on the availability of high-quality antibodies. Moreover, there is no need to use cross-linking reagents, eliminating the risk of cross-linking artefacts. Third, DamID can be performed on about 10^6 cells, which is 10- to 100-fold less material than typically used in ChIP experiments [36]. However, DamID has also limitations. It requires an exogenous fusion protein, whereas ChIP can be performed with the endogenous protein; some proteins lose their genomic binding specificity when fused to Dam. DamID is also less suitable for the detection of rapid changes in protein binding (e.g. during the cell cycle), because the methylation patterns obtained in a typical DamID experiment represent the average of a time period of about 24 h or more. Last, DamID cannot be used to map post-translational modifications such as histone modifications; for such applications, ChIP-on-chip is more suitable.

Surface plasmon resonance (SPR)

Surface plasmon resonance measurement [37, 38] is an optical technique that allows the study of the interaction between an immobilised molecule and an analyte that is in solution. No labelling is required. The method relies on the change in the refractive index of solutions adjacent to a surface upon an increase in mass, which is caused by analyte binding. In studies aimed at DNA–protein interactions, either the DNA molecule is attached, for instance biotinylated DNA on streptavidin-covered surface, or a protein is immobilised, either by tags such as poly-histidine or glutathione S-transferase or directly via free amino groups.

The system offers real-time recording of the association and dissociation of the analyte at the immobilised ligand, thus permitting rapid and accurate stoichiometric kinetic, affinity and thermodynamic measurements [39]. With regard to quantitative measurements, this procedure is currently the gold standard for protein interaction analysis.

Systematic evolution of ligands by exponential enrichment (SELEX)

SELEX has been used to identify high-affinity nucleic acid ligands for a large number of proteins. It exploits the power of genetic selection and the advantages of in vitro biochemistry. The assay is done by selecting a subset of oligonucleotides from a complex mixture of nucleic acid sequences. This is achieved by incubation of the DNA with the investigated protein, separation of bound molecules from the unbound fraction, release and amplification. This process is repeated iteratively until the nucleic acid sequences that bind to the protein with high affinity are enriched significantly [40]. The SELEX method enabled in vitro selection of the optimal binding sites of several transcription factors [41], for example.

A limitation is the fact that the system does not allow one to define the exact in vivo selectivity of proteins but is aimed at the identification of the best binding DNA targets. This fact also limits its ability for even semi-quantitative analysis.

Yeast one-hybrid system

The yeast one-hybrid system [42] is conceptually similarly designed to the yeast two-hybrid system that is used for the detection of protein–protein interactions. A DNA sequence of interest (the DNA bait) is cloned upstream of a reporter gene and integrated into the yeast genome by site-specific recombination. At the protein side, a hybrid protein is

generated by fusion of the prey protein to a transcription activation domain. When the prey protein and the DNA bait physically interact with each other, the reporter gene expression is activated. By generating libraries of DNA fragments and fusion proteins, complex analyses can be performed. With this system, 283 interactions between 72 *C. elegans* digestive tract gene promoters and 117 proteins could be identified [43], for example. A similar system but based on bacteria rather than yeast has also been described [44].

A big disadvantage of the system is its experimental complexity. Appropriate libraries need to be generated. Also, since the analysis occurs in vivo, chances of false results are substantial, since other processes may trigger or inhibit the transcriptional activity that is utilised for detection.

Microarray-based assays

For the achievement of high throughput, microarray technology was an obvious candidate, since it permits parallel analyses. Both DNA and protein microarrays were utilised to this end.

DNA microarrays

As reported above, ChIP-on-chip and microarray-based DamID improved the throughput of the basic methods considerably and are currently still the most widely used array-based methods for the identification of transcription factor binding sites. However, DNA microarrays can also be used for a direct analysis of DNA–protein interactions. Yet, since most sequence-specific DNA-binding proteins bind to double-stranded DNA, processes had to be developed to create double-stranded DNA molecules on chip surfaces. DNA microarrays used for other purposes are usually made of single-stranded DNA.

In some studies, PCR products were printed on microarrays [45]. As an alternative, and superior in terms of investigating the specificity of binding, analyses were performed on oligonucleotide arrays. Double-strand formation can be achieved by synthesising long oligonucleotides, which consist of self-complementary sequences, therefore forming hairpin structures (Fig. 3). Alternatively, the second DNA strand can be produced enzymatically by means of a sequence that is common to the 3′-end of all single-stranded oligonucleotides on the array. Upon addition of a complementary primer molecule and an appropriate polymerase, the initially single-stranded oligonucleotides are converted to DNA duplexes [46]. Also, a mixture of both methods can be applied, with an individual hairpin that occurs at the 3′-terminus of each oligonucleotide probe acting as an intramolecular primer for second strand synthesis. Critical to this is an attachment of the oligonucleotides via their 5′-end, which requires a special chemistry for in situ microarray synthesis, e.g. [47].

For monitoring protein binding, first the protein of interest is expressed. Frequently, an expression system has been used that adds an epitope tag. Such a tag has two functions: first, it is used to isolate the protein by affinity purification and, second, it permits detection by means of an epitope-specific reporter, such as an antibody. Alternatively, directly labelled proteins can be used in the assay. The protein is incubated on the microarray and the signal intensities obtained at the various array features are measured (e.g. Fig. 4). The method was applied successfully in several studies for an analysis of proteins that had been uncharacterised before [48, 49].

In terms of coverage, microarrays were used that contained PCR products, which covered entire genomes [48, 50]. Microarrays made of long PCR products have the advantage that they cover much sequence space with relatively few microarray features. At the same time, however, they have the disadvantage that the probability of calling a binding event correctly is less for a single binding site, which is embedded in a long rather than a short sequence. Moreover, depending on the number and types of binding sites within a single region, interaction may occur once or several times by one or several proteins at various degrees of affinity [51]. From the blend of information gained from such measurements, it may well be impossible to extract how many interactions are involved, let alone any accurate information about strength and specificity. For such ends, arrays that consist of short synthetic double-stranded oligonucleotides exhibit superior performance [52, 53]. While coverage of all binding sites in a genome is unlikely to be achieved with oligonucleotide arrays—and unnecessary because of the advent of sequencing techniques that are better suited for this purpose—arrays have been produced that are comprehensive with respect to the binding sequences, presenting all 10-bp sequences possible, for example [54, 55].

Protein microarrays

While analyses on DNA microarrays promise a rather comprehensive identification of the target sequences of a specific transcription factor or other DNA-binding protein, they are not able to identify the various proteins, which recognize a particular sequence of interest. This is made possible, however, by reversing the assay format. Studies on protein arrays that represented 282 potential yeast transcription factors [56], 802 *Arabidopsis* transcription factors [57] or 4,191 non-redundant human DNA-binding proteins [13], for example, demonstrated the potential of

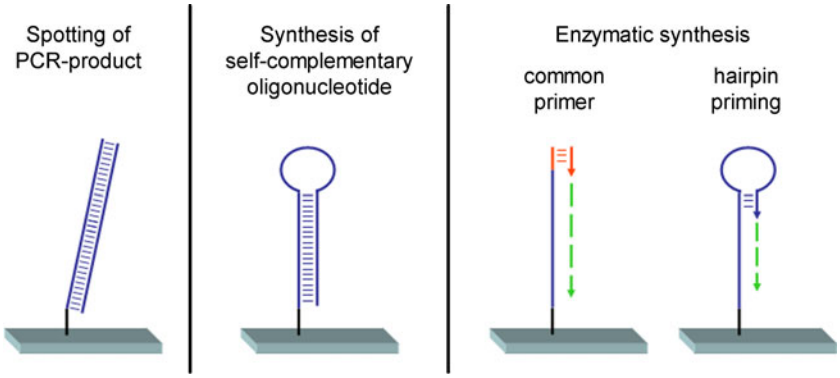

Fig. 3 Schematic presentation of different processes for producing microarrays with double-stranded DNA fragments. Either double-stranded DNA is spotted under conditions that do not denature the DNA on the chip surface. Alternatively, an oligonucleotide is spotted or synthesised, which is partly self-complementary and thus forms a hairpin structure. Duplexes can also be produced by using oligonucleotides, which have a common sequence at their 3′-end, which can be used for priming a polymerase reaction. Last, a short hairpin structure that is specific for each oligonucleotide could be utilised for priming

this approach. The arrays were created using epitope-tagged proteins, mostly by fusion to glutathione S-transferase. Subsequent to purification by affinity chromatography, the proteins were attached to the chip surface and incubated with double-stranded oligonucleotides. Each oligomer was made of three or four repeats of the relevant sequence motif. In humans, for example, a total of 17,718 DNA–protein interactions between 460 DNA motifs and 4,191 human proteins were identified [13]. Among them were a large number of interactions of transcription factors with DNA sequences, which had not been anticipated before. Also, the binding characteristics of transcription factors were identified this way.

Recent developments toward the production of comprehensive protein arrays [58–60] should permit an extension of the format. However, the structural integrity of the proteins or at least of their DNA-binding domains is of critical importance for a successful application and is unlikely to be conserved on a microarray surface for all molecules. In addition, problems could occur with proteins that need to form multimers or complexes with other proteins in order to exhibit their binding activity.

Proximity ligation

A technique that permits an analysis even at single-molecule level is proximity ligation [61]. It is a method for very sensitive solution-phase detection of interaction partners. While mostly useful for studying protein–protein interactions, the DNA binding of proteins can also be investigated. To this end, three probe molecules are required; one is an antibody against the investigated protein, the second is the DNA recognition sequence in the form of a double-stranded DNA. Both molecules have attached a single-stranded DNA tag segment. The third molecule needed is an oligonucleotide that is complementary to the ends of both DNA segments. The target protein binds to the DNA probe and is subsequently detected by the antibody. In consequence, the two DNA segments attached to antibody and DNA fragment will come in close proximity. Only then they can hybridise to the oligonucleotide, which acts as a connector (Fig. 5). In the resulting double-stranded sequence, the DNA segments of antibody and recognition sequence can be joined by an enzymatic ligation. The resulting sequence can be specifically ampli-

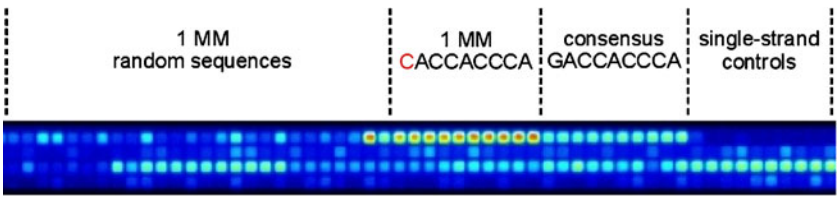

Fig. 4 Binding analysis on a DNA microarray. Analysis of the binding of transcription factor Gli2 to various DNA sequences was studies. Double-stranded oligonucleotides synthesised on a microarray. Directly fluorescence-labelled protein was incubated on the array. After washing, the signal intensities were recorded, which are equivalent to the respective affinity of Gli2. In the *top row*, particular sequences were present several times, including the consensus sequence GACCACCCA, a single-base variation (CACCACCCA) that exhibited even better binding of Gli2, and a number of random sequence permutations in one base that resulted in different signal intensities. At the *far right*, single-stranded controls are shown. The *spots* in *rows 2–4* also present permutations of the consensus sequence (unpublished data from our own studies)

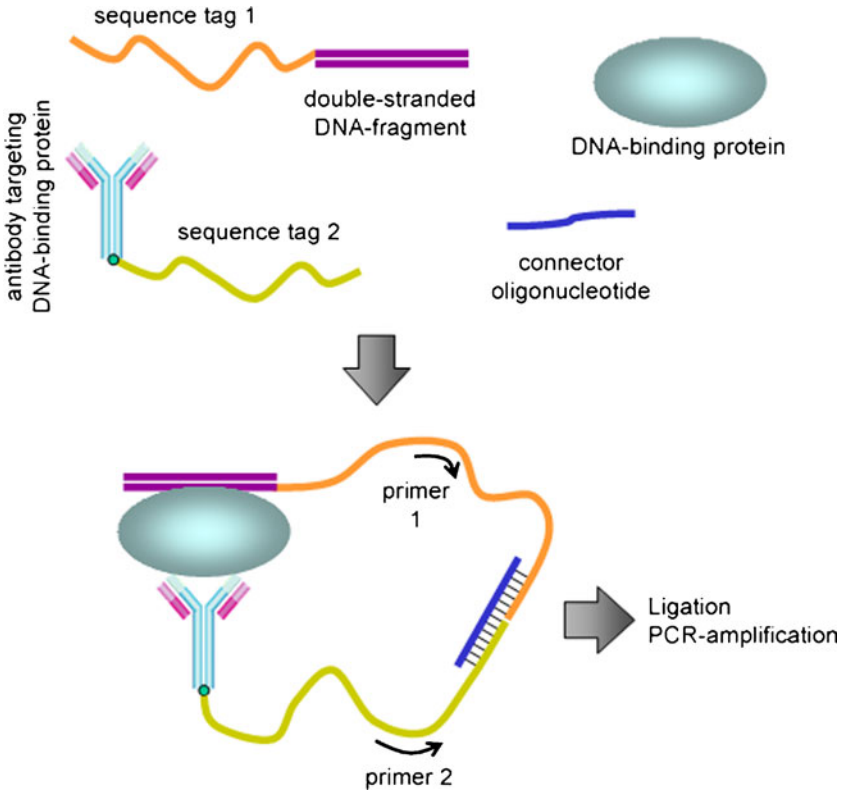

Fig. 5 Schematic presentation of a proximity ligation assay. Upon incubation of a DNA fragment, an appropriate protein and an antibody that binds the protein, a complex of the three molecules is formed. Both the DNA fragment and the antibody are tagged with single-stranded DNA segments. After addition of an oligonucleotide that is complementary to both ends of the tag sequences, the DNA molecules are joined by ligation and can then be amplified by PCR

fied and detected in real time by PCR or rolling circle amplification [62].

While this technique is powerful in terms of specificity and sensitivity, it is very dependent on the availability of specific antibodies. Also, the analysis requires extensive preparative steps. In principle, the approach permits a parallel investigation as long as specific sequences are attached to each pair of probe molecules. Also, the method is less suited for the identification of new binding sites, since no connector molecule would be available for their detection. For in vivo studies of defined protein–DNA interactions at the level of single-molecule sensitivity, however, the method could become enormously useful because of its sensitivity and specificity.

Other technologies

Atomic force microscopy (AFM) detects the nanometric features of both DNA and proteins. AFM imaging could therefore be used to locate protein binding sites and provide detailed structural analysis of binding protein-induced DNA conformations, such as bending or looping [63, 64]. Unfortunately, the resolution is currently still insufficient to read directly the DNA sequences or identify the proteins. However, the technique could be useful for the identification of the positioning of protein binding sites in large genomic areas, and the structural consequences caused by interaction or its absence.

X-ray crystallography and nuclear magnetic resonance (NMR) spectrometry [65, 66], on the other hand, are two techniques, which permit a detailed study the structures of defined DNA–protein complexes. However, they are of no use for the definition of interaction pairs or for the identification of their genomic locations.

Conclusions

Apart from the mere identification of more interaction pairs of DNA element and binding protein, which is an important objective for a comprehensive description of regulation at this level, studies in this area could also contribute more directly to biomedical applications. Several diseases are associated with functional single nucleotide polymorphisms (SNPs) or mutations in promoter regions, which could create, delete or modify binding sites and consequently increase or decrease the promoter activity [67–70]. Many

more such functionally relevant sequence variations will be found in future by programmes such as the International Cancer Genome Consortium [71]. In consequence, the demand for a functional characterisation of disease-associated mutations or SNPs in a promoter region will increase substantially and could add to the aim of accurate diagnosis and personal therapy. Even quantitative differences could be studied in a parallel manner [72], thus adding an analogous level of regulation on top of the digital information of binding or no binding. Recent developments could also improve sensitivity down to single-molecule detection, either by applying relevant signal amplification schemes [73] or technical developments that permit detection of individual fluorophore molecules [74], for example. In addition, the advances in sequencing technology will revolutionise studies further toward really genome-wide identification and analyses of regulative active DNA elements. In combination, high sensitivity and quantitative rather than purely qualitative or relative measurements will add another level of quality to the understanding of the regulation and functional consequences and may permit the study of the DNA–protein interactome comprehensively not only in terms of identification of participating molecules but also in terms of interaction strength and competition.

Overall, technology for the detection of DNA–protein interaction has progressed enormously during the last 40 years. From initial qualitative in vitro assays on individual interaction partners, through to high-throughput, single-molecule sensitivity, in vivo detection and quantification have been achieved, although not yet combined in a single system. Such a system, however, is unlikely to be feasible. More sensible and applicable will be assays that combine a few of the above-mentioned factors according to the actual requirements of the analyses performed. This combination, possibly including even the measurement of detailed structural information on the way, is the challenge of the years to come.

Acknowledgements We are grateful to Christian Kraus for provision of Fig. 4. Work on protein–DNA interaction in our laboratory is funded by the German Federal Ministry of Education and Research (BMBF) as part of the NGFNplus *PaCaNet* project as well as the *Regulatory Genomics* and *DropTop* projects funded by the European Commission. R.H. is supported by a PhD fellowship of the Egyptian Government.

References

1. Bain G, Maandag EC, Izon DJ, Amsen D, Kruisbeek AM, Weintraub BC, Krop I, Schlissel MS, Feeney AJ, van Roon M, Vandervalk M, Teriele HPJ, Berns A, Murre C (1994) E2A proteins are required for proper B cell development and initiation of immunoglobulin gene rearrangements. Cell 79:885–892
2. Dynlacht BD (1997) Regulation of transcription by proteins that control the cell cycle. Nature 389:149–152
3. Accili D, Arden KC (2004) FoxOs at the crossroads of cellular metabolism differentiation and transformation. Cell 117:421–426
4. Lemon B, Tjian R (2000) Orchestrated response: a symphony of transcription factors for gene control. Genes Dev 14:2551–2569
5. Furney SJ, Higgins DG, Ouzounis CA, Lopez-Bigas N (2006) Structural and functional properties of genes involved in human cancer. BMC Genomics 7:3
6. Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W, Cherry JM, Henikoff S, Skupski MP, Misra S, Ashburner M, Birney E, Boguski MS, Brody T, Brokstein P, Celniker SE, Chervitz SA, Coates D, Cravchik A, Gabrielian A, Galle RF, Gelbart WM, George RA, Goldstein LSB, Gong FC, Guan P, Harris NL, Hay BA, Hoskins RA, Li JY, Li ZY, Hynes RO, Jones SJM, Kuehl PM, Lemaitre B, Littleton JT, Morrison DK, Mungall C, O'Farrell PH, Pickeral OK, Shue C, Vosshall LB, Zhang J, Zhao Q, Zheng XQH, Zhong F, Zhong WY, Gibbs R, Venter JC, Adams MD, Lewis S (2000) Comparative genomics of the eukaryotes. Science 287:2204–2215
7. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254
8. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M (2003) Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science 301:71–76
9. Hughes JR, Cheng JF, Ventress N, Prabhakar S, Clark K, Anguita E, De Gobbi M, de Jong P, Rubin E, Higgs DR (2005) Annotation of cis-regulatory elements by identification subclassification and functional assessment of multispecies conserved sequences. Proc Natl Acad Sci U S A 102:9830–9835
10. Venkatesh B, Yap WH (2005) Comparative genomics using fugu: a tool for the identification of conserved vertebrate cis-regulatory elements. Bioessays 27:100–107
11. Cliften PF, Hillier LW, Fulton L, Graves T, Miner T, Gish WR, Waterston RH, Johnston M (2001) Surveying Saccharomyces genomes to identify functional elements by comparative DNA sequence analysis. Genome Res 11:1175–1186
12. 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 Res 14:2041–2047
13. Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, Rho HS, Woodard C, Wang H, Jeong JS, Long SY, He XF, Wade H, Blackshaw S, Qian J, Zhu H (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. Cell 139:610–622
14. Riggs AD, Bourgeois S, Newby RF, Cohn M (1968) DNA binding of the lac repressor. J Mol Biol 34:365–368
15. Riggs AD, Bourgeois S, Cohn M (1970) The lac repressor-operator interaction. 3. Kinetic studies. J Mol Biol 53:401–417
16. Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci U S A 76:4350–4354
17. Woodbury CP Jr, von Hippel PH (1983) On the determination of deoxyribonucleic acid-protein interaction parameters using the nitrocellulose filter-binding assay. Biochemistry 22:4730–4737
18. Beattie KL, Wiegand RC, Radding CM (1977) Uptake of homologous single-stranded fragments by superhelical DNA. II. Characterization of the reaction. J Mol Biol 116:783–803
19. Tullius TD (1989) Physical studies of protein-DNA complexes by footprinting. Annu Rev Biophys Biophys Chem 18:213–237
20. Galas DJ, Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5:3157–3170

21. Brenowitz M, Senear DF, Shea MA, Ackers GK (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. Methods Enzymol 130:132–181
22. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560
23. Siebenlist U, Gilbert W (1980) Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. Proc Natl Acad Sci U S A 77:122–126
24. Yang VW (1998) Eukaryotic transcription factors: identification characterization and functions. J Nutr 128:2045–2051
25. Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9:6505–6525
26. Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. Nucleic Acids Res 9:3047–3060
27. Carey J (1991) Gel retardation. Methods Enzymol 208:103–117
28. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc 2:1849–1861
29. Fried MG, Bromberg JL (1997) Factors that affect the stability of protein-DNA complexes during gel electrophoresis. Electrophoresis 18:6–11
30. Massie CE, Mills IG (2008) ChIPping away at gene regulation. EMBO Rep 9:337–343
31. Hoffman BG, Jones SJ (2009) Genome-wide identification of DNA-protein interactions using chromatin immunoprecipitation coupled with flow cell sequencing. J Endocrinol 201:1–13
32. Wu J, Smith LT, Plass C, Huang TH (2006) ChIP-chip comes of age for genome-wide functional analysis. Cancer Res 66:6899–6902
33. Mardis ER (2007) ChIP-seq: welcome to the new frontier. Nat Methods 4:613–614
34. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431:99–104
35. Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. Nat Protoc 2:1467–1478
36. Lee TI, Johnstone SE, Young RA (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc 1:729–748
37. Johne B, Gadnell M, Hansen K (1993) Epitope mapping and binding kinetics of monoclonal antibodies studied by real time biospecific interaction analysis using surface plasmon resonance. J Immunol Methods 160:191–198
38. Buckle M, Williams RM, Negroni M, Buc H (1996) Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance. Proc Natl Acad Sci U S A 93:889–894
39. Rich RL, Myszka DG (2006) Survey of the year 2005 commercial optical biosensor literature. J Mol Recognit 19:478–534
40. 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. Mol Cell Biol 9:2944–2949
41. 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. Nat Biotechnol 20:831–835
42. Deplancke B, Dupuy D, Vidal M, Walhout AJ (2004) A gateway-compatible yeast one-hybrid system. Genome Res 14:2093–2101
43. Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stamm L, Reece-Hoyes JS, Hope IA, Tissenbaum HA, Mango SE, Walhout AJM (2006) A gene-centered C elegans protein-DNA interaction network. Cell 125:1193–1205
44. Noyes MB, Christensen RG, Wakabayashi A, Stormo GD, Brodsky MH, Wolfe SA (2008) Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. Cell 133:1277–1289
45. Bulyk ML (2006) DNA microarray technologies for measuring protein-DNA interactions. Curr Opin Biotechnol 17:422–430
46. Bulyk ML, Gentalen E, Lockhart DJ, Church GM (1999) Quantifying DNA-protein interactions by double-stranded DNA arrays. Nat Biotechnol 17:573–577
47. Beier M, Stephan A, Hoheisel JD (2001) Synthesis of photolabile 5′-O-phosphoramidites for the production of microarrays of inversely oriented oligonucleotides. Helv Chim Acta 84:2089–2095
48. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. Science 290:2306–2309
49. Lieb JD, Liu X, Botstein D, Brown PO (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28:327–334
50. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional regulatory networks in Saccharomyces cerevisiae. Science 298:799–804
51. Berger MF, Bulyk ML (2006) Protein binding microarrays (PBMs) for rapid high-throughput characterization of the sequence specificities of DNA binding proteins. Methods Mol Biol 338:245–260
52. 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. Nat Genet 36:1331–1339
53. Berger MF, Badis G, Gehrke AR, Talukder S, Philippakis AA, Pena-Castillo L, Alleyne TM, Mnaimneh S, Botvinnik OB, Chan ET, Khalid F, Zhang W, Newburger D, Jaeger SA, Morris QD, Bulyk ML, Hughes TR (2008) Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell 133:1266–1276
54. 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. Proc Natl Acad Sci U S A 103:867–872
55. Berger MF, Bulyk ML (2009) Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat Protoc 4:393–411
56. Ho SW, Jona G, Chen CT, Johnston M, Snyder M (2006) Linking DNA-binding proteins to their recognition sequences by using protein microarrays. Proc Natl Acad Sci U S A 103:9940–9945
57. Gong W, He K, Covington M, Dinesh-Kumar SP, Snyder M, Harmer SL, Zhu YX, Deng XW (2008) The development of protein microarrays and their applications in DNA-protein and protein-protein interaction analyses of Arabidopsis transcription factors. Mol Plant 1:27–41
58. He M, Taussig MJ (2001) Single step generation of protein arrays from DNA by cell-free expression and in situ immobilisation (PISA method). Nucleic Acids Res 29:E73
59. Ramachandran N, Hainsworth E, Bhullar B, Eisenstein S, Rosen B, Lau AY, Walter JC, LaBaer J (2004) Self-assembling protein microarrays. Science 305:86–90

60. Angenendt P, Kreutzberger J, Glokler J, Hoheisel JD (2006) Generation of high density protein microarrays by cell-free in situ expression of unpurified PCR products. Mol Cell Proteomics 5:1658–1666
61. Gustafsdottir SM, Schlingemann J, Rada-Iglesias A, Schallmeiner E, Kamali-Moghaddam M, Wadelius C, Landegren U (2007) In vitro analysis of DNA-protein interactions by proximity ligation. Proc Natl Acad Sci U S A 104:3067–3072
62. Conze T, Shetye A, Tanaka Y, Gu JJ, Larsson C, Goeransson J, Tavoosidana G, Soederberg O, Nilsson M, Landegren U (2009) Analysis of genes, transcripts, and proteins via DNA ligation. Ann Rev Anal Chem 2:215–239
63. Bustamante C, Rivetti C (1996) Visualizing protein-nucleic acid interactions on a large scale with the scanning force microscope. Annu Rev Biophys Biomol Struct 25:395–429
64. Hansma HG (2001) Surface biology of DNA by atomic force microscopy. Annu Rev Phys Chem 52:71–92
65. Conlin RM, Brown RS (2001) Reconstitution of protein-DNA complexes for crystallization. Methods Mol Biol 148:547–556
66. Moss T (2001) DNA-protein interaction: principles and protocols, 2nd edn. Humana, Totowa, NJ
67. Masotti C, Armelin-Correa LM, Splendore A, Lin CJ, Barbosa A, Sogayar MC, Passos-Bueno MR (2005) A functional SNP in the promoter region of TCOF1 is associated with reduced gene expression and YY1 DNA-protein interaction. Gene 359:44–52
68. Taulan M, Lopez E, Guittard C, Rene C, Baux D, Altieri JP, DesGeorges M, Claustres M, Romey MC (2007) First functional polymorphism in CFTR promoter that results in decreased transcriptional activity and Sp1/USF binding. Biochem Biophys Res Commun 361:775–781
69. Champigny MJ, Mitchell M, Fox-Robichaud A, Trigatti BL, Igdoura SA (2009) A point mutation in the neu1 promoter recruits an ectopic repressor Nkx32 and results in a mouse model of sialidase deficiency. Mol Genet Metab 97:43–52
70. Hoff JH, te Morsche RH, Roelofs HM, van der Logt EM, Nagengast FM, Peters WH (2009) COX-2 polymorphisms -765G→C and -1195A→G and colorectal cancer risk. World J Gastroenterol 15:4561–4565
71. The International Cancer Genome Consortium (2010) International network of cancer genome projects. Nature 464:993–998
72. Marcy Y, Cousin PY, Rattier M, Cerovic G, Escalier G, Bena G, Gueron M, McDonagh L, le Boulaire F, Benisty H, Qeisbuch C, Avarre JC (2008) Innovative integrated system for real-time measurement of hybridization and melting on standard format microarrays. Biotechniques 44:913–920
73. Ericsson O, Jarvius J, Schallmeiner E, Howell M, Nong RY, Reuter H, Hahn M, Stenberg J, Nilsson M, Landegren U (2008) A dual-tag microarray platform for high-performance nucleic acid and protein analyses. Nucleic Acids Res 36:e45
74. Hesse J, Jacak J, Kasper M, Regl G, Eichberger T, Winklmayr M, Aberger F, Sonnleitner M, Schlapak R, Howorka S, Muresan L, Frischauf AM, Schütz GJ (2006) RNA expression profiling at the single molecule level. Genome Res 16:1041–1045