Acta Biochimica Polonica

Vol. 51 No. 1/2004

1–8

QUARTERLY

Minireview

DNA microarrays, a novel approach in studies of chromatin structure✪

Piotr Widłak✉

Department of Experimental and Clinical Radiobiology, Center of Oncology, Gliwice, Poland

Received: 11 December, 2003; revised: 24 January, 2004; accepted: 26 January, 2004

Key words: DNA microarray, genomics, epigenomics, chromatin, nucleosomes

The DNA microarray technology delivers an experimental tool that allows surveying expression of genetic information on a genome-wide scale at the level of single genes — for the new field termed functional genomics. Gene expression profiling — the primary application of DNA microarrays technology — generates monumental amounts of information concerning the functioning of genes, cells and organisms. However, the expression of genetic information is regulated by a number of factors that cannot be directly targeted by standard gene expression profiling. The genetic material of eukaryotic cells is packed into chromatin which provides the compaction and organization of DNA for replication, repair and recombination processes, and is the major epigenetic factor determining the expression of genetic information. Genomic DNA can be methylated and this modification modulates interactions with proteins which change the functional status of genes. Both chromatin structure and transcriptional activity are affected by the processes of replication, recombination and repair. Modified DNA microarray technology could be applied to genome-wide study of epigenetic factors and processes that modulate the expression of genetic information. Attempts to use DNA microarrays in studies of chromatin packing state, chromatin/DNA-binding protein distribution and DNA methylation pattern on a genome-wide scale are briefly reviewed in this paper.

Completion of the Human Genome Project has opened a new era in studies of functions of cells and organisms. Identification of the thousands of genes forming genomes brings us to the next frontier: elucidation of the functions of these genes and their interactions —

✪This work was supported by the Ministry of Science and Informatics, Grant KBN 4T11F01824.

✉Address for correspondence: Piotr Widłak, Department of Experimental and Clinical Radiobiology, Center of Oncology, Wybrzeże AK 15, Gliwice 44-100, Poland, tel.: (48 32) 278 9672; fax: (48 32) 231 3512; e-mail: widlak@io.gliwice.pl

to "functional genomics". An experimental tool that allows surveying expression of the genetic information on a genome-wide scale at the level of single genes has been developed just a few years ago thanks to the microarray technology. The basic concept behind DNA microarrays is the precise positioning of DNA probes at a high density in a way that they can work as molecular detectors. Different variations of this technology are in use, yet the most common are cDNA microarrays, where DNA fragments are spotted on a solid surface, and oligonucleotide microarrays, where oligonucleotides are synthesized *in situ*. In original protocols immobilized DNA probes were hybridized to cDNA obtained after reverse transcription of RNA samples, either poly(A)mRNA or total RNA. In the initial procedure probes were labeled directly, whereby reverse transcription of mRNA was primed using a poly(dT) primer in the presence of fluorescently labeled nucleotides. More advanced protocols involve repeated rounds of RNA amplification, which increases the sensitivity and reduces the required amount of starting RNA (reviewed in: Holloway *et al.*, 2002). Gene expression microarrays allow one to study transcripts of thousands of genes simultaneously, generating "gene expression profiles". Such "expression profiles" have been successfully used in medical research and biotechnology. On the medical field, expression microarrays allow detailed documentation of responses of cells and tissues to both disease and therapeutic treatment, and facilitate classification of samples (reviewed in: Gerhold *et al.*, 2002). Gene expression profiling – the primary application of DNA microarray technology – generates monumental amounts of information concerning the functioning of a cell type or tissue specimen. However, the expression of genetic information is regulated by several factors that cannot be directly targeted by standard gene expression profiling.

The genetic material of eukaryotic cells is packed into a nucleoprotein complex termed chromatin. This packing of DNA provides the compaction and organization for the replication, repair and recombination processes, and is the major epigenetic factor determining the expression of genetic information. The fundamental structural unit of chromatin is the nucleosome, which contains 146 base pairs of DNA wrapped around an octamer of core histones (so called core particle). In addition, the nucleosome consists of a linker region of variable length (generally less than 50 base pairs), which interacts with the linker histone H1 and/or other non-histone proteins (e.g. certain HMG proteins). The polynucleosomal chain is further looped and folded into various higher order structures. The organization of chromatin domains (or loops) seems to be maintained by anchorage of specific DNA sequences into a protein network of the nucleoskeleton. Nucleosomes positioned in the regulatory regions of genes often hinder the accessibility of binding sites for transcription factors. In addition, formation of specific chromatin structures leads to transcriptional repression of chromatin domains or regions (*e.g.* formation of highly packed/condensed heterochromatin). Among the different mechanisms that act on the level of chromatin to activate transcription are: non-covalent ATP-dependent remodeling of nucleosome structure and covalent post-translational modifications of histones (so called "histone code"). The most common post-translational modification of core histones is acetylation/deacetylation of conserved lysine residues catalyzed by specific histone acetyltransferases and deacetylases. At the moment, it is commonly accepted that structural transitions in chromatin caused by histone modifications and/or chromatin remodeling "machines" facilitate the binding of regulatory proteins to gene promoters, which allows assembly of the RNA polymerase complex to activate transcription. It is generally believed that the open/uncondensed chromatin state, which allows access of transcription factors and RNA polymerases to the tem-

plate, is typical for regions where active (or potentially active) genes are located. On the other hand, non-active repressed genes are located primarily in regions of packed/condensed chromatin (heterochromatin) (reviewed in: Groudine & Felsenfeld, 2003; Fry & Peterson, 2001).

Because of technical limitations, the knowledge about the actual state of chromatin packing/condensation and its relationship to transcriptional activity was until recently restricted to a small number of genes studied in a few model organisms. The DNA microarray technology delivered the unique opportunity to survey the chromatin structure on a genome-wide scale at the resolution of single genes. In fact, modified DNA microarray technology has already been applied to genome-wide study of epigenetic factors and processes that regulate the expression of genetic information (reviewed in: Pollack & Iyer, 2002). This new field could be termed "epigenomics" (Novik *et al.*, 2002). This paper briefly describes attempts to use DNA microarrays in studies of chromatin structure on a genome-wide scale.

METHODOLOGY

The initial implementation of DNA microarray technology into genome structural research was comparative genomic hybridization (CGH) array, which allowed high resolution analysis of gene copy number (Solinas-Toldo *et al.*, 1997; Pinkel *et al.*, 1998). The primary difference between gene expression microarrays and the CGH array is replacement of RNA samples with DNA ones as a starting material. Two DNA samples are labeled with different fluorophores and co-hybridized to a DNA microarray, and their fluorescence ratio represents the relative DNA copy number. Similar strategies could be applied to study other aspects of genome structure: "test" and "reference" DNA samples that are differentially labeled and co-hybridized to a DNA microarray, either "standard" or "specialized" (e.g. microarrays of promoter sequences or CpG islands). DNA could be fluorescence labeled either during PCR amplification or without amplification. The most essential step in such "structural" array protocols is initial isolation/fractionation of genomic DNA in a way that would reflect the problem to be analyzed. Several principles that lie behind such fractionation procedures are listed below.

Differential physicochemical characteristics of nucleoprotein complexes

One of such strategies, originally described by Garrard and coworkers (reviewed in: Huang & Garrard, 1988), has been used to fractionate chromatin based on differential solubility of histone H1-containing and histone H1-free nucleosomes. Isolated nuclei were briefly incubated at "physiological" ionic strength with micrococcal nuclease, which specifically cleaves internucleosomal linker DNA. That treatment solubilized 10–20% of the chromatin, which was collected as the first supernatant fraction termed S1. After removal of salt an additional 50–60% of the chromatin was solubilized, which was collected as the second supernatant fraction termed S2. The S1 fraction contained primarily mononucleosomes lacking histone H1 while S2 consisted of histone H1-containing oligonucleosomal particles. Another strategy to fractionate genomic DNA based on specific nucleoprotein complexes that seems to be potentially applicable to DNA microarray analysis would be isolation of nuclear matrix-attached DNA (Sumer *et al.*, 2003). The nuclear matrix is a putative skeletal structure isolated from nuclei after removal of the majority of DNA and chromatin proteins. Such a residual fraction obtained after treatment of nuclei with nucleases and high salt buffers contains 5–10% of the total genomic DNA putatively involved in chromatin organization and regulation of the genome metabolism.

Immunoprecipitation

This strategy is based on the chromatin immunoprecipitation (ChIP) technique and allows one to study the interactions of genomic DNA with any specific protein. In the standard ChIP procedure proteins are crosslinked to their binding sites in chromatin by formaldehyde, the chromatin is sheared to possess DNA fragments of about 500 bp in length, and then the chromatin containing the proteins of interest is immunoprecipitated using specific antibody. After reversal of the crosslinks and removal of proteins, amplified and labeled DNA is hybridized to a microarray. This allows identification of DNA sequences enriched in a specific protein-bound fraction, and could be used to study binding distribution of transcription factors, chromatin proteins, replication/recombination proteins, DNA methylases and any other proteins interacting with the genome (see below).

DNA size fractionation

This strategy could be used to identify DNA sequences that are substrates of specific DNA fragmentation factors (e.g. structure-specific nucleases), either in chromatin or after DNA purification. Digested fragments could be isolated by size fractionation (e.g. by gel electrophoresis), and then labeled and hybridized to microarrays. Sequences containing nuclease-sensitive structures or located within nuclease-sensitive domains would be enriched in such "shortened" fraction. An alternative procedure includes PCR amplification from ligated linkers. After linker-ligation, randomly fragmented genomic DNA is digested with structure-specific nucleases (e.g. methylation-sensitive restriction enzymes), PCR-amplified/labeled and hybridized to microarrays. Fragments containing nuclease-sensitive sites are not amplified and under-represented in hybridization profiles. The size-dependent strategy was successfully used to identify DNA methylation pattern (Huang *et al.,* 1999) and chromatin nuclease-sensitivity (Weil *et al.,* 2004).

GLOBAL SURVEY OF CHROMATIN STRUCTURE

Chromatin packing arrays

In order to isolate the open/loosely packed chromatin fraction putatively enriched in active (or potentially active) genes and the condensed/packed chromatin fraction putatively enriched in repressed/inactive genes we took advantage of differential solubility of nucleosomes containing or lacking histone H1, because this histone is more abundant in condensed chromatin (Garrard, 1991). DNA purified from the S1 (histone H1-deficient) or S2 (histone H1-containing) chromatin fraction was labeled and co-hybridized with RNA or total genomic DNA to a cDNA microarray (Widłak & Fujarewicz, 2003; Weil *et al.,* 2004). To identify genes enriched in the condensed/packed chromatin DNA signals were compared between the S2 and S1 chromatin fractions or between the S2 chromatin fraction and total non-fractionated genomic DNA. Although a large portion of genes showed random distribution between the fractions, clusters of genes enriched in the S1 and in the S2 fractions could be distinguished. We found a clear correlation between the packing/condensation state and the transcription activity: inactive genes were enriched in the S2 fraction while active ones were depleted (Widłak & Fujarewicz, 2003; Weil *et al.,* 2004). The results obtained confirmed the initial speculation that the nuclear fraction containing the histone H1-rich packed/condensed chromatin would be enriched in inactive genes while the histone H1-deficient less condensed/open chromatin would be enriched in active genes.

Another experimental approach that has been used in a global survey of chromatin packing/condensation state is based on nuclease sensitivity. Sites in chromatin that

have already bound transcription factors, or that will allow their binding, are experimentally detected as nuclease hypersensitive sites. Hypersensitive sites are usually indicators of regulatory regions of potentially active or active genes, and transcriptionally active chromatin is generally more sensitive to nucleases as compared to inactive heterochromatin (Gross & Garrard, 1988). A method of chromatin fractionation was implemented that based on differential accessibility to DNase I and recovery of the packed/condensed fraction by fragment length selection for the highest mass and therefore the most protected DNA fragments. The fraction of nuclease resistant DNA was co-hybridized to cDNA micrarrays together with total genomic DNA, which allowed identification of genes enriched in the packed/condensed chromatin fraction. As expected, the expression of genes was inversely proportional to their packing state. More interestingly, fractions of genes enriched in the packed/condensed chromatin fraction isolated according to either the nuclease resistance or the chromatin solubility methods essentially (80%) overlapped thereby validating the concept of chromatin packing arrays (Weil *et al.,* 2004). The fact that the chromatin condensation state established using "chromatin packing arrays" correlates with the transcriptional activity state is particularly important in the case of genes whose expression level is below the detection threshold and could enable the discovery of new expressed genes.

Global distribution of specific chromatin proteins

Chromatin structure is strongly influenced by the presence of specific histone variants, their post-translational modifications and non-histone proteins. Heterochromatin usually contains hypo-acetylated and hyper-methylated core histones (and histone H1 as well). Consequently, the presence of hyper-acetylated histones and/or specific histone acetyltranferases might be indicative for open transcriptionally active chromatin. The "ChIP-on-microarray" (or "ChIP on CHIP") strategy has been used to map the genomic distribution of acetylated domains in yeast based on immunoprecipitation of acetylated histones H3 and H4, and histone acetyltransferases or deacetylases (Reid *et al.*, 2001; Robyr *et al.*, 2002; Kurdistani *et al.*, 2002). It has been evidenced that different histone deacetylases and acetyltransferases are specific for different genes and chromatin domains. The distribution of the Sir2 deacetylase and other Sir proteins responsible for chromatin silencing in yeast has also been mapped (Lieb *et al.*, 2001). Interestingly, similar experiments confirmed the association of Sir proteins with the telomere-specific proteins Rap and Rif at chromosome ends and subtelomeric regions (Smith *et al.*, 2003). The same "ChIP-on-microarray" strategy has been used to map the interactions of the RSC chromatin remodeling complex. Identification of about 700 RSC physiological targets in the yeast genome showed that the complex is generally recruited to Pol III promoters while recruitment to Pol II ones required specific transcriptional activators and repressors (Ng *et al.*, 2002). It has also been shown that the RSC complex is involved in modulation of expression of genes regulated by stress (Damelin *et al.*, 2002).

Binding distribution of transcription factors

The expression of genes is regulated by both chromatin structure and binding of specific transcription factors. To understand the roles of individual transcription factors in the regulation of expression of genetic information it is absolutely essential to map their physical interactions with their *in vivo* targets. The DNA microarray technology combined with immunoprecipitation of regulatory proteins and associated DNA fragments is the method

of choice to study global binding distribution of transcription factors. The knowledge about the global distribution of transcription factors binding sites would allow one to identify their directly targeted genes and determine actual clustering of different transcription factors in particular regulatory regions of genes *in vivo*. Such genome-wide maps of binding sites have been established for several yeast transcription factors (Ren *et al.*, 2000). Data on the binding distribution of the yeast cell-cycle regulators SBF and MBF enabled the clarification of the network of interactions involved in the regulation of the cell cycle (Iyer *et al.*, 2001). Combination of expression profiling with binding distribution data proved that indeed cell cycle transcriptional activators that function during one stage of the cycle directly targeted activators specific for the next stage (Simon *et al.*, 2001).

DNA replication, recombination and repair

The "ChIP-on-microarray" strategy can be applied to map the genomic distribution of proteins specifically involved in genome metabolism: the replication, recombination and repair. DNA replication origins are essential elements contributing to the propagation of chromosomes. Such elements could be identified due to specific proteins that form pre-replicative complexes. The binding distribution of ORC and MCM proteins allowed the mapping of about 400 replication origins in yeast chromosomes (Wyrick *et al.*, 2001). Similarly, association of the Spo11 recombinase with recombination initiation sites allowed the indentification of meiotic recombination hotspots and coldspots in the yeast genome (Gerton *et al.*, 2000).

CpG methylation arrays

In addition to protein modification, gene expression is regulated by DNA modifications, mostly methylation. Mammalian DNA can be methylated at cytosines and indeed a large portion of genomic DNA is methylated in CpG dinucleotides. The essential exception are the GC-rich "CpG islands" in active gene promoters that are generally hypo-methylated. The pattern of DNA methylation specific for particular cells is maintained by DNA cytosine-5-methyltranferases (like Dnmt1), which preferentially modify hemimethylated DNA resulting from replication. Aberrations in DNA methylation patterns could be indicators of gene malfunctions and frequently contribute to the development of cancer. DNA methylation and chromatin modification structurally and functionally cooperate in repression of gene expression. It has been shown that methyl-CpG-binding proteins (like MeCP2) recruit histone deacetylases and other co-repressors to targeted sequences. Thus, DNA methylation could pattern other chromatin modifications (reviewed in: Ng & Bird, 1999).

Genome-wide screening of hypermethylated CpG islands was allowed by a DNA array-based method called differential methylation hybridization (DMH) arrays. This technique involves cleavage of DNA with methylation-sensitive restriction enzymes and hybridization of PCR-amplified/labeled fragments to a microarray comprised of cloned CpG islands (Huang *et al.*, 1999). DMH arrays were used to detect methylation patterns in human cancer. It has been shown that hypermethylation of CpG islands correlates with the stage of breast (Yan *et al.*, 2000) and ovarian (Wei *et al.*, 2002) cancer. An alternative technique to study DNA methylation patterns is called methylation-specific oligonucleotide (MSO) array. In this method genomic DNA is modified with bisulphite, which converts unmethylated cytosines to thymine. Modified DNA is then PCR-amplified/labeled and hybridized to an oligonucleotide microarray designed to discriminate between the bisulphite-converted TpG dinucleotides and the methylation-protected CpG dinucleotides (Gitan *et al.*, 2002).

The author wishes to thank William T. Garrard for ideas, comments and help in preparation of the manuscript.

REFERENCES

Damelin M, Simon I, Moy TI, Wilson B, Komili S, Tempst P, Roth FP, Young RA, Cairns BR, Silver PA. (2002) The genome-wide localization of Rsc9, a component of the RSC chromatin-remodeling complex, changes in response to stress. *Mol Cell.*; **9**: 563–73.

Felsenfeld G, Groudine M. (2003) Controlling the double helix. *Nature.*; **421**: 448–53.

Fry CJ, Peterson CL. (2001) Chromatin remodeling enzymes: who's on first? *Curr Biol.*; **11**: R185–97.

Garrard WT. (1991) Histone H1 and the conformation of transcriptionally active chromatin. *Bioessays.*; **13**: 87–8.

Gerhold DL, Jensen RV, Gullans SR. (2002) Better therapeutics through microarrays *Nat Genet.*; **32** Suppl: 547–52.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. (2000) Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A.*; **97**: 11383–90.

Gitan RS, Shi H, Chen CM, Yan PS, Huang TH. (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.*; **12**: 158–64.

Gross DS, Garrard WT. (1998) Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem.*; **57**: 159–97.

Holloway AJ, van Laar RK, Tothill RW, Bowtell DDL. (2002) Options available — from start to finish — for obtaining data from DNA microarrays II. *Nat Genet.*; **32** Suppl: 481–9.

Huang SY, Garrard WT. (1989) Electrophoretic analyses of nucleosomes and other protein–DNA complexes. *Methods Enzymol.*; **170**: 116–42.

Huang TH, Perry MR, Laux DE. (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet.*; **8**: 459–70.

Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature.*; **409**: 533–8.

Kurdistani SK, Robyr D, Tavazoie S, Grunstein M. (2002) Genome-wide binding map of the histone deacetylase Rpd3 in yeast. *Nat Genet.*; **31**: 248–54.

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**: 327–34.

Ng HH, Bird A. (1999) DNA methylation and chromatin modification. *Curr Opin Genet Dev.*; **9**: 158–63.

Ng HH, Robert F, Young RA, Struhl K. (2002) Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. *Genes Dev.*; **16**: 806–19.

Novik KL, Nimmrich I, Genc B, Maier S, Piepenbrock C, Olek A, Beck S. (2002) Epigenomics: genome-wide study of methylation phenomena. *Curr Issues Mol Biol.*; **4**: 111–28.

Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet.*; **20**: 207–11.

Pollack JR, Iyer VR. (2002) Characterizing the physical genome. *Nat Genet.*; **32** Suppl: 515–21.

Reid JL, Iyer VR, Brown PO, Struhl K. (2000) Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Mol Cell.*; **6**: 1297–307.

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**: 2306–9.

Robyr D, Suka Y, Xenarios I, Kurdistani SK, Wang A, Suka N, Grunstein M. (2002) Microarray deacetylation maps determine genome-wide functions for yeast histone deacetylase. *Cell.*; **109**: 437–46.

Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell.*; **106**: 697–708.

Smith CD, Smith DL, DeRisi JL, Blackburn EH. (2003) Telomeric protein distribution and remodeling through the cell cycle in *Saccharomyces cerevisiae*. *Mol Biol Cell.*; **14**: 556–70.

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Dohner H, Cremer T, Lichter P. (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer.*; **20**: 399–407.

Sumer H, Craig JM, Sibson M, Choo KH. (2003) A rapid method of genomic array analysis of scaffold/matrix attachment regions (S/MARs) identifies a 2.5-Mb region of enhanced scaffold/matrix attachment at human neocentromere. *Genome Res.*; **13**: 1737–43.

Wei SH, Chen CM, Strathdee G, Harnsomburana J, Shyu CR, Rahmatpanah F, Shi H, Yan PS, Nephew KP, Brown R, Huang TH. (2002) Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin Cancer Res.*; **8**: 2246–52.

Weil MR, Widlak P, Minna JD, Garner HR. (2004) Global survey of chromatin accessibility using DNA microarrays. *Genome Res.*; in press.

Widłak P, Fujarewicz K. (2003) The analysis of chromatin condensation state and transcriptional activity using DNA microarrays. *J Med Inform Technol.*; **6**: 13–9.

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM. (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science.*; **294**: 2357–60.

Yan PS, Perry MR, Laux DE, Asare AL, Caldwell CW, Huang TH. (2000) CpG islands arrays: an application toward deciphering epigenetic signatures of breast cancer. *Clin Cancer Res.*; **6**: 1432–8.