DNA microarrays, chromatin, transcription

Piotr WIDŁAK[*], Krzysztof FUJAREWICZ[**]

THE ANALYSIS OF CHROMATIN CONDENSATION STATE AND TRANSCRIPTIONAL ACTIVITY USING DNA MICROARRAYS

The DNA microarray-based technique has been developed to semi-quantitatively measure the in vivo global chromatin condensation state at the resolution of a gene. Chromatin was fractionated due to the differential solubility of histone H1-containing and histone H1-free nucleosomes. A set of genes non-randomly distributed between histone H1-free (uncondensed or open) and histone H1-containing (condensed or closed) chromatin fractions has been identified. The transcript levels have been measured for the same group of genes. The correlation between transcriptional activity and chromatin fraction distribution of particular genes has been established.

1. INTRODUCTION

The genetic material of eukaryotic cells is packed into a nucleoprotein complex termed chromatin. This packaging of the template provides the compaction and organization for transcription, replication, repair and recombination processes. The fundamental structural unit of chromatin is the nucleosome, which comprises 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 linker histone (e.g., 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. Chromatin structure is the most important factor regulating expression of the genetic material. Nucleosomes affect the action of transcription factors and transcription elongation, and mechanisms that activate transcription due to the modulation of the structure of nucleosomes (non-covalent remodeling and covalent modification of histones) are well known. Furthermore, the formation of specific chromatin structures leads to transcriptional repression of chromatin domains. It is generally believed that the open or uncondensed chromatin state, which allows access of transcription factors and RNA polymerases to the template, 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 condensed (or closed) chromatin (so called heterochromatin). However, knowledge about the actual state of chromatin condensation and its relationship to transcriptional activity is limited to a small number of genes studied in a few model organisms [1].

[*] Department of Experimental and Clinical Radiobiology, Center of Oncology, Gliwice, Poland; widlak@io.gliwice.pl

[**] Department of Automatic Control, Institute of Automation, Silesian Technical University, Gliwice, Poland; kfujarewicz@ia.polsl.gliwice.pl

Earlier techniques have allowed the analyses of DNA methylation and chromatin condensation states, important epigenetic factors involved in regulation of transcription, either at the level of the whole genome or for a few individual sub-kilo base regions. Currently, the way to survey global chromatin condensation or methylation states with a gene to gene resolution is available due to advances in a DNA microarray technology. Functional genomics using cDNA microarray technology is a powerful technique that allows quantitative analysis of global gene expression. This technique is based on hybridization of total cDNA preparred from cellular mRNA with specific DNA (or oligonucleotide) probes spotted (or synthesized) on silica microplates [2]. However, this technique has several limitations, one of them being the lack of ability to detect rare transcripts (a few mRNA molecules per cell). It is known that among such rare transcripts are those coding for essential transcription factors or signaling molecules. Transcription factors and signaling molecules are frequently expressed in a cell cycle-dependent mode, which add more problems when such genes are studied in populations of unsynchronized cells.

Here we aimed to use a DNA array-based technique to semi-quantitatively survey the in vivo global chromatin condensation state at the resolution of single genes. First, we have separated nuclear material into uncondensed (open) and condensed (closed) chromatin fractions and established the relative enrichment of particular genes in each fraction by hybridization of DNA purified from these fractions to DNA microarrays. Second, we checked for a correlation between the "enrichment" of particular genes in the uncondensed chromatin fraction and their level of transcription. We expected that genes with high transcription rates would be preferentially located in the uncondensed chromatin fraction. This fraction may also be enriched in potentially active genes. Such a correlation, if it exists, would allow one to establish the transcriptional activity state of a given gene based on its distribution among chromatin fractions, which may be particularly useful in the case of genes expressed at low levels or in a cell cycle-dependent mode.

2. RESULTS AND DISCUSSION

Several techniques have been described that allow one to experimentally differentiate between uncondensed and condensed chromatin states. Among these are the susceptibility of specific sequences to cleavage by nucleases (uncondensed chromatin being generally more accessible) and the presence of histone H1 (more abundant in condensed chromatin) [3]. Here, we have taken advantage the chromatin fractionation technique described by Garrard and coworkers [4], based on differential solubility of histone H1-containing and histone H1-free chromatin fragments. Briefly, isolated nuclei are incubated with micrococcal nuclease (MNase), which specifically cleaves internucleosomal linker DNA, at physiological ionic strength (100 mM KCl and 3 mM $MgCl_2$). This treatment solubilizes 10-20 % of the chromatin, which is collected as the first supernatant fraction termed S1. The remnant nuclei are then incubated in the low ionic strength buffer (2 mM EDTA), which solubilizes an additional 50-60% of the chromatin (collected as the second supernatant fraction termed S2). The S1 fraction lacks histone H1 while the S2 fraction consists of histone H1 in addition to core particles. When the mouse plasmacytoma MPC-11 cell line was used as a research model, heavily transcribed immunoglobulin genes were enriched in the S1 fraction while the non-transcribed beta-globin gene was enriched in the S2 fraction [4]. We have chosen the same experimental model for our study. Nuclei isolated from MPC-11 cells were digested with

MNase and then the S1 and the S2 fractions were prepared as described above. The protein and DNA composition was analyzed electrophoretically in both fractions [Figure 1]. As expected, the S1 fraction contained primarily mononucleosomes lacking histone H1. In the contrast, the S2 consisted of histone H1 containing oligonucleosomal particles.

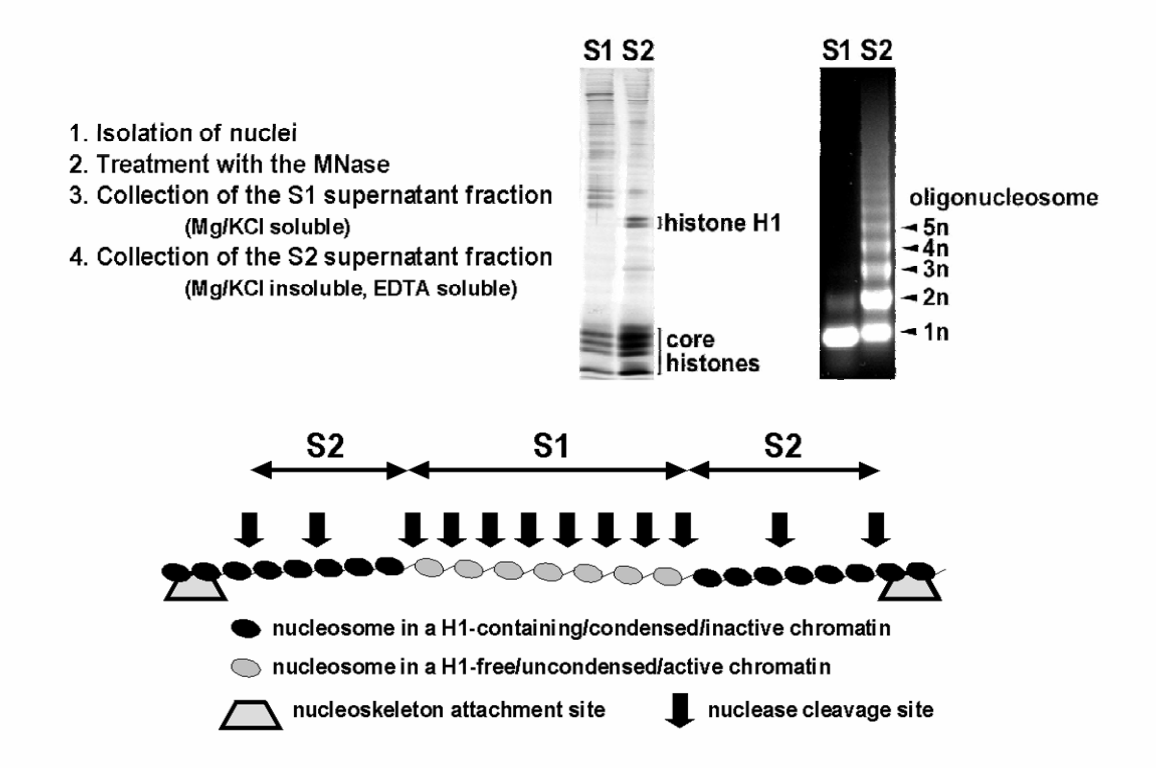

Fig.1. Chromatin fractionation; experimental procedure and analyses of protein and DNA content in the S1 and S2 chromatin fractions.

DNA was then purified from the S1 and the S2 fractions and fluorescence-labeled by incorporation of Cy5 using the standard DNA random primer procedure (Amersham Pharmacia). Total RNA was also purified from the portion MPC-11 cell using Trizol extraction and labeled by incorporation of Cy3 using the Cyscribe kit (Amersham). The mixtures of labeled DNA (from the S1 or S2 fractions) and the same RNA preparation (10μg each) were hybridized with mouse cDNA microarrays (about 22,000 spotted genes; produced by the microarray core at the University of Texas Southwestern Medical Center, Dallas) according to the protocol published by Pollack [5]. The microarrays were scanned and the levels of RNA and DNA that hybridized to each spotted gene were analyzed using MArcC-V Microarray Calculation and Visualization v. 2.91 software. Both Cy5 (scanned at 635nm; representing DNA) and Cy3 (scanned at 532nm; representing RNA) values were normalized by subtraction of background signals (mean pixel intensities) and the DNA/RNA proportion was presented as a normalized log ratio for each gene separately. The experiment with the S1 fraction DNA was repeated with reversed label combinations to exclude dye dependent variability. Of the 22,000 probes on the microarrays 2,685 contained detectably transcribed genes of MPC-11 cells and were characterized by sufficient signal/noise ratios in all three experiments.

For each of such 2,685 expressed genes the relative enrichment/depletion in the S1/S2 chromatin fractions have been established. To do so, we have correlated DNA/RNA log ratios between two repeat experiments with the S1 fraction and between experiments with the S1 and S2 fractions. In the case of the repeat experiments with the S1 fraction, the expected correlation between experiments was 1.0. In the ideal situation, dots representing each of the 2,685 analyzed genes would lay along the diagonal line in Figure 2. Data presented in Figure 2 show the positions of experimental points (genes). The calculated inter-experimental correlation coefficient between genes was 0.7897.

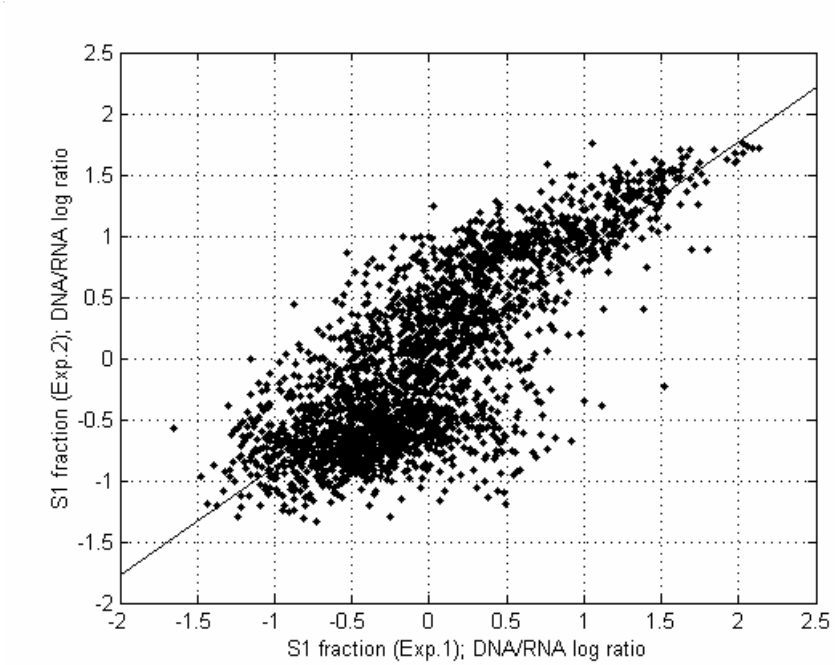

Fig.2. The inter-experimental correlation between two experiments with the S1 fraction.

A similar analysis that correlates the DNA/RNA log ratios between experiments with the S1 and the S2 fraction has been also done [Figure 3]. The location of experimental points (genes) and the slope of the linear regression curve represents the similarity between these fractions. We expected that the correlation between these two experiments would be significantly different from 1.0. This difference would indicate dissimilarity between the S1 and S2 chromatin fractions. Figure 3 shows the location of experimental points in such an analysis. The calculated inter-experimental correlation coefficient between genes was 0.5382 when fractions S1 and S2 were compared. The data indicate the relative similarity of the 2,685 analyzed genes with respect to their distribution between the S1 and S2 fractions. In other words, only a small portion of the analyzed genes show relative enrichment or depletion in the S1 compared to the S2 chromatin fraction.

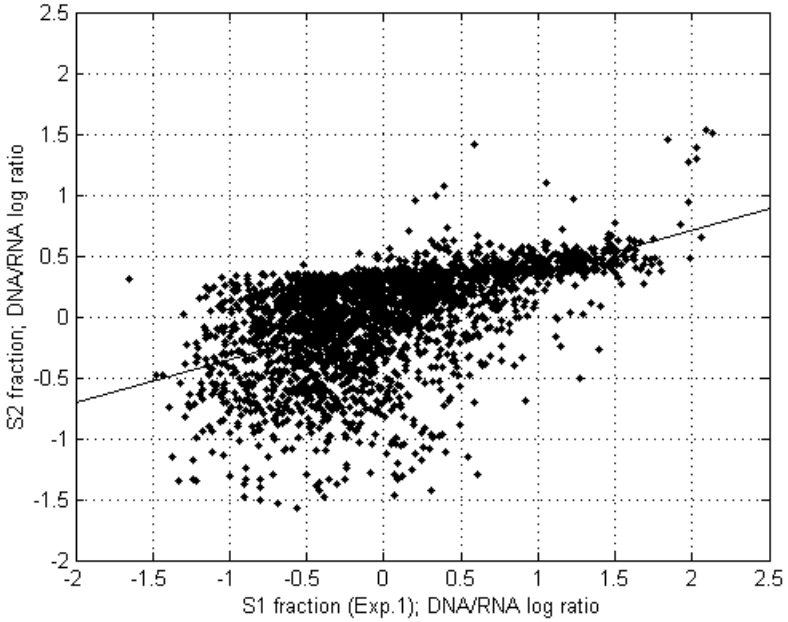

Fig.3. The inter-experimental correlation between experiments with the S1 and S2 fractions.

As the second step of our analyses, we have searched for particular genes whose distribution between chromatin fractions was the most similar and the most dissimilar. Based on signal levels obtained for DNA in the experiments that analyzed the S1 and the S2 chromatin fractions we calculated the S1/S2 dissimilarity value D:

$$D = \log_{10} \frac{DNA_{S1}}{RNA} - \log_{10} \frac{DNA_{S2}}{RNA} \qquad (1)$$

where DNA_{S1}/ RNA values are averaged signals obtained in two experiments. The calculated D value was correlated with the transcript level T for each gene (transcript level of a gene was expressed as the proportion a normalized value of RNA signal to the mean value for all 2,685 genes and averaged for all three experiments). The D values were plotted against T values for each of the analyzed genes (Figure 4). A correlation between the D and T values has been found: an increase of the D value corresponded to an increase of the T value. The correlation coefficient was 0.2719 (or 0.0962, when D was correlated with $\log_{10}$ of T). A similar analysis was performed when genes with particularly low transcript levels (T<0.3) were eliminated. In this case, the correlation coefficient increased to 0.3978 (or 0.4679, when D was correlated with the $\log_{10}$ of T). The data indicate that the expression levels of genes enriched in the S1 fraction (high D values) are higher then those enriched in the S2 fraction (low D values).

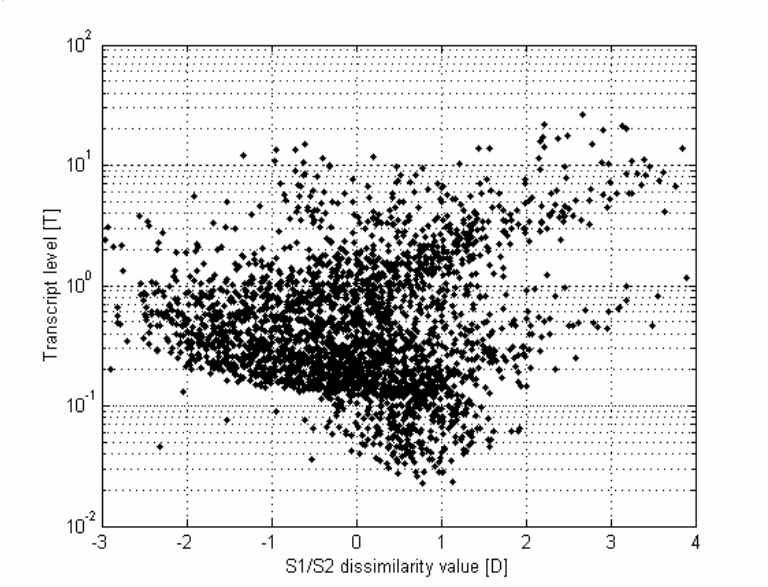

Fig.4. The correlation between the gene distribution among the S1 and S2 chromatin fractions [D] and the level of transcripts [T].

In further analyses we have looked for three classes of genes: (i) equally distributed between fractions (D equal or closest to 0); (ii) over-represented in the S1 fraction (maximal D); (iii) under-represented in the S1 fraction (minimal D). We have selected 363 genes with the most similar distributions between fractions (D in the range –0.2±0.2), 100 genes over-represented (D above 2) and 94 genes under-represented (D below –2). Figure 5 shows the correlation between D and T values for these selected groups of genes. The identification and biological characterization of genes from such isolated classes will be performed in the future.

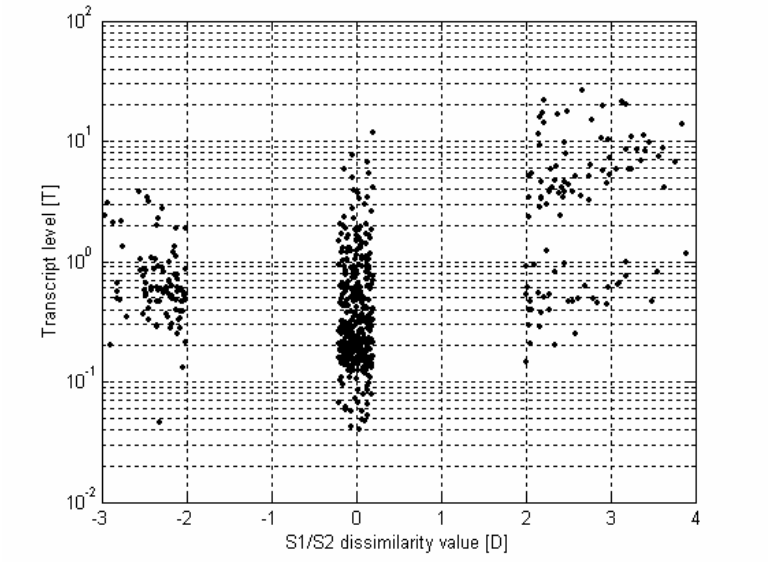

Fig.5. The correlation between the gene distribution among the S1 and S2 chromatin fractions [D] and the level of transcripts [T]. Shown are genes over-represented (left) or underrepresented in the S1 (right), or equally distributed between the S1 and S2 chromatin fractions (middle).

In this work we have identified genes that are equally or non-randomly distributed between chromatin fractions differing in their condensation state. At the onset we had speculated that the nuclear fraction containing the histone H1-deficient, less condensed chromatin would be enriched in highly-transcribed genes. In fact, we have found such a correlation. However, the established correlation was weaker than we expected. This may indicate that potentially active genes (with an open chromatin state but not currently transcribed) dominate in the S1 chromatin fraction over currently transcribed ones. On the other hand, the quality of experimental data used for the analyses was very low (only about 2,600 out of 22,000 genes were suitable for analyses) that possibly affected these final results.

BIBLIOGRAPHY

[1] WIDLAK P., GARRARD WT., Nucleosomes and regulation of gene expression. Structure of the HIV-1 5'LTR. Acta Biochimica Polonica, vol. 45, pp. 209-219, 1998.
[2] CHURCHILL GA., Fundamentals of experimental design for cDNA microarrays. Nature Genetics, supplement, vol. 32, pp. 490-495, 2002.
[3] GARRARD WT., Histone H1 and the conformation of transcriptionally active chromatin. Bioessays, vol. 13, pp. 87-88, 1991.
[4] HUANG SY., GARRARD WT., Electrophoretic analyses of nucleosomes and other protein-DNA complexes. Methods in Enzymology, vol. 170, pp. 116-142, 1989.
[5] POLLACK JR., Parallel analysis of gene copy number and expression using cDNA microarrays. Methods in Molecular Biology, vol. 224, pp. 89-97, 2003.

This work was supported by Committee for Scientific Research (KBN) Grants 4T11F01824 and PBZ-KBN-040/P04/2001. The chromatin fractionation and microarray hybridization were conducted at UT Southwestern Medical Center in Dallas, TX in the laboratory of William Garrard under the support of NIH grants GM29935 and GM59809.