(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 November 2003 (20.11.2003)

PCT

(10) International Publication Number
WO 03/095608 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/15032

(22) International Filing Date: 13 May 2002 (13.05.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/290,036 11 May 2001 (11.05.2001) US

(71) Applicant *(for all designated States except US)*: **STAMATOYANNOPOULOS, John, A.** [US/US]; 15 Milford Street, #1, Boston, MA 02118 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **MCARTHUR, Michael** [GB/GB]; Mill House, Mill Lane, Rocklands All Saints NR17 1XR (GB).

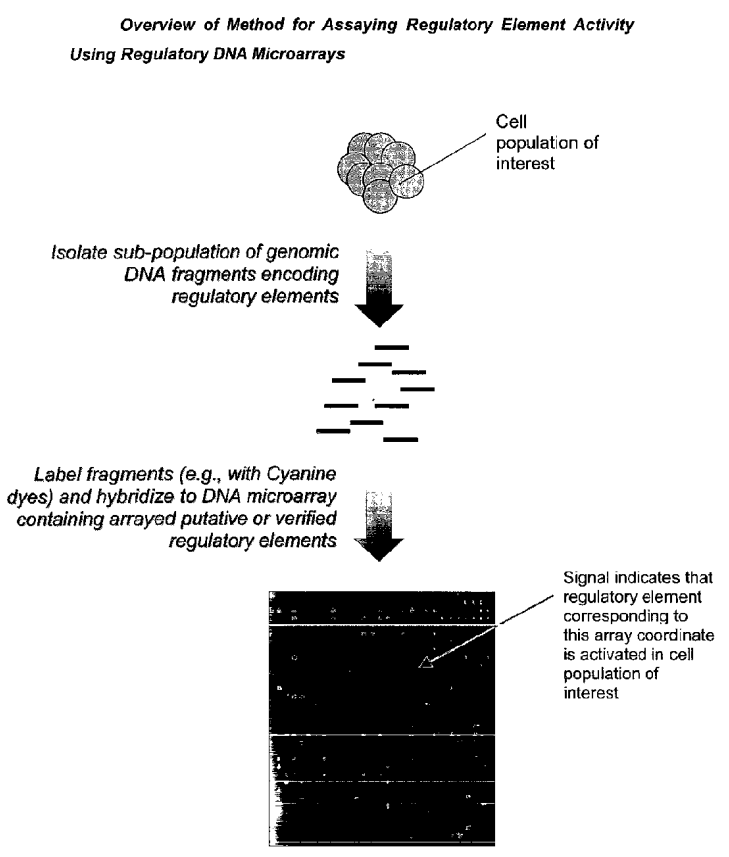
(74) Agents: **ANTLER, Adriane, M.** et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036-2711 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

*[Continued on next page]*

(54) Title: DNA MICROARRAYS COMPRISING ACTIVE CHROMATIN ELEMENTS AND COMPREHENSIVE PROFILING THEREWITH




(57) Abstract: Methods are disclosed for the construction of and interrogation of DNA arrays containing Active Chromatin Elements, and, thereby active genetic regulatory sequences.. Further methods are disclosed for interrogation of such arrays in order to reveal the pattern of genetic regulatory activity within any given cell or tissue type under a variety of conditions.

WO 03/095608 A2

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# DNA Microarrays Comprising Active Chromatin Elements and Comprehensive Profiling Therewith

## 1. Field of the Invention

The invention relates to DNA arrays for simultaneous detection of multiple nucleic acid sequences, their manufacture and use. The invention further concerns array methods and devices for detecting patterns of active chromatin elements, and particularly genetic control elements active in eukaryotic cells.

## 2. Background of the Invention

Conventional gene expression studies generally employ immobilized DNA molecules that are complementary to gene transcripts (either the entire transcript or to selected regions thereof) that are transcribed and spliced into mRNA. Recent advances in this field utilize arrays or microarrays of such molecules that enable simultaneous monitoring of multiple distinct transcripts (see, e.g., Schena et al., Science 270:467-470 (1995); Lockhart et al., Nature Biotechnology 14:1675-1680 (1996); Blanchard et al., Nature Biotechnology 14, 1649 (1996); and U.S. Pat. No. 5,569,588, issued Oct. 29, 1996 to Ashby et al. entitled "Methods for Drug Screening."). Such arrays have the potential to detect transcripts from virtually all actively transcribed regions of a cell or cell population, provided the availability of an organism's complete genomic sequence, or at least a sequence or library comprising all of its gene transcripts. In the case of the Human where a complete gene set remains unclear, such arrays may be employed to monitor simultaneously large numbers of expressed genes within a given cell population.

The simultaneous monitoring technologies particularly relate to identifying genes implicated in disease and in identifying drug targets (see, e.g., U.S. Patent Nos. 6,165,709; 6,218,122; 5,811,231; 6,203,987; and 5,569,588 all incorporated herein by reference for all purposes). Unfortunately, these array technologies generally rely on direct detection of

expressed genes and therefore reveal only indirectly the activity of genetic regulatory pathways that control gene expression itself. On the other hand, a detection system directed toward sensing the activity of particular genetic regulatory pathways or cis-acting regulatory elements could provide deeper information concerning a cell's regulatory state. Accordingly, the detection of active regulatory elements, particularly in related and interacting groups, potentially could become extremely important for delineation of regulatory pathways, and provide critical knowledge for design and discovery of disease diagnostics and therapeutics.

Most research in the area of gene regulation has focused on finding and using individual sequences either upstream or downstream of individual coding gene targets. Generally, the presence of absence of a particular DNA sequence is linked with increased or decreased expression of a nearby gene when determining the regulatory effect of the sequence. For example, the beta-like globin gene was shown to contain four major DNAase I hypersensitive sites of possible regulatory function by studies that removed or added these sequences and that looked for an effect on gene expression in erythroid cells. See Grosveld et. al. U.S. No. 5,532,143. From related studies Townes et al. asserted that two of the four DNAse hypersensitive sites might control genes generally in cells of erythroid lineage. Although an interesting development, these observations generally are limited to detection of effects on nearby coding sequences of known genes. Multiple regulatory units, which behave coordinately, are not readily amenable to analysis by these techniques.

Multiple gene and protein elements interact for even simple biological processes. Because of this, a one at a time strategy for targeting a single coding gene and nearby non-coding sequences to determine their effects on the preselected gene insufficiently addresses the true *in vivo* situation. Accordingly, any tool that can provide simultaneous regulation system

information would give rich benefits in terms of improved diagnosis, clinical treatment and drug discovery.

### 3. Summary of the Invention

The present invention overcomes the problems and disadvantages associated with current strategies and designs with methods and materials that enable the use of nucleic acid arrays for profiling large numbers of active chromatin elements ('ACE'), and hence active genetic regulatory units.

One embodiment of the invention is directed to methods for manufacturing an array of genomic regulatory elements. Since virtually all active genomic regulatory regions are contained within ACEs, an array of ACEs constitutes an array of regulatory elements. Generally, a nucleic acid microarray is made having spots that contain copies of sequences corresponding to a genomic DNA sequence that encodes an ACE or a putative genomic regulatory element. The nucleic acid sequences are obtained by amplifying sequences from a library using the polymerase chain reaction, and depositing material with a microarraying apparatus, or synthesizing *ex situ* using an oligonucleotide synthesis device, and subsequently depositing using a microarraying apparatus, or synthesizing *in situ* on the microarray using a method such as piezoelectric deposition of nucleotides.

Another embodiment of the invention is directed to methods for analyzing ACEs comprising: preparing chromatin from a target cell population; treating said chromatin with an agent that induces modifications at hypersensitive sites in chromatin such as a non-specific restriction endonuclease to induce single and double stranded cleavage at such locations in marked preference to other locations within the genome; modifying the fragment ends through the ligation of a linker adapter or similar means to tag the sequences in a manner such that they can be separated from the mixture; modifying the fragments to reduce the average fragment

size by digest with a restriction enzyme or by sonication or an equivalent procedure; labeling the fragment subpopulation containing hypersensitive site sequences with a fluorescent dye or other marker sufficient for detection through an automated apparatus such as a DNA microarray reader; and incubating the labeled fragment population with the microarray and recording the signal intensity at each array coordinate.

Yet another embodiment of the invention is a procedure for profiling ACEs from an organism, comprising a first step of constructing a DNA microarray that contains genomic regulatory elements, and a second step of probing the microarray to assay regulatory element activation. The first step involves constructing a DNA microarray having spots with one or more copies of a DNA sequence corresponding to a genomic DNA sequence that encodes a nuclease hypersensitive site or a putative genomic regulatory element. The DNA sequences may be obtained or deposited alternative ways: by amplifying the DNA sequence using PCR from a library containing such sequences, and subsequently depositing with a microarraying apparatus; synthesizing the DNA sequence *ex situ* with an oligonucleotide synthesis device, and subsequently depositing with a microarraying apparatus; or by synthesizing the DNA sequence *in situ* on the microarray by, for example, piezoelectric deposition of nucleotides. The number of sequences deposited on the array may vary between 10 and several million depending on the technology employed to create the array.

In another embodiment of the invention a DNA microarray containing genomic DNA sequences corresponding to established or putative regulatory elements is assayed in five steps. In step one, chromatin from a target cell population is prepared and treated with an agent that induces modifications at ACEs. For example, the non-specific restriction endonuclease DNAse may be used to induce single and double stranded cleavage at such locations in marked preference to other locations within the genome. Secondly, the fragment ends are modified through the ligation of a linker

adapter, enzymatic labeling or similar means to tag the sequences in a manner such that they can be separated from the mixture. Thirdly, the DNA fragments may be modified further to reduce the average fragment size by digest with a restriction enzyme, by sonication or an equivalent procedure. Fourthly, the DNA fragment subpopulation containing hypersensitive site sequences is labeled with a fluorescent dye or other marker sufficient for detection through an automated apparatus such as a DNA microarray reader. A last step is incubation of the labeled fragment population with the DNA microarray and recording the signal intensity at each array coordinate.

Other embodiments and advantages of the invention are set forth in part in the description which follows, and in part, will be obvious from this description, or may be learned from practice of the invention.

## 4. Description of the Figures

Figure 1 is an overview of an embodiment for assaying ACE activity using ACE DNA microarrays.

Figure 2 indicates profiling results for ACE activity using a two-dye system to increase signal-to-noise ratio.

Figure 3 depicts profiling of differential ACE representation.

Figure 4 shows the use of ACE arrays to screen drugs and/or small molecule compounds.

Figure 5 indicates a correlation of ACEs with gene expression obtained by an embodiment of the invention.

Figure 6 shows the use of an embodiment for controlling quality of conventional expression arrays.

## 5. Detailed Description of the Invention

Nuclease hypersensitive sites from chromatin lack protein coding sequences and generally lack highly repetitive sequences. These sequences are putative regulatory sites and as such are part of the set of regulatory elements that suffice to control the entire programme of the genome within a cell (hereinafter termed "ACE").

An Active Chromatin Element ('ACE') may be defined as a genomic DNA locale which, in the context of nuclear chromatin, serves as a template for the binding of one or more proteins or protein complexes sufficient to produce a focal alteration in the nucleosomal structure. Such ACEs typically, but not exclusively, range from between 16 base pairs to 200 base pairs to up to 1500 base pairs in extent (e.g., J Biol Chem 2001 Jul 20;276(29):26883-92).

An ACE at a particular genomic locale may be revealed through its differential sensitivity ('hypersensitivity') to the action of DNA modifying agents such as for example the non-specific endonuclease DNAse (e.g., EMBO J 1995 Jan 3;14(1):106-16). However, whereas all DNAse Hypersensitive Sites are, by definition, ACEs, not all ACEs may be detected through a DNAse Hypersensitivity assay. ACEs may also be revealed through methods which rely on the detection of epigenetic modifications in chromatin such as histone acetylation and cytosine methylation. Treatments which may exert selective effects at ACEs include one or more of the following DNA-modifying agents: nucleases (both sequence-specific and non-specific); topoisomerases; methylases; acetylases; chemicals; pharmaceuticals (e.g., chemotherapy agents); radiation; physical shearing; nutrient deprivation (e.g., folate deprication), etc..

An alternative approach is to modify the proteins that bind to a given ACE (or set of ACEs) so they induce DNA modification such as strand breakage. Proteins can either be modified by many means, such as incorporation of $^{125}I$, the radioactive decay of which would cause strand breakage (e.g., Acta Oncol. 39: 681-685 (2000)), or modifying cross-linking reagents such as 4-azidophenacylbromide (e.g., Proc. Natl. Acad. Sci. USA 89: 10287-10291) which form a cross-link with DNA on exposure to UV-light. Such protein-DNA cross-links can subsequently be converted to a double-stranded DNA break by treatment with piperidine.

Yet another approach relies on antibodies raised against specific proteins bound at one or more ACEs, such as transcription factors or architectural chromatin proteins, and used to isolate the DNA from the nucleoprotein complexes associated with ACEs *in vivo*. An example of a currently used technique cross-links proteins and DNA within the eukaryotic genome following treatment with formaldehyde. After isolation of the chromatin and following either sonication or digestion with nucleases the sequences of interest are immunoprecipitated (Orlando *et al.* Methods 11: 205-214 (1997)).

Alterations to the epigenetic pattern are also known to correlate with alterations with the activity of the ACEs. One of the most closely studied types of modification is cytosine methylation. The global pattern of methylation is relatively stable but certain genes become methylated if they are silenced or conversely demethylated if activated. Differential methylation can be detected by use of pairs of restriction endonucleases that cut the same site differently according to whether or not it is methylated (Tompa *et al.* Curr. Biol. 12: 65-68 (2002)). Alternatively it is possible to generically distinguish between a methylated and non-methylated cytosine by genomic sequencing (a methodology developed by Pfeifer *et al.* Science 246: 810-813 (1989)) that converts cytosine to uracil, which behaves similarly to thymine in sequencing reactions, and leaves methyl-cytosine unmodified.

This material can be used as a template in PCR with primers sensitive to the C to U transition. Alternatively the potential mismatch (G:U) between oligonucleotide and template can be cleaved by *E.coli* Mismatch Uracil DNA Glycosylase, and that fragment removed from the population.

Additionally the enzymatic machinery which gives rise to or maintains the epigenetic patterns can also be labeled as described above so that it can be induced to cause detectable DNA modifications such as double stranded DNA breaks. Target proteins for this kind of approach would include the recently described HATs (Histone-Acetyl Transferases), HDACs (Distone De-Acetylase Complexes) whose effect on transcriptional induction has been recently described (Cell 108: 475-487 (2002)), as well as DNA methyltransferases and structural proteins that bind to the sites of methylation, such as MeCP1 and MeCP2. Histones, and transcription factors are also known to become methylated, phosphorylated and ubiquinated. A range of covalent modifications, some of which have yet to be described, may be made to the structural and enzymatic machinery of transcription, replication and recombination. Current understanding indicates that such modifications have a regulatory role and it has been demonstrated that these modifications can be positively and negatively correlated with the functional activity of the underlying sequence (Science 293: 1150-1155). The potential for combinations of modifications of the ACEs overlays another layer of complexity of regulation on the underlying genome and it is possible to dynamically follow these epigenetic changes with the immunoprecipitation of the DNA sequences from *in vivo* nucleoprotein complexes.

ACEs define certain features of the nuclear architecture which plays a large role in regulation of genomic processes. Increasingly the molecules, including proteins and RNAs, which control the structure of the nucleus are being identified, and these are also used as targets to identify ACEs.

Moreover, cytologically distinct region of interphase nuclei have been described such as the nucleoli which contain the heavily transcribed rRNA genes (Proc. Natl. Acad. Sci. USA 69: 3394-3398 (1972)) and active genes may be preferentially associated with clusters of interchromatin granules (J. Cell Biol. 131: 1635-1647 (1995)). Specific regulatory regions may become localized to distinct areas within the nucleus on transcriptional induction (Proc. Natl. Acad. Sci. USA 98: 12120-12125 (2001)). By contrast specific areas of eukaryotic nuclei have been shown to be transcriptionally inert (Nature 381: 529-531 (1996)) and associated with heterochromatin. Fractionation of the nucleus on the basis of such and similar physical properties can be used to capture sets of ACEs implicated in these processes.

The number and location of ACEs differs between and among cell types, as may the number and identity of the proteins that bind to the genomic locale to create a given ACE. Certain ACEs may be specific to a particular tissue cell type or to a restricted set of tissue or cell types ('Tissue-specific ACEs'). Another set may form in co-ordination with the cell cycle or due to environmental stimuli. Other ACEs may be present in all tissue or cell types ('Constitutive ACEs') (e.g., Mol Cell Biol 1999 May;19(5):3714-26).

The total number of potential ACEs within a given cell depends largely on the cell type and state, but is generally equal to at least the number of active genes within that cell, and may be many times that number as active genes may be surrounded by (or contain within, e.g., their introns) more than one ACE. ACEs may function alone or in combination with other ACEs to modulate the expression of a cis-linked gene (e.g., Mol Cell Biol 1999 Nov;19(11):7600-9), or even a receptive gene *in trans*.

The superset of ACEs is expected to contain within it active units from virtually all known classes of genetic regulatory elements including

promoters, enhancers, silencers, locus control regions, domain boundary elements, and other elements having chromatin remodeling activities. Each of the aforementioned units may in turn be comprised of one or more ACEs (e.g., Trends Genet 1999 Oct;15(10):403-8). In addition other processes may be controlled by a subset of the ACEs or interactions between them. These include, but may not be limited to, DNA replication, recombination and the structure of the genomic DNA within the nucleus such as regions of specialized chromatin structure and three-dimensional topology of the chromatin fibre. As such, the complete set of ACEs across all cells and tissue types will contain substantially all of the regulatory elements necessary to define the transcriptional program of the genome, in any state of differentiation or in response to any stimulus.

The inventors synthesized primers associated with assembled sets of such sites and discovered that the primers were useful for either preparing libraries of sequences or directly detecting ACEs from other cell samples. That is, a library of ACE sequences or sequence locations generated with arrays of the invention provides rich and highly valuable information concerning the gene regulatory state of the cells from which the chromatin had been isolated. Further, two or more arrays or profiles (information obtained from use of an array) of such sequences are useful tools for comparing a sample set of hypersensitive sites with a reference, such as another sample, synthesized set, or stored calibrator. In using an array, individual nucleic acid members typically are immobilized at separate locations and allowed to react for binding reactions.

In many embodiments made possible from this discovery, genomic regulatory information is extracted from a biological sample without foreknowledge of genetic locus or marker information. That is, exemplified methods can identify *en mass*, hypersensitive sites for which no genetic marker has been identified previously. After identification, DNA containing sequences of the hypersensitive sites may be used as probes to identify

complementary genomic DNA sequences to find proteins and protein complexes having regulatory activity, and to discover pharmaceutical drug activities for compounds that can influence one or multiple regulatory systems. In addition, knowledge of these sequences allow the mapping and detection of naturally occurring mutations in the genome which are implicated in causing, potentially pathogenic, changes to the transcriptional programme of the cell, such as single nucleotide polymorphisms (SNPs). In many embodiments the sequences are grouped into libraries, which can be converted or abstracted into arrays to probe multiple regulatory systems simultaneously.

A library (or array, when referring to physically separated nucleic acids corresponding to at least some sequences in a library) of ACEs has very desirable properties as further detailed below. These properties can be associated with specific cell types and cell conditions, and may be characterized as regulatory profiles. A profile, as termed here refers to a set of members that provides regulatory information of the cell from which the ACEs are obtained. A profile in many instances comprises a series of spots on an array made from deposited ACE sequences from ACEs. Without wishing to be bound by any one theory of this embodiment of the invention, it is believed that a eukaryotic cell such as a Human cell contains many potential ACEs and that only a portion of the ACE potential regulatory elements are formed at any given time. By sampling and profiling the ACEs an array presents a snapshot of the cell's regulatory status.

An array profile of a cell's regulatory status typically concerns at least 10, more preferably at least 100, 250, 500, 1000, 2000, 5,000 and even more than 10,000 ACEs in some cases. Profile information from a test sample may be more or less detailed depending on the number of ACEs required to distinguish the profile from others. For example, a profile designed to examine the presence of a particular chromosomal breakage, crosslinkage or other defect may need to detect only 2 - 3, 2-10, 3-5, 10-20

or other small number of ACEs. With present techniques, the activation state (defined by an ability to form a nuclease hypersensitive site in chromatin) of only one or a very limited number of such sequence elements may be detected in an single experiment, such as a southern blot analysis.

A characteristic profile generally is prepared by use of an array. An array profile may be compared with one or more other array profiles or other reference profiles. The comparative results can provide rich information pertaining to disease states, developmental state, susceptibility to drug therapy, homeostasis, and other information about the sampled cell population. This information can reveal cell type information, morphology, nutrition, cell age, genetic defects, propensity to particular malignancies and other information. Accordingly, particularly desirable embodiments were explored that use arrays for creating ACE libraries, as detailed below.

***Libraries that Contain Descriptive Information of Cell Populations***

The simultaneous detection of multiple hypersensitive sites using arrays provides a wide range of methods for a variety of advantages. In some embodiments an array contains one or more internal references and the data profile is used directly without further comparison with reference data. In other embodiments a library of sites (either sequences, position locations or both) is obtained from a sample and then compared with another library, such as a pre-existing "type" library. A type library may be characteristic for a cell type, a development status type, a disease type such as a genetic disease, or a morphologic type associated with the presence of factor(s) such as hormones, nutrients, pharmacologically active compounds and the like. The comparison to a type library may generate an output set of difference "profile information" for the library.

The term "library" as used here means a set of at least 10, preferably 50, 100, 200, 300, 500, 1000, 2000, 5000, 10,000, 20,000 30,0000 or even at least 50,000 members of nucleic acids having characteristic sequences. The library may be an information library that contains a) ACE DNA sequences, b) location information for ACEs in the genome; or c) both sequence information and matching location information. As an information library, the members preferably are stored in a computer storage medium as sequences and/or gene position locations. As a physical DNA library, the members may exist as a set of nucleic acids, clones, phages, cells or other physical manifestations of DNA in a form useful for simultaneous manipulation.

A library of nucleic acid molecules conveniently may be maintained as separate cloned vectors in host cells. Preferably each member is physically isolated from the other members, although a mixture of members within a common vessel may be suitable, particularly for assays wherein members become separated based on a physical property such as by hybridization with specific members on a solid support.

An ACE library member in most instances comprises a sequence at least 16 bases long and less than 1500 bases long. More preferably the sequence comprises between 60 bases and 400 bases. Yet more preferably the sequence comprises between 75 bases and 300 bases. The term "mean sequence length of the hypersensitive DNA sequences" means the numeric average of all DNA sequences in the respective library or array. Experimental results indicate that most ACEs are about 50 to 400 bases long and more generally about 150 to 300 bases long. Methods for replicating DNA (or RNA) sequences and maintaining copies of those sequences in libraries are well known and have been used for some years. See for example the procedures described in U.S. Nos. 4,987,073; 5,763,239; 5,427,908; 5,853,991.

***ACE Profiling and Reference Libraries***

In preferred embodiments of the invention a set of at least 10 hypersensitive sequences and/or locations obtained from a sample are combined to form a profile of the sample. Typically an array is made that can detect the sequences and generate a data profile indicating at least a) the presence or absence of each sequence or ACE site in a sample or b) the relative abundance of active (hypersensitive) sites from a sample. It was discovered that "detection" of (i.e. determination of the presence and/or relative abundance of) at least some of the hypersensitive ACEs of a sample as a group profile on an array can reveal useful characteristics of the sample. Such characteristics include, for example, whether the sample contains a DNA break that increases the risk of particular malignancies or has a highly expressed region with respect to a normal state.

In another embodiment, a sample is processed to determine ACE usage and a profile is obtained from binding reactions between nucleic acid sequences obtained from the sample and other nucleic acid references. Advantageously either the reference nucleic acids or the sample nucleic acids are first bound in an array and the array exposed to the other set. In an embodiment at least 10, more preferably at least 100, 1000, 10,000, or even more than 20,000 reference nucleic acids are used in this embodiment.

In yet another embodiment a sample is processed to generate nucleic acids corresponding to sequences of ACEs and the nucleic acids identified by sequencing, mass spectrometry and/or another method. Profile results obtained advantageously are compared to known values.

Yet another embodiment of the invention provides a master organism reference library that substantially contains all possible ACEs of a cell. The phrase "substantially contains" in this context means at least 50% of all possible hypersensitive sites, including every site that can be found in one

situation (cell type, cell morphology, or other condition) or another. Preferably "substantially contains" refers to at least 75% of all possible hypersensitive sites, and more preferably refers to at least 90%, 95% and even at least 99% of all sequences and/or site locations. In an embodiment such library is made by mapping ACEs from at least 3 different cell types of an organism and more preferably 4, 5, 6, or even more than 10 types of different cells, and compiling all of the different ACEs into a "organism specific" set of ACEs. One version of a library includes sequences corresponding to each ACE. Yet another version of the library includes position information of each ACE. Either or both versions of data are very useful tools for diagnostic tests and other studies.

Yet another embodiment is a cell type specific reference library that "substantially contains" all ACEs of that specific type of cell. The term "substantially contains" in this context means at least 50% of all ACEs that behave as hypersensitive sites under one or more conditions experienced by that cell type. Preferably "substantially contains" refers to at least 75% of all possible hypersensitive sites, and more preferably refers to at least 90%, 95% and even at least 99% of all sequences and/or site locations. By way of example, a Human cell line was found to contain approximately 30,000 hypersensitive site ACEs, when examined in late log stage of growth.

***Generation and Use of Library Members in MicroArrays***

Many uses of the invention arise from the ability to generate, manipulate and analyze large amounts of information through libraries and their use in microarrays to provide information. Arrays generally are made and used by a variety of methods that can be discussed in terms of i) preparation of arrays; ii) sample preparation and conversion into fragment libraries, iii) manipulating the fragments by for example amplifying and cloning them, and iv) profiling libraries (i.e. either the entire set of prepared fragments or a subset of them) by detection on arrays.

### *i. Preparation of Arrays Containing ACEs*

Microassays, also called "biochips" or "arrays" are miniaturized devices typically with dimensions in the micrometer to millimeter range for performing chemical and biochemical reactions and are particularly suited for embodiments of the invention. Arrays may be constructed via microelectronic and microfabrication techniques known in the semiconductor industry and in the biochemistry industry.

Microarrays are particularly desirable for their virtues of high sample throughput and low cost for generating profiles and other data. A DNA microassay typically is constructed with spots that comprise nucleic acid with ACE sequences. In a preferred embodiment immobilized DNAs have sequences that hybridize to ACE hypersensitive sites such as putative genomic regulatory elements.

Microarrays according to embodiments of the invention may include immobilized biomolecules such as oligonucleotides, cDNA, DNA binding proteins, RNA and/or antibodies on their surfaces. Advantageous embodiments of the invention have immobilized nucleic acid on their surfaces. The nucleic acid participates in hybridization binding to nucleic acid prepared from hypersensitive sites. Such chips can be made by a number of different methodologies. For example, the light-directed chemical synthesis process developed by Affymetrix (see, U.S. Pat. Nos. 5,445,934 and 5,856,174) may be used to synthesize biomolecules on chip surfaces by combining solid-phase photochemical synthesis with photolithographic fabrication techniques. The chemical deposition approach developed by Incyte Pharmaceutical uses pre-synthesized cDNA probes for directed deposition onto chip surfaces (see, e.g., U.S. Pat. No. 5,874,554).

Other useful technology that may be employed is the contact-print method developed by Stanford University, which uses high-speed, high-

precision robot-arms to move and control a liquid-dispensing head for directed cDNA deposition and printing onto chip surfaces (see, Schena, M. et al. Science 270:467-70 (1995)). The University of Washington at Seattle has developed a single-nucleotide probe synthesis method using four piezoelectric deposition heads, which are loaded separately with four types of nucleotide molecules to achieve required deposition of nucleotides and simultaneous synthesis on chip surfaces (see, Blanchard, A. P. et al. Biosensors & Bioelectronics 11:687-90 (1996)). Hyseq, Inc. has developed passive membrane devices for sequencing genomes (see, U.S. Pat. No. 5,202,231). These methods and adaptations of them as well as others known by skilled artisans may be used for embodiments of the invention.

Arrays generally may be of two basic types, passive and active. Passive arrays utilize passive diffusion of sample molecule for chemical or biochemical reactions. Active arrays actively move or concentrate reagents by externally applied force(s). Reactions that take place in active arrays are dependant not only on simple diffusion but also on applied forces. Most available array types, e.g., oligonucleotide-based DNA chips from Affymterix and cDNA-based arrays from Incyte Pharmaceuticals, are passive. Structural similarities exist between active and passive arrays. Both array types may employ groups of different immobilized ligands or ligand molecules. The phrase "ligands or ligand molecules" refers to bio/chemical molecules with which other molecules can react. For instance, a ligand may be a single strand of DNA to which a complementary nucleic acid strand hybridizes. A ligand may be an antibody molecule to which the corresponding antigen (epitope) can bind. A ligand also may include a particle with a surface having a plurality of molecules to which other molecules may react. Preferably the reaction between ligand(s) and other molecules is monitored and quantified with one or more markers or indicator molecules such as fluorescent dyes. In preferred embodiments a matrix of ligands immobilized on the array enables the reaction and monitoring of multiple analyte molecules. For example, an array having an immobilized

library of ACE fragments may be tested for binding with one or more putative DNA binding proteins. A two dimensional array is particularly useful for generating a convenient profile that may be imaged, as exemplified in Figures 1 through 6.

More recent developments in array manufacture and use specifically are contemplated. For example, electronic arrays developed by Nanogen can manipulate and control sample biomolecules by electrical fields generated with microelectrodes, leading to significant improvement in reaction speed and detection sensitivity over passive arrays (see, U.S. Pat. Nos. 5,605,662, 5,632,957, and 5,849,486). Another active array procedure contemplated in some embodiments is the technology described in U.S. No. 6,355,491 and issued to Zhou et al. entitled "Individually addressable micro-electromagnetic unit array chips." This latter technology provides an active array wherein individually addressable (controllable) units arranged in an array generate magnetic fields. The magnetic forces manipulate magnetically modified molecules and particles and promote molecular interactions and/or reactions on the surface of the chip. After binding, the cell-magnetic particle complexes from the cell mixture are selectively removed using a magnet. (See, for example, Miltenyi, S. et al. "High gradient magnetic cell-separation with MACS." Cytometry 11:231-236 (1990)). Magnetic manipulation also is used to separate tagged ACE sequences during sample preparation in desirable embodiments, before application of DNA to a test array.

Arrays can be used to compare reference libraries as well as profiling based on as little as a single nucleotide difference. The chemistry and apparatus for carrying out such array profiling and comparisons are known. See for example the articles "Rapid determination of single base mismatch mutations in DNA hybrids by direct electric field control" by Sosnowski, R. G. et al. (Proc. Natl. Acad. Sci., USA, 94:1119-1123 (1997)) and "Large-scale identification, mapping and genotyping of single-nucleotide polymorphisms in

the Human genome" by Wang, D. G. et al. (Science, 280: 1077-1082 (1998)), which show recent techniques in using arrays for manipulation and detection of sequence alternations of DNA such as point mutations. "Accurate sequencing by hybridization for DNA diagnostics and individual genomics." by Drmanac, S. et al. (Nature Biotechnol. 16: 54-58 (1998)), "Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy" by Shoemaker, D. D. et al. (Nature Genet., 14:450-456 (1996)), and "Accessing genetic information with high density DNA arrays." by Chee, M et al., (Science, 274:610-614 (1996)) also show known array technology used for DNA sequencing.

Further examples of technology contemplated for use in making and using arrays are provided in "Genome-wide expression monitoring in Saccharomyces cerevisiae." by Wodicka, L. et al. (Nature Biotechnol. 15:1359-1367 (1997)), "Genomics and Human disease--variations on variation." by Brown, P. O. and Hartwell, L. and "Towards Arabidopsis genome analysis: monitoring expression profiles of 1400 genes using cDNA microarrays." by Ruan, Y. et al. (The Plant Journal 15:821-833 (1998)).

***ii. Interrogation of Arrays Containing ACEs (sample preparation via marking ACEs and conversion into fragment libraries).***

A first step in the generation and use of library members is to mark multiple hypersensitive sites. A site may be marked by a biochemical alteration that can be used to identify or separate the site for sequencing. This alteration often will involve breaking or making a covalent bond within specific ACEs. For example, a nuclease may mark by cutting the ACE. In a preferred embodiment non-specific nuclease such as DNAse I cuts DNA at the hypersensitive sites.

In a particularly desirable embodiment DNAse I is used to mark hypersensitive sites by cutting DNA strands at these sites. Following isolation and optional amplification of the DNA segments that flank the

hypersensitive cut sites the fragments are sub-cloned into a suitable vector as a commercially available bacterial plasmid. To effect this, the fragments are digested with restriction enzymes, cut sites of which have been engineered into the linker regions. Following incorporation into suitable bacterial plasmids, colonies are recovered which contain bacteria in which the plasmid replicates.

Other agents and methods that may be used to mark eukaryotic DNAs at hypersensitive sites include, for example, radiation such as ultraviolet radiation, chemical agents such as chemotherapeutic compounds that covalently bind to DNA or become bound after irradiation with ultraviolet radiation, other clastogens such as methyl methane sulphonate, ethyl methone sulphonate, ethyl nitrosourea, Mitomycin C, and Bleomycin, enzymes such as specific endonucleases, non-specific endonucleases, topoisomerases, topoisomerase II, single-stranded DNA-specific nucleases such as S1 or P1 nuclease, restriction endonucleases, *Eco*R1, *Sau*3a, DNase 1, StyI, methylases, histone acetylases, histone deacetylases, and any combination thereof.

As will be appreciated by skilled artisans, clastogens may be used to break DNA and the broken ends tagged and separated by a variety of techniques. Compounds that covalently attach to DNA are particularly useful as conjugated forms to other moieties that are easily removable from solution via binding reactions such as biotin with avidin. The field of antibody or antibody fragment technology has advanced such that antibody antigen binding reactions may form the basis of removing labeled, nicked or cut DNA from a hypersensitive ACE site.

In many embodiments, after forming a break or directly binding to the DNA, the affected DNA sequence around the site may be isolated and determined and/or the site mapped to a location in the genome. For example, an agent that forms a covalent bond with DNA may be conjugated

to a binding member such as biotin or a hapten. After bond formation, endonuclease may be used to generate smaller DNA fragments. Fragments that contain the marked ACE may be isolated by a specific binding reaction with a conjugate binding member (avidin or an antibody/antibody fragment respectively in this case), for example, on a solid phase that immobilizes the ACE fragments and allows removal of the other fragments.

Sample preparation begins with chromatin from cellular material. Preferably the chromatin is extracted from a eukaryotic cell population such as a population of animal cells, plant cells, virus-infected cells, immortalized cell lines, cultured primary tissues such as mouse or Human fibroblasts, stem cells, embryonic cells, diseased cells such as cancerous cells, transformed or untransformed cells, fresh primary tissues such as mouse fetal liver, or extracts or combinations thereof. Chromatin may also be obtained from natural or recombinant artificial chromosomes. For example, the chromatin may have been assembled *in vitro* using previously sub-cloned large genomic fragments or Human or yeast artificial chromosomes.

In many embodiments multiple ACE sequences and/or location sites are obtained from a eukaryotic cell sample by first extracting and purifying nuclei from the sample as for example, described in U.S. No. 09/432,576. Briefly, a sample is treated to yield preferably between about 1,000,000 to 1,000,000,000 separated cells. The cells are washed and nuclei removed, by for example NP-40 detergent treatment followed by pelleting of nuclei. An agent that preferentially reacts with genomic DNA at ACEs is added and marks the DNA, typically by cutting or binding to the DNA. In a particularly advantageous embodiment DNAse I is used to form two single strand breaks near each other, and typically within 5 bases of each other. After reaction with hypersensitive DNA sites the reacted DNA is, if not already, converted into smaller fragments and the reacted fragments optionally are amplified and separated into a library. Preferably breaks on both strands within up to

10 base pairs from each other are detected after extraction by cloning one or both sides of the site.

### *iii. Manipulation of fragments*

Isolation of DNA after marking and fragmentation may be accomplished by a number of techniques. Exemplary methods include: adaptive cloning linkers that facilitate selective incorporation into a cloning vector or PCR; streptavidin/biotin recovery systems; magnetic beads, silicated beads or gels; dioxygenin/anti-dioxygenin recovery systems; or a variety of other methods. Once isolated (or even before isolation), fragments can be labeled with a detectable label. Suitable detectable labels include fluorescent chemicals, magnetic particles, radioactive materials, and combinations thereof.

Amplification of isolated DNA fragments may be required in the event that the quantities of DNA recovered from this isolation step are insufficient to effect efficient cloning of the desired segments, or simply to produce a more efficient process.

In a desirable embodiment described in Example 1 a biotin-labeled linker is added after formation of cut ends by DNase I and binds to the cut ends. The mixture is digested with one or more restriction endonucleases such as Sau3a or StyI to create smaller fragments and the biotin labeled fragments recovered by a binding reaction to immobilized avidin followed by removal of unbound fragments. An amplification step such as polymerase chain reaction ("PCR") optionally may be performed.. To render the fragments fit for PCR, another linker can be incorporated at the opposite end from that of the biotinylated linker.

Newer variations of PCR and related DNA manipulations such as those described in U.S. Nos. 6,143,497 (Method of synthesizing diverse collections of oligomers); 6,117,679 (Methods for generating polynucleotides

having desired characteristics by iterative selection and recombination); 6,100,030 (Use of selective DNA fragment amplification products for hybridization based genetic fingerprinting, marker assisted selection, and high throughput screening); 5,945,313 (Process for controlling contamination of nucleic acid amplification reactions); 5,853,989 (Method of characterization of genomic DNA); 5,770,358 (Tagged synthetic oligomer libraries); 5,503,721 (Method for photoactivation); and 5,221,608 (Methods for rendering amplified nucleic acid subsequently un-amplifiable) are desirable. The contents of each cited patent which pertains to methods of DNA manipulation are most particularly incorporated by reference.

DNA samples thus prepared by marking and amplification may be further manipulated and applied to an array in a number of ways. For example, the DNA sequence may be amplified using the polymerase chain reaction from a library containing such sequences, and subsequently deposited using a microarraying apparatus. In another way the DNA sequence is synthesized *ex situ* using an oligonucleotide synthesis device, and subsequently deposited using a microarraying apparatus. In yet another way the DNA sequence may be synthesized *in situ* on the microarray using a method such as piezoelectric deposition of nucleotides. The number of sequences deposited on the array generally may vary upwards from a minimum of at least 10, 100, 1000, or 10, 000 to between 10,000 and several million depending on the technology employed.

A DNA fragment subpopulation containing ACE sequences advantageously may be detected by fluorescence measurements by labeling with a fluorescent dye or other marker sufficient for detection through an automated DNA microarray reader. The labeled fragment population generally is incubated with the surface of the DNA microarray onto which has been spotted different binding moieties and the signal intensity at each array coordinate is recorded. Fluorescent dyes such as Cy3 and Cy5 are particularly useful for detection, as for example, reviewed by Integrated DNA

Technologies (see "Technical Bulletin at http://www.idtdna.com/ program/techbulletins/Dark_Quenchers.asp) and as provided by Amersham (See Catalog # PA53022, PA55022 and related description).

DNA arrays that contain sequences such as those described here, their complementary sequences, or other sequences derived from them may be prepared by a wide variety of technologies, as discussed next.

***iv. Profiling Libraries on Arrays***

As described above libraries may exist in silico as DNA sequences or in vitro as physical elements that contain DNA. In other embodiments libraries are profiled on arrays. Data obtained from large assemblages of library elements are useful for many purposes. In principle, two or more arrays are prepared under similar conditions with one array acting as a control or reference for the other(s). For example, alteration of expression induced by a test compound such as a drug candidate may be determined by creating two arrays, one that corresponds to cells that have been treated with the test compound and a second that corresponds to the cells before treatment.

Differences in array data profiles can reveal which ACEs are affected by the test compound. An ACE may be more hypersensitive in the presence of the drug, as seen by more abundant hits at that ACE site during the nuclei incubation/reaction step leading to a stronger ACE signal in a profile. An ACE may be found less hypersensitive if, in comparison to a no drug control, a weaker signal were produced for that ACE spot in the array. In another example, an array profile obtained from a malignant tissue sample may be compared with an array profile obtained from a control or normal tissue sample. An inspection of the hypersensitive ACE differences between the arrays may reveal a genetic cause in the disease or a genetic factor in the disease progression.

A wide variety of diseases have genetic components and may be diagnosed by ACE profiling according to embodiments of the invention. In a preferred embodiment a test ACE profile (usually an array result) from a test sample is compared with a reference ACE profile from a healthy tissue. In another embodiment, the test ACE profile is taken before and after treatment of the cells before nuclei extraction, with one or more pharmacological compounds designed to combat one or more disease states. In yet another embodiment a library of known polymorphisms is used as a data base for comparison with sequences obtained from one or more ACEs associated with the disease state. In this way, a drug treatment regimen may be more individually tailored to the genetic regulatory profile of the patient. This latter embodiment is particularly useful for diagnosing a malignancy, particularly where a DNA breakage or transposition event has occurred. In a particularly advantageous embodiment a profile of ACEs associated with chromosomal breakage sites is determined for a patient. The profile is compared with a reference to obtain a diagnosis and to determine a possible treatment.

Many DNA breakages are associated with malignancies and occur at ACEs. Accordingly, the profiling and comparison of profiles from biopsy samples are particularly helpful for analyses of malignancies that are associated with DNA strand breakage. This profiling and analysis is particularly useful in conjunction with other known therapies, as it can provide clinically valuable information leading to selection of superior pharmacologic and (where appropriate) chemotheapeutic agents. For example, some malignancies are characterized by progressive DNA strand breakage and/or activation (up regulation) of protective systems (such as efflux of xenobiotics) against chemical therapies. An ACE profile obtained from such a sample can distinguish between regulatory systems sensitive to drug intervention of the disease from those that are not, or that have lost sensitivity to a pharmaceutical. Furthermore, combination drug therapy is conveniently optimized by the profiling of multiple regulatory systems at once.

An ACE profile may be as simple as a small set of 6, 7, 8, 10, 10 to 25, 25 to 100, or 100 to 500 ACEs. The procedures and materials illustrated in "Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays." by Cronin, M. T. et al. (Human Mutation, 7:244-255 (1996)), and "Polypyrrole DNA chip on a silicon device: Example of hepatitis C virus genotyping." by Livache, T. et al. (Anal. Biochem. 255:188-194 (1998)) are particularly contemplated for determining differences between a reference sequence or library sequence and that obtained from a sample. These documents are specifically incorporated by reference and illustrate the knowledge of skilled artisans in this field.

In another embodiment an array generates data that reveal ACE copy number. As will be readily appreciated, some ACEs are more hypersensitive than others for a given cell state and this character can be seen as a higher copy number, or (where appropriate) a greater detection signal compared to another ACE or reference sample. According to an embodiment of the invention the relative copy numbers of one or more ACEs are compared to a reference or set of references to determine a relative activity of the ACE.

Without wishing to be bound by any one theory of this embodiment of the invention it is believed that ACE profiling in this manner often yields a more accurate determination of gene regulation than measuring transcribed mRNA or a protein product of a gene because "hypersensitivity" itself is a more direct measure of whether a regulatory system is on or off. In contrast, mere quantitation of a transcription or translation product generally reflects more variables and may be less tightly associated with the biochemical operation of the corresponding regulatory unit. One embodiment of the invention is an improvement in previous diagnostic and quantitative tests for gene regulation wherein one or more ACEs and/or a ACE profile is determined by an array and correlated with a particular protein function or other biological effect.

Another embodiment of the invention is a set of primers corresponding to a library of ACEs and which can form an array. Preferably the library contains at least 10, 100, 250, 500, 1,000 5,000 or even more than 10,000 primers that correspond to specific ACEs. In an advantageous method a library of ACE specific primers are used to selectively amplify or detect ACE sequences corresponding to a particular desired profile. A library profile may be as small as a set of 5 or 10 ACE sequences. In this case 5 or 10 primers with sequences corresponding to the desired ACEs may be used with a DNA sample to selectively amplify those ACEs for further analysis.

The library profiling and comparison techniques of the invention are useful for discovery of drugs that interact with regulatory mechanisms mediated by one or more ACEs. A respective embodiment directly screens for drugs by exposing a microarray of ACE sequences to potential drugs. Another embodiment scores the effect of a chemical on an intact nucleus by exposing the nucleus to the drug and then deriving a library of ACEs from the treated nucleus. Representative techniques and materials useful in combination for this embodiment are found in "Selecting effective antisense reagents on combinatorial oligonucleotide arrays." by Milner, N. et al. (Nature Biotechnol., 15:537-541 (1997)), and "Drug target validation and identification of secondary drug target effects using DNA microarray." by Marton, M. J. et al. (Nature Medicine, 4:1293-1301 (1998)).

While many embodiments of the invention concern profiled information from arrays, the fragment libraries and derivatives of them are independently valuable tools. A fragment library prepared by marking and separating out ACEs from chromatin contains valuable information that may be extracted and used in a variety of forms. For example, the fragments can be sequenced and their profile information entered into a computer or other data base for comparison *in silico* with one or more reference libraries. The

fragments may be cloned and used for drug discovery via one or more screening techniques. Isolated fragments may be cloned by any of a number of techniques using any number of cloning vectors. Exemplary techniques include: introduction into self-replicating bacterial plasmid vectors; introduction into self-replicating bacterophage vectors; and introduction into yeast shuttle vectors.

Generally, the fragment library may be converted by an array manipulation in silico or in vitro into other valuable libraries by a variety of techniques. For example, members of the library having highly repetitive sequences may be deleted from computer memory by pattern matching and removal of matched sequences. Highly repetitive sequences and/or other undesirable sequences/sites such as those found by random breaks during DNA isolation. Such fragment libraries, either as computer data base set or as physical DNA containing sets of vessels, molecules, plasmids, cells or organisms, are valuable items of commerce. For example, a library obtained from tissue of a patient with a particular disease will represent a snapshot of the active ACE profile associated with the disease and has significant value for drug discovery and for diagnosis. Both a computer based data set library and physical embodiments of that set such as a library of clones has great utility and may be sold for a variety of purposes.

The following specific examples are provided to illustrate embodiments of the invention, and should not be viewed as limiting the scope of the invention.

## Examples

Many of the exemplified processes utilize combinations of new and old techniques and yield libraries of sub-cloned DNA fragments containing nuclease hypersensitive sites, as exemplified in Figures 1 and 2. A more specific example, as represented below and illustrated in Figure 2 is a

method that generates libraries of sub-cloned DNA fragments representing the complement of nuclease hypersensitive sites present in the chromatin of cells from erythroid cell lines.

Examples 1-3 set forth a general, but preferred, method for producing a hypersensitive site library from cultured hematopoetic cell lines. This method embodies the process illustrated in Figure 2.

**Example 1 Preparation of DNA microarrays containing ACEs**

Primer pairs were designed to allow amplification of approximately 500 bp PCR products from human genomic DNA. Following two rounds of amplification, where in the second one-hundredth volume of the original PCR reaction is used as a template, the PCR products are purified (using Millipore Multi-screen PCR purification plates), quantified (A260) and their concentration established to be between 50 ng/ l – 150ng/ul. The size of the PCR products is checked by agarose gel eletrophoresis before the microarrays are printed (in 50% DMSO) onto mirrored slides (RPK0331, Amersham) using Amersham's Lucidea Arrayer. The PCR products are crosslinked to the slides with 500mJ, using Stratagene's Stratalinker. The slides are stored desiccated until use.

**Example 2 Preparation of DNA that contains one or more single-stranded or double-stranded cleavage sites within domains defined by ACEs.**

K562 cells were grown to confluence (5 x 105 cells per cubit milliliter as assayed by hemocytometer). Nuclei were prepared from a suitable volume (e.g., 100ml) and nuclei were prepared as described (Reitman et al MCB 13:3990). Briefly, Nuclei were resuspended at a concentration of 8 OD/ml with 10 microliters of 2 U/microliter DNaseI [Sigma] at 37°C for 3 min. The DNA was purified by phenol-chloroform extractions and ethanol precipitated. The DNA was repaired in a 100 microliter reaction containing 10 microgram DNA and 6 U T4 DNA polymerase (New England Biolabs) in the manufacturer's recommended buffer and incubated for 15 min at 37°C and then 15 min at 70°C. 1.5 U Taq polymerase (Roche) was added and the incubation continued at 72°C for a further 10 min. The DNA was recovered using a Qiagen PCR Clean-up Kit and the DNA eluted in 50 microliter of 10 mM Tris.HCl, pH8.0

**Example 3 Isolation of DNA fragments associated with ACEs.**

DNA was mixed in a 100 microliter reaction volume containing 50 pmol of PS003 adapter (created by annealing equimolar amounts of oligonucleotides 5' biotinylated PS003f and 5' phosphorylated PS003r, to create an adapter containing a *NotI* site) and 40 U T4 DNA ligase (New England Biolabs) in the manufacturer's recommended buffer for 16 h at 4°C. The reaction was incubated at 65°C for 20 min before the DNA was isopropanol precipitated in the presence of 0.3 M NaOAc and after ethanol washing resuspended in 20 microliter TE buffer (10 mM Tris.HCl, 1 mM EDTA, pH8.0). The DNA was digested in a 50 microliter reaction volume containing 20 U *Hsp92* II (Promega) in the manufacturer's recommended buffer by incubation at 37°C for 2 h, afterwhich a further 20 U of enzyme was added and the incubation continued for 1 h and then heated to 72°C for 15

min. The DNA was captured on M-270 Dynal beads as per manufacturer's instructions.

The beads were finally washed in 200 microliter of ligation buffer before capture and resuspension in a 100 microliter reaction volume containing 50 pmol of Hsp adapter (made by annealing equimolar amounts of oligonucleotides fHsp and rHsp) supplemented with 6 U T4 DNA ligase (New England Biolabs) in the manufacturer's recommended buffer and incubated at 16°C for 16 h. The reaction was heated to 65°C for 15 min prior to capture of the beads. The beads were washed in 1 x NEB3 buffer (New England Biolabs) and then resuspended in a reaction volume of 100 microliter of the same buffer supplemented with 40 U *NotI* (New England Biolabs) and incubated for 37°C for 1 hour with occasional mixing. Afterwhich the beads were captured and the supernatant retained. The beads were washed once and the resultant supernatant combined with the first and isopropanol precipitated in the presence of 20 microgram glycogen and 0.3 M NaOAc. After ethanol washing the DNA was resuspended in 10 microliter of 10 mM Tris.HCl, pH8.0.

It will be clear to those skilled in the art that fragments isolated by the procedure above, or modifications thereof, may be used as reagents for the isolation or identification of genomic DNA segments that flank the site of DNA modification by combination with separately prepared population of genomic DNA that has been fragmented by other methods.

In the case of this specific embodiment/example, it is desirable to perform an amplification step prior to subcloning. It is anticipated that such a step may be required in some, but by no means all instances of the application of the process of the invention, as mentioned above. To perform amplification of the recovered DNA fragments prior to cloning, PCR may be employed or other methods of amplification, such as RCA (Rolling Circle Amplification) or versions of it. To render the fragments fit for PCR for

example, another linker can be incorporated at the opposite end from that of the biotinylated linker mentioned above. A PCR amplification is then carried out.

To confirm that the DNA segments isolated with the above procedure contain ACE regions that would be expected in an erythroid cell line such as K562, the products are probed for the presence of nuclease ACEs known to be present in this cell type.

**Example 4 Labeling of DNA fragments associated with ACEs**

Two μg of DNA were diluted into a volume of 24 μl with water and 20 μl of 2.5 x Random Primers Solution (Invitrogen, constituent of BioPrime Labeling Kit) and the mixture heated to 95°C for 5 min. The mixture is cooled on ice for 5 min before 2 ml dNTP solution (consisting of 5 mM Promega's dATP, dGTP, dTTP and 1 mM dCTP) and 3 μl of either 1 mM dCTP-Cy3 or dCTP-Cy5 (Amersham) and 1 μl of 40 U/ml Klenow (Invitrogen). The mixture was incubated at 37°C for 2.5 h before being stopped by the addition of 5 μl of 0.5 M EDTA. The probes were purified on Qiagen QIAquick columns and eluted in 100 μl of EB. The amount of incorporation was calculated by reading the absorbance at 550 nm (for Cy3) and 650 nm (for Cy5) and probes were mixed at a dye molar ratio of 4:1 (pmol Cy3: pmol Cy5). Typically 200 pmol of Cy3 labeled probe was used and 50 pmol Cy5.

**Example 5 Preparation and labeling of control DNA fragments**

Genomic DNA was isolated from K562 nuclei which had not been treated with a nuclease (1 ml of nuclei with an $A_{260}$ of 8 OD/ml) and had been subsequently digested with *Nla*III to completion and the DNA purified using a Qiagen Dneasy column. The concentration of the DNA was corrected to 150 ng/μl. These probes were labeled with Cy3..

**Example 6 Hybridization of ACE-associated and control DNA fragments to ACE-containing DNA microarrays**

The calculated amounts of probes were mixed and dried down in the dark. The paired probes are resuspended thoroughly in 8.5 μl 4 x Hybridization buffer (Amersham, #RPK0325) and 8.5 μl water and then mixed with 17 μl formamide and vortexed. The mixture is heated at 95°C for 3 min then cooled by spinning at 13K for 2 min. 30 μl of this hybridization solution was dispensed in a thin line across a slide and spread evenly over the surface by laying on of a coverslip and incubated at 42°C for 16 h in a humid and darkened hybridization chamber.

The slides are washed in the dark with gentle agitation. The washes used were 5 min at 37°C in Wash 1 (1 x SSC, 0.2% SDS), two 5 min washes at 37°C in Wash 2 (0.1 x SSC, 0.2% SDS) and two 5 min washes at room temperature in Wash 3 (0.1 x SSC). The slides were air-dried and scanned immediately using Packard Biosciences ScanArray 4000.

**Example 7 Overview processes**

An overview of a representative process is illustrated in Figure 1. This figure shows how the structural integrity of ACEs within a sample may be determined in a two step process: A probing reagent is created and compared to a query population. To create the reagent, cells are treated by a procedure developed to isolate and label a population of DNA fragments from the genome that is enriched in those structurally formed ACEs or a functional subset of them, such as transcriptional promoters, or a structural subset, such as methylated sequences. In this example these DNA fragments are used as a probe to hybridize against a population of sequences on a microarray. Those sequences may be a set of previously characterized ACEs, may physically span a section of the genome or be a

large enough combination of oligonucleotides to allow discretion of complex binding patterns. Following analysis the presence and intensity of the signal reflects the extent to which that particular ACE has formed within that population of cells.

Alternatively, the process may be carried out in parallel using two different markers in order to reveal a differential expression pattern. This process may be employed to increase the signal-to-noise ratio as illustrate in Figure 2. Here, the sensitivity and accuracy of microarray hybridization will be maximized by comparing the signal of two populations of probes generated by the same procedure but isolated from a treated and non-treated population. In this example the probe labeled with Cy3 is enriched for ACEs whilst the Cy5-labeled probe will contain ACEs at the same frequency as they occur in the genome. As the probes are generated the same way they will share similar physical characteristics, such as length and labeling efficiency. Therefore the ratio of intensity seen on a co-ordinate in the array will accurately reflect enrichment of the sequence in one of the probing populations. In this example a structurally formed ACE in the cell population would give rise to a green (Cy3) spot, whilst an unformed site would be yellow (equal amounts of Cy3 and Cy5 bound) or red (Cy5).

Several further additional applications of the invention are illustrated in figures 3 through 6. These include:

i. Differential profiling of regulatory elements (i.e., between two different cell populations). An overview of this process is illustrated in Figure 3. Figure 3 shows how the technology can be used to examine the dynamic nature of ACE formation. In this example two cell types are treated with a similar procedure to generate from each a differently labeled probe population enriched in ACEs. As in Figure 2 the probes will have similar physical characteristics which allows their direct comparison. Hence an ACE formed in one tissue but not the other will label its spot predominately red or green, whilst those formed in both tissues will colour yellow. The exact ratio

of Cy3 to Cy5 will provide information about the relative abundance of that ACE in the tissues. Any ACEs that are absent from both tissues will not be lit up on the array.

ii. Screening for compounds or treatments that impact the regulatory element activity profile. An overview of this process is illustrated in Figure 4. As seen here, profile changes may be monitored to show changes in the pattern of ACE in response to stimuli. Comparative hybridization, as described in Figure 3, can be used to determine, in this example, which ACEs are induced or repressed by treatment with a drug or small molecule. A probe population is prepared from a reference population of untrerated cells and compared to that of a differently labeled probe from the cells following treatment following hybridization to the microarray.

iii. Correlation of regulatory element activation patterns with gene expression patterns to construct regulatory network maps. An overview of this process is illustrated in Figure 5, which establishes a correlation between ACE and expression data. Parallel analysis of gene expression, as detected by use of expression arrays, and ACE structural integrity will give information about ACEs implicated in transcriptional control of specific genes. Such correlation will also enable improved quality control for conventional expression arrays.

iv.Correlation of regulatory element activation with gene expression to provide a powerful biological quality control assay for gene expression arrays. An overview of this process is illustrated in Figure 6.

Other embodiments and uses of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. All references cited herein, including all U.S. and foreign patents and patent applications including U.S. Provisional patent number 60/108,206, and U.S. Patent application number 09/432,576, are specifically and entirely hereby incorporated herein by reference. It is intended that the specification and examples be considered exemplary only,

with the true scope and spirit of the invention indicated by the following claims.

Claims:

1. A nucleic acid array comprising a plurality of active chromatin elements

2. The array of claim 1 wherein each active chromatin element contains a nuclease hypersensitive site.

3. The array of claim 2 which further comprises one or more sets set of nucleic acid sequences that tile across one or more hypersensitive sites.

4. The array of claim 3 wherein the one or more sets each comprise sequences within 200 nucleotides of said hypersensitive site.

5. The array of claim 1 wherein the plurality comprises sequences derived from an organism.

6. The array of claim 3 wherein the organism is selected from the group of organisms consisting of Homo sapien, rat, mouse, zebrafish, drosophila, yeast, C. elegans, and combinations thereof.

7. The array of claim 1 wherein the plurality comprises nucleic acid sequences with lengths from about 16 nucleotides to about 1,500 nucleotides.

8. The array of claim 1 wherein the plurality comprises nucleic acid sequences with lengths from about 100 nucleotides to about 350 nucleotides.

9. The array of claim 1 wherein the plurality comprises at least 100, at least 1,000, at least 10,000. at least 100,000, or at least 1,000,000 active chromatin elements.

10. The array of claim 1 which further contains nucleic acids that represent transcribed sequences.

11. The array of claim 1 which further contains sequences that flank active chromatin elements.

12. The array of claim 1 which further contains repetitive sequences.

13. The array of claim 12 wherein repetitive sequences comprise less than five percent of the total nucleic acids of the array.

14. The array of claim 1 prepared by a process comprised of treating cells with an agent that induces modifications in the nucleic acid.

15. The array of claim 14 wherein the modification is selected from the group consisting of cleavage, methylation, radiation, and combinations thereof.

16. The array of claim 14 wherein the modified nucleic acids are subtracted from nuclease treated unmodified nucleic acids.

17. The array of claim 14 further comprising the step of attaching biotin to the modified nucleic acids.

18. The array of claim 14 further comprising the step of amplifying the modified nucleic acid by PCR.

19. A method for forming the array of claim 1 comprising:

treating genomic DNA with an agent that induces modifications in said DNA;

treating a portion of the modified DNA with nuclease;

subtracting nuclease treated DNA from the modified DNA; and

obtaining an array of active chromatin elements.

20. The method of claim 19 wherein the modifications comprises cleavages that create DNA fragments.

21. The method of claim 20 wherein the DNA fragments are ligated to a linker.

22. The method of claim 21 wherein the linker-ligated DNA fragments are isolated.

23. The method of claim 20 wherein the fragments are cut into smaller sizes by a procedure selected from the group consisting of digestion with a restriction enzyme and sonication.

24. A method for determining the active chromatin element profile of nuclear chromatin of a cell comprising:

treating a portion of said chromatin with an agent that preferentially modifies DNA at hypersensitive sites to form a first set of nucleic acids;

treating another portion of said chromatin with another agent that non-preferentially modifies DNA to form a second set of nucleic acid; and

comparing the first and second sets to obtain said active chromatin element profile.

25. The method of claim 24 wherein the first and second sets are compared by hybridization.

26. The method of claim 25 wherein the first or second set is amplified by PCR.

27. The method of claim 25 wherein the first or second set is labeled with a fluorescent dye.

28. A method for identifying a profile of DNA regulatory elements in a eukaryotic cell comprising:

treating said cell with an agent that modifies DNA of said cell at DNA hypersensitive sites; and

identifying the DNA hypersensitive sites from said reaction with the agent, wherein the nucleotide sequences of said DNA hypersensitive sites and the locations thereof in the DNA of said type of cells constitute a profile of DNA regulatory elements in said type of cells.

29. A method for producing a profile of DNA regulatory elements in eukaryotic cells, comprising:

treating said cells with an agent that modifies eukaryotic DNA at DNA hypersensitive sites;

identifying the DNA hypersensitive sites from said reaction with said agent wherein the nucleotide sequences of said DNA hypersensitive sites and the locations thereof in the DNA of said type of cells constitute a profile of DNA regulatory elements in said type of cells; and

isolating the nucleotide sequences of said hypersensitive sites.

30. The method of claim 29 wherein one or more oligonucleotide linkers are ligated into said nucleotide sequences.

31. The method of claim 30 wherein said oligonucleotide linkers are biotinylated and wherein said isolating is performed using streptavidin-coated magnetic beads.

32. The method of claim 30 further comprising amplifying said nucleotide sequences by polymerase chain reaction.

33. The method of claim 29 wherein the eukaryotic cells are selected from the group consisting of primary cell cultures, cell lines, newly isolated cells from an organism, and combinations thereof.

34. The method of claim 29 wherein the eukaryotic cells are normal cells or abnormal cells.

35. The method of claim 34 wherein the abnormal cells are cancer cells.

36. The method of claim 29 wherein said agent is selected from the group consisting of radiation, a chemical agent, an enzyme, and combinations thereof.

37. The method of claim 36 wherein the radiation comprises UV light radiation.

38. The method of claim 36 wherein the chemical agent is a clastogen.

39. The method of claim 36 wherein the enzyme is selected from the group consisting of specific endonucleases, non-specific endonucleases, topoisomerases, methylases, histone acetylases, histone deacetylases, and combinations thereof.

40. The method of claim 39 wherein the specific endonuclease comprises one or more four-base restriction endonucleases, one or more six-base restriction endonucleases, or combinations thereof.

41. The method of claim 40 wherein the four-base restriction endonuclease is selected from the group consisting of Sau3a, StyI, Nla III, Hsp 92, and combinations thereof.

42. The method of claim 40 wherein the six-base endonuclease is selected from the group consisting of EcoRI, HindIII, and combinations thereof.

43. The method of claim 39 wherein the non-specific endonuclease is DNase I.

44. The method of claim 39 wherein the topoisomerase is topoisomerase II.

45. A profile of DNA regulatory elements in eukaryotic cells as produced by the method of claim 29, said profile comprising isolated nucleotide sequences of the hypersensitive sites.

46. The profile of claim 45 wherein the eukaryotic cells are selected from the group consisting of primary cell cultures, cell lines, newly isolated cells from an eukaryotic species, and combinations thereof.

47. The profile of claim 46 wherein the eukaryotic cells are normal cells or abnormal cells.

48. The profile of claim 47 wherein said abnormal cells are cancer cells.

49. The profile of claim 45 wherein the nucleotide sequences are labeled with a fluorescent dye, a radioactive nucleotide. a magnetic particle, or a combination thereof.

50. A nucleotide array having spotted thereon the profile of claim 45.

51. The nucleotide array of claim 50 wherein the array is fixed to a slide, a chip, or a membrane filter.

52. The nucleotide array of claim 50 wherein one or more copies of said nucleotide sequences of the hypersensitive sites are spotted on said array.

53. A method for detecting DNA regulatory elements in eukaryotic cells comprising:

isolating mRNAs from said cells, converting said mRNA's to cDNA and probing an array to generate a profile;

isolating active regulatory elements from said cells and probing an array to generate a profile, and

comparing the profile from the cDNA probe with the profile from the active regulatory elements probe to correlate regulatory element activity with gene activity.

54. A method for detecting DNA regulatory elements in eukaryotic cells comprising:

isolating mRNAs from the cells;

contacting said isolated mRNAs to the array of claim 23 to detect hybridization signals, wherein the nucleotide sequences of hybridized spots represent the DNA regulatory elements of said cells.

55. A sequence library of active chromatin elements encoding fragments suitable for preparing a profile to determine the regulatory status of a eukaryotic cell sample.

56. The library of claim 55 wherein the fragments are obtained by the step of marking hypersensitive sites of nuclei of the eukaryotic cells of the sample.

57. The library of claim 55 wherein the marking step is carried out by incubating DNAse I with the nuclei to form nicks in DNA at the hypersensitive sites.

58. The library of claim 55 wherein less than five percent of the fragments contain repetitive DNA sequences.

59. The library of claim 55 wherein each fragment comprises a first end generated by cleavage with DNase I and a second end generated by cleavage with another nuclease.

60. The library of claim 61 wherein the library exists in silico.

61. The library of claim 61 wherein the library exists in a vector.

62. The library of claim 67 wherein the vector is selected from the group consisting of microbial cell culture, plasmid vectors and eukaryote cell culture.

63. A library of active chromatin element primers, prepared by obtaining a library of active chromatin element fragments and determining sequences outside the active chromatin element fragments suitable for cloning the active chromatin element fragments.

64. The library of claim 63 which contains at least 10, at least 100, at least 1,000, at least 10,000, at least 100,000 or at least 1,000,000 active chromatin element primers.

65. The library of claim 63 wherein the library is in silico.

66. A method for profiling active chromatin elements from a sample that contains nucleic acid, comprising:

obtaining one or more purified or labeled active chromatin elements from the sample;

contacting the active chromatin elements from step a) with a DNA microarray containing DNA species in separate locations that match sites of the genome; and

detecting binding between the active chromatic elements and sites of the microarray.

67. The method of claim 66 wherein detecting comprises a detection system that involves fluorescence or chemiluminescence to determine position location in the array..

68. The method of claim 66 wherein the DNA microarray comprises immobilized oligonucleotide probes between 5 and 40 nucleotides in length occupying separate known sites of the array.

69. The method of claim 68 wherein the immobilized DNA oligonucleotide probes comprise at least two sets of probes wherein a first set that is exactly complementary to at least one reference sequence and comprises probes that span the reference sequence and which sequentially overlap each other, and at least one additional set of probes, each additional set of which is identical to the first set but for at least one different nucleotide, which different nucleotide is located in the same position in each additional set but which is a different nucleotide in each set.

70. The method of claim 68 wherein the immobilized DNA oligonucleotide probes comprise at least two sets of probes, a first set that is exactly complementary to at least one reference sequence and comprises probes that span the reference sequence and which overlap each other in sequence, and at least one additional set of probes, each additional set of which is identical to the first set but for at least one different nucleotide addition or deletion.

71. The method of claim 66 wherein the DNA species of the DNA microarray are genomic elements.

72. The method of claim 66 wherein the detected binding of step c) is recorded as a reference profile in a computer memory device.

73. A method of ascertaining the effect of an chemical or other environmental perturbation on a regulatory profile of a tissue obtained from a eukaryotic organism comprising;

obtaining a first profile for binding between active chromatic elements of the tissue that is unexposed to the perturbation and a microarray as described in any of claims 1 to 7;

obtaining a second profile for binding between active chromatic elements of the tissue and a microarray as described in claim 66 after exposure of the tissue to the perturbation; and

comparing the first profile with the second profile to determine genetic elements that are effected by the perturbation.

74. The method of claim 73 wherein the perturbation occurs before obtaining the tissue from the organism and wherein the environmental perturbation is selected from the group consisting of an infection of the eukaryotic organism from a microorganism, loss in immune function of the eukaryotic organism, exposure of the tissue to high temperature, exposure of the tissue to low temperature, cancer of the tissue, cancer of another tissue in the eukaryotic organism, irradiation of the tissue, exposure of the tissue to a chemical or other pharmaceutical compound; and aging.

75. The method of claim 73 wherein the perturbation occurs after obtaining the tissue from the organism and wherein the perturbation is selected from the group consisting of exposure of the tissue to high temperature, exposure of the tissue to low temperature, irradiation of the

tissue, exposure of the tissue to a chemical or other pharmaceutical compound, and aging.

76. The method of claim 75 wherein the perturbation is the addition of one or more compounds.

77. The method of claim 76 further comprising the addition of at least one known pharmaceutical compound to the tissue prior to obtaining a profile for binding between active chromatic elements of the tissue and a microarray.

78. A method of discerning at least one set of co-regulated genes in cells of a eukaryotic organism, comprising:

obtaining a first profile for binding between active chromatic elements of the tissue under controlled culture conditions;

obtaining a second profile for binding between active chromatic elements of the tissue under conditions where a known regulator of at least one of the genes is altered with respect to the controlled culture conditions; and

comparing the first profile with the second profile from b) to determine which genetic elements are effected by the alteration of the known regulator.

79. The method of claim 78 wherein the regulator is a hormone, nutrient, or pharmacologically active chemical.

80. A nucleotide array having spotted thereon a set of nucleic acids between 5 and 75 nucleotides long obtained from the profile of any of claims 1 to 70.

81. The nucleotide array of claim 80, wherein said array is a slide, a chip, or a membrane filter.

82. The method of any of claims 19 to 44, wherein the sample is selected from the group consisting of primary cell cultures, cell lines, newly isolated cells from an eukaryotic species, and combinations thereof.

83. A method for profiling active chromatin elements from a sample that contains nucleic acid, comprising:

a) obtaining one or more purified active chromatin elements from the sample and label them;

b) contacting the labeled active chromatin elements from step a) with a DNA microarray containing DNA species in separate locations that match putative or verified regulatory elements; and

c) detecting binding between the active chromatic elements and sites of the microarray.

84. The method of claim 83, wherein detecting comprises a detection system that involves fluorescence or chemiluminescence to determine binding.

85. The method of claim 83, wherein the DNA microarray comprises immobilized oligonucleotide probes between 5 and 40 nucleotides in length occupying separate known sites of the array.

86. The method of claim 85, wherein the immobilized DNA oligonucleotide probes comprise at least two sets of probes wherein a first set that is exactly complementary to at least one reference sequence and comprises probes that span the reference sequence and which sequentially overlap each other, and at least one additional set of probes, each additional set of which is identical to the first set but for at least one different nucleotide, which different nucleotide is located in the same position in each additional set but which is a different nucleotide in each set.

87. The method of claim 85, wherein the immobilized DNA oligonucleotide probes comprise at least two sets of probes, a first set that is exactly complementary to at least one reference sequence and comprises

probes that span the reference sequence and which overlap each other in sequence, and at least one additional set of probes, each additional set of which is identical to the first set but for at least one different nucleotide addition or deletion.

88. The method of claim 83, wherein the DNA species of the DNA microarray are known regulatory sequences.

89. The method of claim 83, wherein the detected binding of step c) is recorded as a reference profile in a computer memory device.

90. A method for profiling active chromatin elements from a sample that contains nucleic acid, comprising:

a) obtaining multiple active chromatin elements from the sample and label them with a first label;

b) obtaining multiple genomic DNA fragments from the sample and label them with a second label;

c) hybridizing the elements from a) and the fragments from b) with a DNA microarray containing DNA species in separate locations that match putative or verified regulatory elements; and

d) determining the ratio of signals from the first and second labels within the array.

91. A method for profiling differential regulatory element activation from two populations that contain nucleic acid, comprising:

a) obtaining multiple active chromatin elements from the first population and labeling them with a first label;

b) obtaining multiple active chromatin elements from the second population and labeling them with a second label;

c) hybridizing the elements from a) and the fragments from b) with a DNA microarray containing DNA species in separate locations that match putative or verified regulatory elements; and

d) determining the ratio of signals from the first and second labels within the array.

92. The method of claim 91, wherein one of the populations is an untreated control, the other population is treated by contact with at least one chemical agent , and the signal ratios obtained in step d) provide an indication of gene regulatory activity by the at least one chemical agent.

93. The method of claim 91, wherein the signal ratios obtained in step d) indicate whether the at least one chemical agent turns on, turns off or has no effect on active chromatin elements.

94. A method for correlating regulatory element activation with gene expression from a sample that contains nucleic acid, comprising:

a) obtaining multiple active chromatin elements from the sample and profiling them on a DNA microarray containing DNA species in separate locations that match putative or verified regulatory elements;

b) isolating RNA from the sample and converting to cDNA;

c) profiling the cDNA on a DNA microarray containing DNA species in separate locations that match putative or verified regulatory elements; and

d) correlate the profile results from a) and c) with gene activity using informatics software.

Figure 1

*Overview of Method for Assaying Regulatory Element Activity Using Regulatory DNA Microarrays*

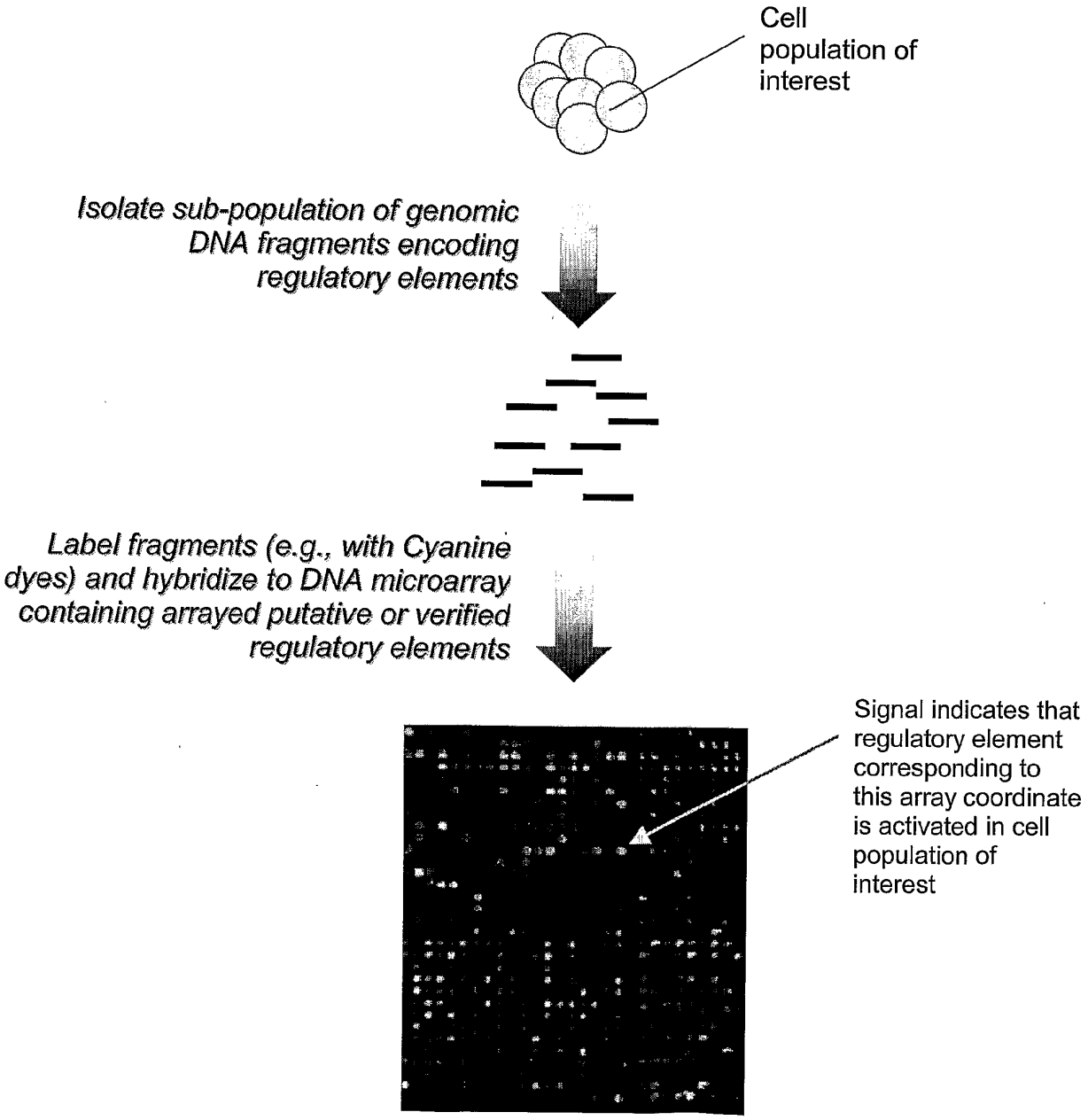


Figure 2

*Profiling Regulatory Element Activity with Two-Dye System to Increase Signal-to-Noise Ratio*

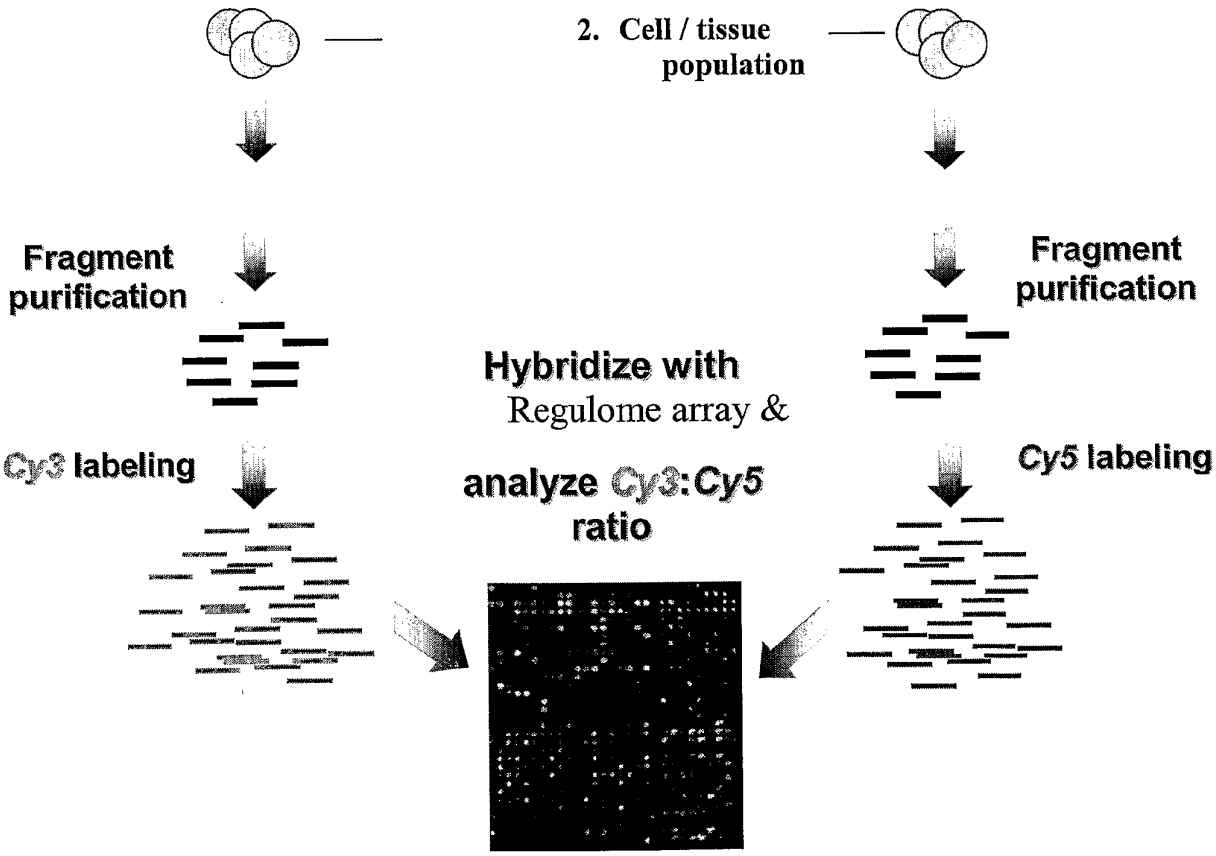


Figure 3

*Profiling of Differential Regulatory Element Activation*

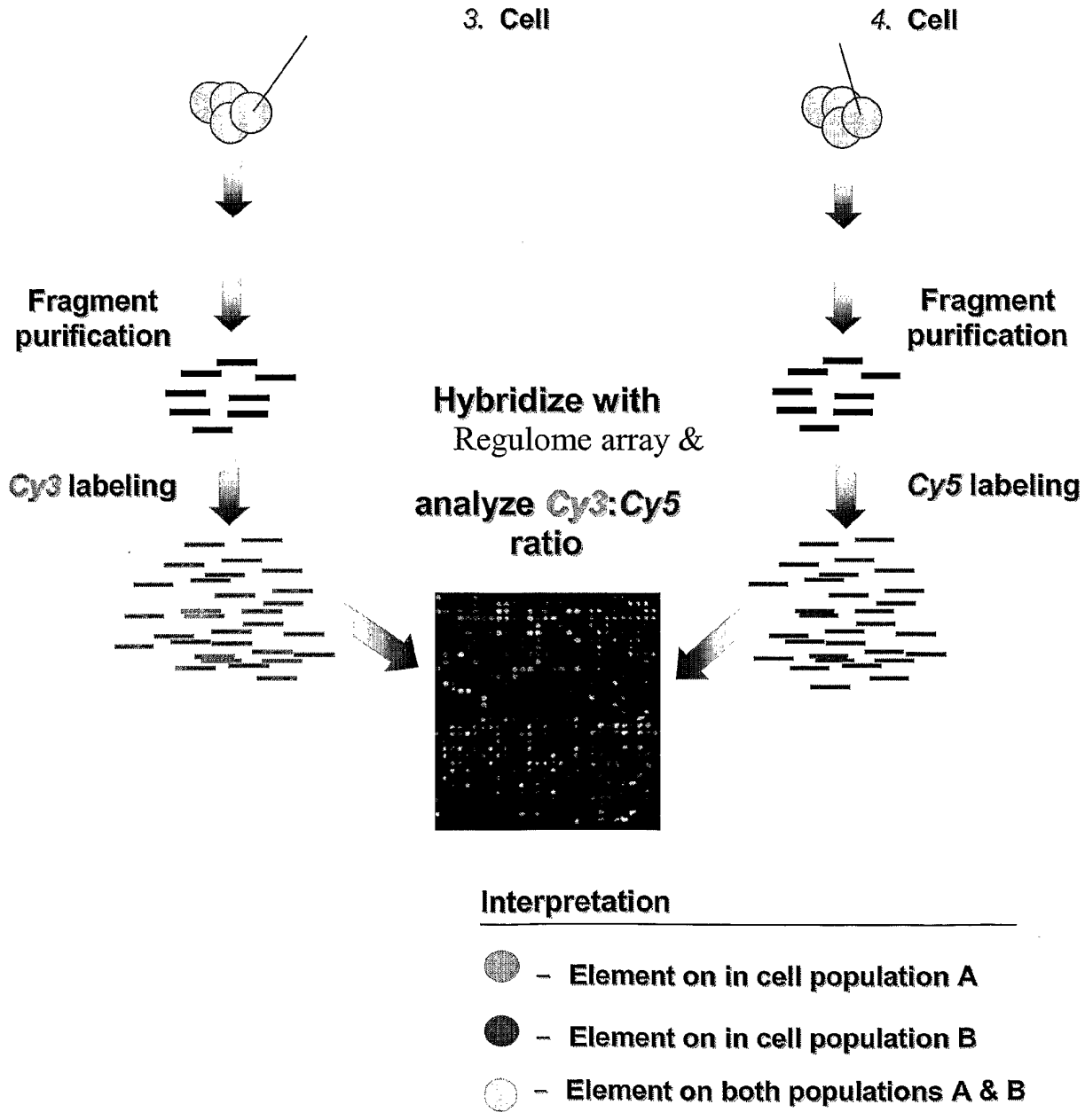


Figure 4

*Use of Regulatory Microarrays to Screen Drugs and/or Small Molecule Compounds*

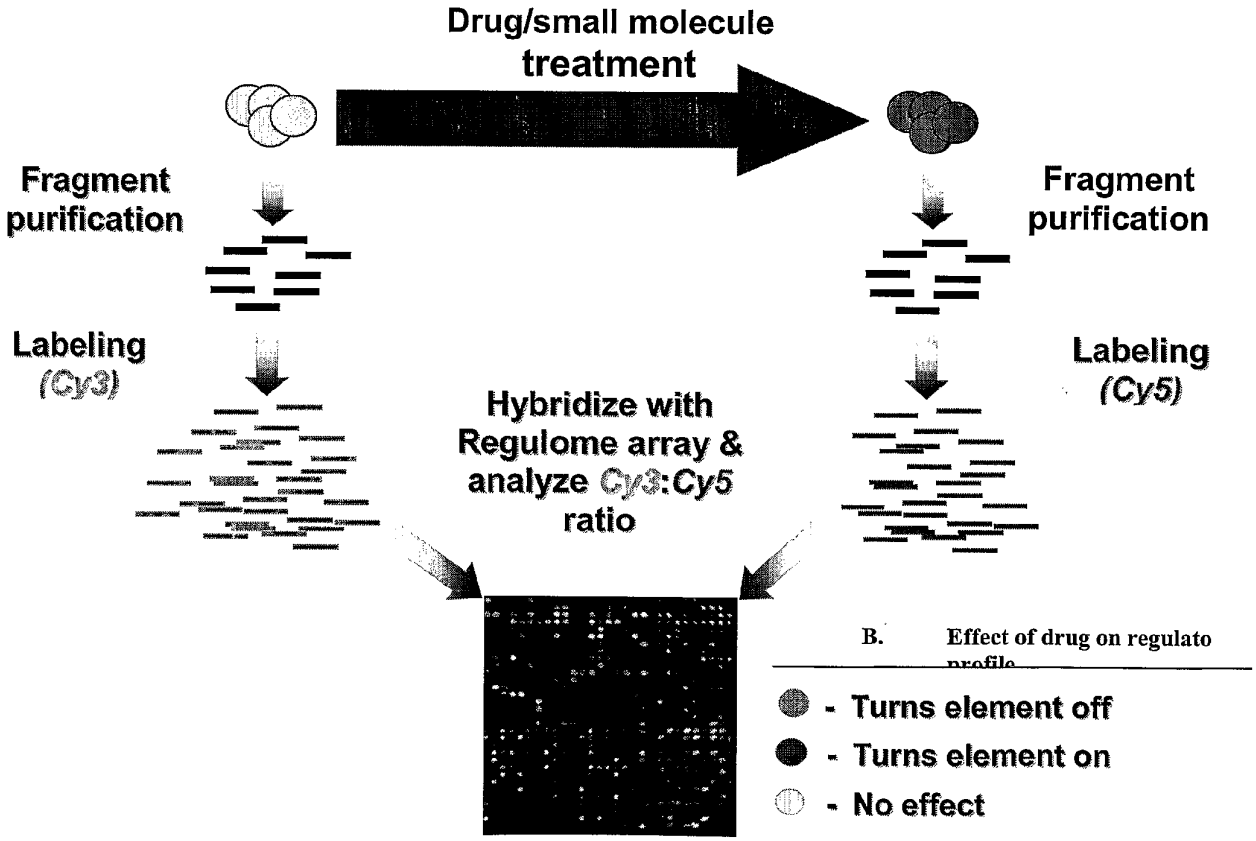


Figure 5

*Correlation of Regulatory Element Activation with Gene Expression*

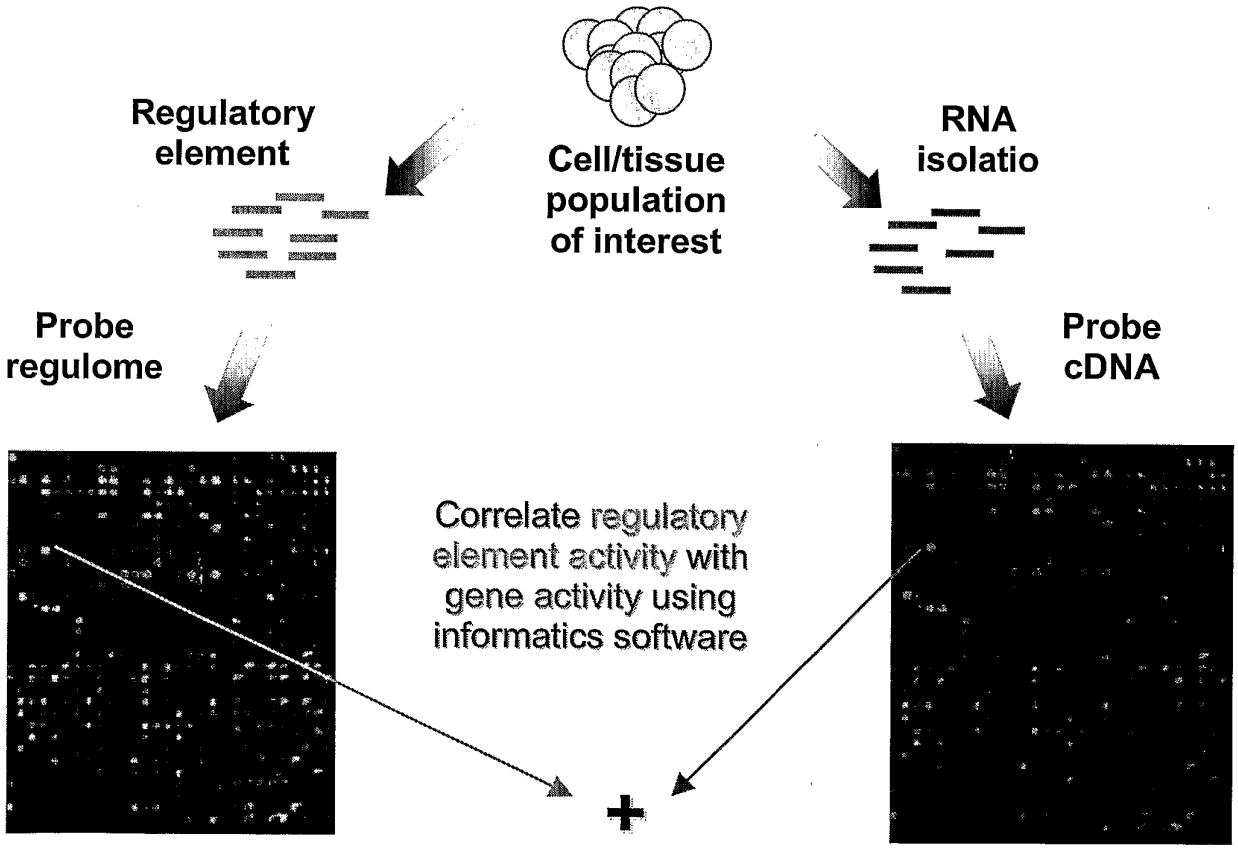
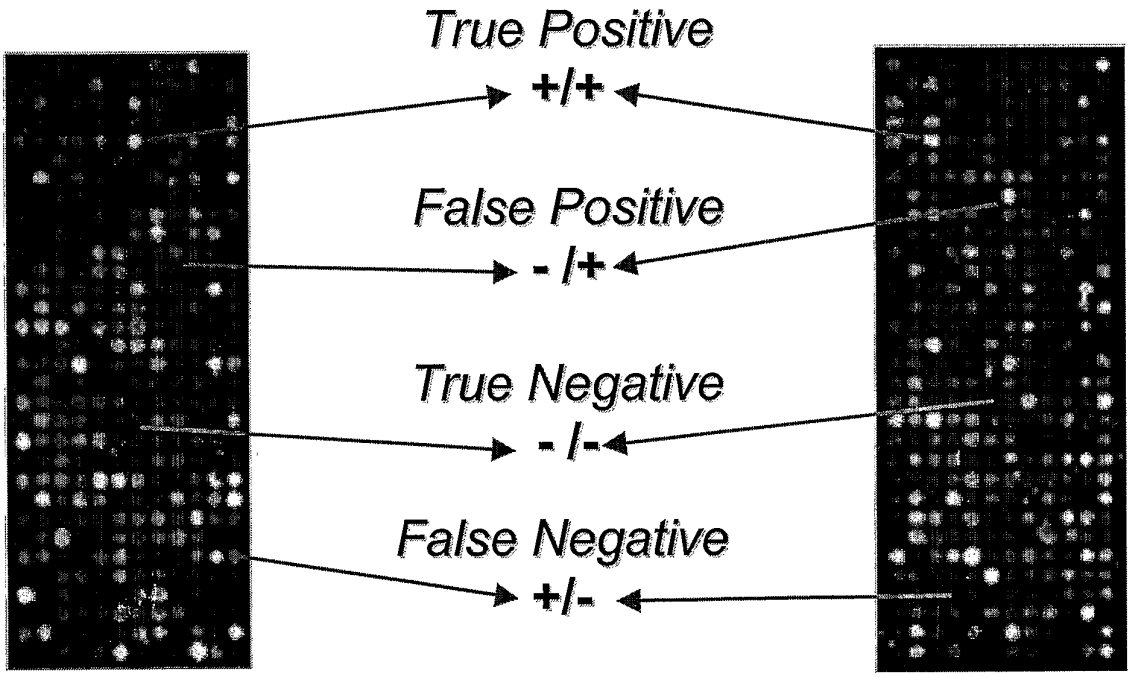


**Figure 6**

***Biological Quality Control for Conventional Expression Arrays***