



- (51) **International Patent Classification:**
C12N 15/11 (2006.01) *C12Q 1/68* (2006.01)
- (21) **International Application Number:**
PCT/US2014/053183
- (22) **International Filing Date:**
28 August 2014 (28.08.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/870,904 28 August 2013 (28.08.2013) US
- (71) **Applicant** (for all designated States except US): **OREGON HEALTH & SCIENCE UNIVERSITY** [US/US];
0690 SW Bancroft Street (mail Code L106TT), Portland,
OR 97239 (US).
- (72) **Inventors:** **MATHERS, William**; 717 8th Street, Lake
Oswego, OR 97229 (US). **YE, Zheng**; 15304 NW Two-
ponds Drive, Portland, OR 97229 (US).
- (74) **Agent:** **BRADLEY, Karri, Kuenzli**; Klarquist Sparkman,
LLP, One World Trade Center, Suite 1600, 121 SW Sal-
mon Street, Portland, OR 97204 (US).
- (81) **Designated States** (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,

BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME,
MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,
OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

- (84) **Designated States** (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,
DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU,
LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,
SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, KM, ML, MR, NE, SN, TD, TG).

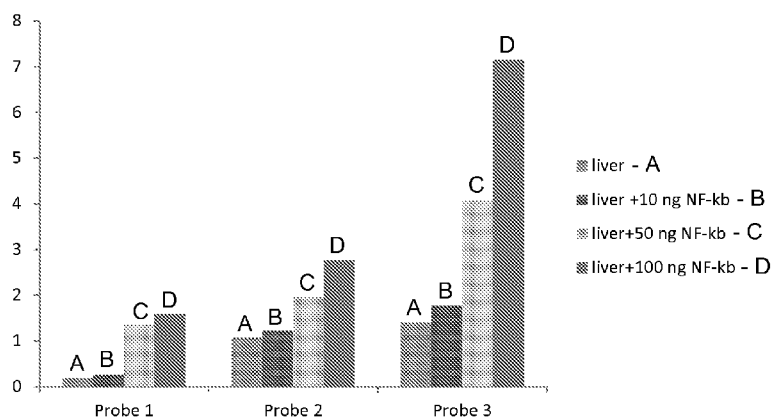
Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
— with sequence listing part of description (Rule 5.2(a))

(54) **Title:** SYNTHETIC OLIGONUCLEOTIDES FOR DETECTION OF NUCLEIC ACID BINDING PROTEINS

FIG. 1

(57) **Abstract:** Synthetic oligonucleotides that comprise a nucleic acid binding protein binding site, PCR primer sequences, and tag sequences that do not bind to nucleic acid binding proteins, with a total length of 85-130 nucleotides are disclosed herein. Also disclosed are libraries and kits comprising the synthetic oligonucleotides as well as methods of detecting nucleic acid binding proteins in a sample using the synthetic oligonucleotides.

SYNTHETIC OLIGONUCLEOTIDES FOR DETECTION OF NUCLEIC ACID BINDING PROTEINS

CROSS REFERENCE TO RELATED APPLICATION

The present application claims the benefit of U.S. Provisional Patent Application No. 61/870,904, entitled "SYNTHETIC OLIGONUCLEOTIDES FOR DETECTION OF NUCLEIC ACID BINDING PROTEINS," filed August 28, 2013, the disclosure of which is incorporated herein by reference in its entirety.

FIELD

Generally, the field is nucleic acid probes and arrays thereof. More specifically, the field is nucleic acid probes that can detect, identify, and quantify nucleic acid binding proteins in a sample.

BACKGROUND

Nucleic acid binding proteins are proteins in the nucleus that bind to nucleic acids such as DNA and RNA to perform any of a number of activities, including the promotion or inhibition of RNA transcription or protein translation. Their binding can be identified with DNA probes that are separated and rendered visible by gel electrophoresis but their detection is limited to one or several binding proteins at one time. Using current methods, it is difficult to quantify nucleic acid binding proteins present in a biological sample. Many currently available technologies comprise the use of probes that self-hybridize, therefore reducing specificity and ease of use. Clearly, new probes that identify nucleic acid binding proteins in biological samples are necessary.

SUMMARY

Disclosed herein are synthetic oligonucleotides that identify DNA binding proteins in a sample. This includes methods of using said oligonucleotides to measure the binding action of a plurality of DNA binding proteins by binding the oligonucleotides to a microarray chip.

The synthetic oligonucleotide comprises sequences that include a nucleic acid binding protein binding site, a sequence that is an RNA polymerase promoter or a binding site for a PCR primer, and a tag sequence that does not bind any nucleic acid binding protein. At least the

nucleic acid binding protein binding site is double stranded and the entire synthetic oligonucleotide can be double stranded. The probe is between 85 and 130 base pairs in length, including between 90 and 98 nucleotides in length. The sequence may comprise T7, SP6, or T5 promoters including SEQ ID NO: 1 and SEQ ID NO: 2. In some examples, the tag sequence is any of SEQ ID NOs: 3-98.

Further disclosed are sets of synthetic oligonucleotides each with different nucleic acid binding protein binding sites as well as kits comprising sets of synthetic oligonucleotides and microarrays comprising oligonucleotides complementary to the tag sequences. The oligonucleotides on the microarray are addressable.

Also disclosed is a method of detecting a DNA binding protein in a sample involving contacting one of the disclosed synthetic oligonucleotides with a sample. The sample is then subjected to conditions that allow binding of any DNA binding proteins in the sample to the synthetic oligonucleotides to form protein/oligonucleotide complexes. The protein/oligonucleotide complexes are purified by electrophoresis. The oligonucleotides are labeled and hybridized to an array. Detection of the label within the context of the array indicates the presence of the DNA binding protein in the sample. The label may be any label, including a fluorescent label. The label may be incorporated into the oligonucleotide by any method including method involving use of RNA polymerase or polymerase chain reaction. Electrophoresis may be performed using any device or method including the devices and methods described in US2012/0160683, now issued U.S. Patent No. 8,496,798, which is hereby incorporated by reference in its entirety.

The foregoing and other objects, features, and advantages of the invention will become more apparent from the following detailed description, which proceeds with reference to the accompanying figure and sequence listing.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a bar graph of the relative strength (Y axis) of three different probes (Probe 1 - SEQ ID NO: 197) (Probe 2 - SEQ ID NO: 198) and Probe 3 - SEQ ID NO: 99 incubated with nuclear extracts comprising the indicated amounts of added nuclear factor kappa-light-chain enhance of activated B cells (NF- κ B), purified by electrophoresis, and bound to a microarray as described herein.

BRIEF DESCRIPTION OF SEQUENCES

The nucleic acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. All sequence database accession numbers referenced herein are understood to refer to the version of the sequence identified by that accession number as it was available on the designated date. In the accompanying sequence listing:

SEQ ID NO: 1 is a universal forward primer sequence binding site.

SEQ ID NO: 2 is a universal reverse primer sequence binding site.

5 SEQ ID NOs: 3-98 are examples of tag sequences

SEQ ID NOs: 99- 100 are synthetic oligonucleotide probes that can bind NF- κ B

SEQ ID NOs: 101-102 are synthetic oligonucleotide probes that can bind P53.

SEQ ID NOs: 103-104 are synthetic oligonucleotide probes that can bind AR (androgen receptor).

10 SEQ ID NOs: 105-106 are synthetic oligonucleotide probes that can bind Creb.

SEQ ID NOs: 107-108 are synthetic oligonucleotide probes that can bind activator protein 1 (AP-1).

SEQ ID NOs: 109-110 are synthetic oligonucleotide probes that can bind early growth response protein 1(EGR-1).

15 SEQ ID NOs: 111-112 are synthetic oligonucleotide probes that can bind activating protein 2 (AP-2).

SEQ ID NOs: 113-114 are synthetic oligonucleotide probes that can bind Nkx-3.1.

SEQ ID NOs: 115-116 are synthetic oligonucleotide probes that can bind prostate-specific antigen (PSA).

20 SEQ ID NOs: 117-118 are synthetic oligonucleotide probes that can bind C-myb.

SEQ ID NOs: 119-120 are synthetic oligonucleotide probes that can bind peroxisome proliferator-activated receptor (PPAR).

SEQ ID NOs: 121-122 are synthetic oligonucleotide probes that can bind PPAR gamma.

SEQ ID NOs: 123-124 are synthetic oligonucleotide probes that can bind Oct-1.

25 SEQ ID NOs: 125- 126 are synthetic oligonucleotide probes that can bind hypoxia-inducible factor 1 alpha (HIF-1 α).

SEQ ID NOs: 127-128 are synthetic oligonucleotide probes that can bind E2F transcription factor 1 (E2F-1).

SEQ ID NOs: 129-130 are synthetic oligonucleotide probes that can bind CEBP (CCAAT-enhancer binding protein).

5 SEQ ID NOs: 131-132 are synthetic oligonucleotide probes that can bind B-cell lymphoma 6 (Bcl-6).

SEQ ID NOs: 133-134 are synthetic oligonucleotide probes that can bind SRE (sterol regulatory element).

SEQ ID NOs: 135-136 are synthetic oligonucleotide probes that can bind Oxo3a.

10 SEQ ID NOs: 137-138 are synthetic oligonucleotide probes that can bind forkhead box a (Foxa).

SEQ ID NO: 139 and 140 are synthetic oligonucleotide probes that can bind forkhead box o (Foxo).

15 SEQ ID NOs: 141 and 142 are synthetic oligonucleotide probes that can bind PR (progesterone receptor).

SEQ ID NOs: 143 and 144 are synthetic oligonucleotide probes that can bind RAR (retinoic acid receptor).

SEQ ID NOs: 145 and 146 are synthetic oligonucleotide probes that can bind Snail.

SEQ ID NOs: 147 and 148 are synthetic oligonucleotide probes that can bind Stat1.

20 SEQ ID NOs: 149 and 150 are synthetic oligonucleotide probes that can bind Stat3.

SEQ ID NOs: 151 and 152 are synthetic oligonucleotide probes that can bind Stat4.

SEQ ID NOs: 153 and 154 are synthetic oligonucleotide probes that can bind Stat5.

SEQ ID NOs: 155 and 156 are synthetic oligonucleotide probes that can bind Stat5 and Stat6.

25 SEQ ID NOs: 157 and 158 are synthetic oligonucleotide probes that can bind Brn-3.

SEQ ID NOs: 159 and 160 are synthetic oligonucleotide probes that can bind CBF (CCAAT binding factor).

SEQ ID NOs: 161 and 162 are synthetic oligonucleotide probes that can bind CDP (CCAAT displacement protein).

30 SEQ ID NOs: 163 and 164 are synthetic oligonucleotide probes that can bind CCCTC-binding factor (CTCF).

SEQ ID NO: 165 and 166 are synthetic oligonucleotide probes that can bind Fast-1.

SEQ ID NOs: 167 and 168 are synthetic oligonucleotide probes that can bind GATA binding protein 2 (GATA2).

SEQ ID NOs: 169 and 170 are synthetic oligonucleotide probes that can bind runt-related transcription factor 3 (RUNX3).

5 SEQ ID NOs: 171 and 172 are synthetic oligonucleotide probes that can bind ETS-related gene (ERG).

SEQ ID NOs: 173 and 174 are synthetic oligonucleotide probes that can bind FLI-1 (Friend leukemia virus integration 1).

10 SEQ ID NOs: 175 and 176 are synthetic oligonucleotide probes that can bind hepatocyte nuclear factor 4 (HNF-4).

SEQ ID NOs: 177 and 178 are synthetic oligonucleotide probes that can bind IRF-1 (interferon regulatory factor 1).

SEQ ID NOs: 179 and 180 is a synthetic oligonucleotide probe that can bind nuclear factor 1 (NF-1).

15 SEQ ID NOs: 181 and 182 are synthetic oligonucleotide probes that can bind nuclear factor, erythroid 2 (NF-e2).

SEQ ID NOs: 183 and 184 are synthetic oligonucleotide probes that can bind upstream stimulatory factor (USF-1).

20 SEQ ID NO: 185 is a synthetic oligonucleotide probe that can bind Interferon-Gamma Activated Sequence (GAS)/ Interferon-Stimulated Response Element (ISRE).

SEQ ID NO: 186 is a synthetic oligonucleotide probe that can bind Smad.

SEQ ID NOs: 187-188 are synthetic oligonucleotide probes that can bind Smad.

SEQ ID NO: 189-192 are synthetic oligonucleotide probes that lack a nucleic acid binding protein binding site used as negative controls.

25 SEQ ID NOs: 193 and 194 are synthetic oligonucleotide probes that can bind Myc-Max.

SEQ ID NO: 195 is a T7 promoter site

SEQ ID NO: 196 is an SP6 promoter site.

SEQ ID NOs: 197-199 are synthetic oligonucleotides that can bind NF- κ b.

SEQ ID NOs: 200-202 are synthetic oligonucleotides with mutated NF- κ b binding sites.

30 SEQ ID NOs: 203-205 are synthetic oligonucleotides that can bind estrogen receptor.

SEQ ID NOs: 206-208 are synthetic oligonucleotides with mutant estrogen receptor binding sites.

SEQ ID NOs: 209-211 are synthetic oligonucleotides that can bind SP1

SEQ ID NOs: 212-214 are synthetic oligonucleotides with mutant SP1 binding sites.

DETAILED DESCRIPTION

I. Terms

5 Unless otherwise noted, technical terms are used according to conventional usage.

Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al. (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive*
10 *Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context
15 clearly indicates otherwise. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of this disclosure, suitable methods and materials are described below. The term “comprises” means “includes.”

20 All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

To facilitate review of the various embodiments of this disclosure, the following
25 explanations of specific terms are provided:

Amplifying a nucleic acid molecule: to increase the number of copies of a nucleic acid molecule, such as a double stranded synthetic oligonucleotide described herein. The resulting products are called amplification products. An example of in vitro amplification is the polymerase chain reaction (PCR). Other examples of in vitro amplification techniques include
30 quantitative real-time PCR, strand displacement amplification (see US Patent Number 5,744,311); transcription-free isothermal amplification (see US Patent Number 6,033,881); repair chain reaction amplification (see WO90/01069); ligase chain reaction amplification (see EP application 320 308); gap filling ligase chain reaction amplification (see US Patent Number

5,427,930); coupled ligase detection and PCR (see USPN 6,027,889); and NASBA™ RNA transcription-free amplification (see USPN 6,025,134).

Amplification of a nucleic acid molecule may also include the production of RNA molecules from a DNA template through a transcription reaction (such as an in vitro transcription reaction). In this case, a sequence to be amplified comprises an RNA polymerase promoter that binds to an RNA polymerase which, in the presence of ribonucleotides, produces multiple copies of an RNA sequence.

Amplification of a nucleic acid sequence may be used for any of a number of purposes, including increasing the amount of a rare sequence to be analyzed by other methods. It may also be used to identify a sequence directly (for example, though an amplification refractory mutation system) or as part of a DNA sequencing method.

Array: An arrangement of molecules, such as biological macromolecules (such as peptides, antibodies, or nucleic acid molecules) or biological samples (such as tissue sections), in addressable locations on or in a solid support or substrate. A "microarray" is an array that is miniaturized so as to require or be aided by microscopic examination for evaluation or analysis. Arrays are sometimes called chips or biochips. The array of molecules ("features") makes it possible to carry out a large number of analyses on a sample at one time. In certain example arrays, one or more molecules (such as an oligonucleotide probe) will occur on the array a plurality of times (such as two or three times), for instance to provide internal controls. The number of addressable locations on the array can vary, for example from at least one, to at least 2, to at least 5, to at least 10, at least 20, at least 30, at least 50, at least 75, at least 100, at least 150, at least 200, at least 300, at least 500, least 550, at least 600, at least 800, at least 1000, at least 10,000, at least 100,000, or more. In particular examples, an array includes nucleic acid molecules, such as oligonucleotide sequences that are at least 10 nucleotides in length.

In particular examples, the array may comprise an oligonucleotide that can hybridize to a tag sequence on a nucleic acid probe. In further examples, the array may be a universal array that can bind a set of tags. Such an array might be stripped and reused as appropriate (see US Patent Application Publication Number 2009/0061424 which is incorporated by reference herein.)

Within an array, each arrayed biomolecule is addressable, in that its location can be reliably and consistently determined within at least two dimensions of the array. The feature application location on an array can assume different shapes. For example, the array can be regular (such as arranged in uniform rows and columns) or irregular. Thus, in ordered arrays the

location of each sample is assigned to the sample at the time when it is applied to the array, and a key may be provided in order to correlate each location with the appropriate target or feature position. Often, ordered arrays are arranged in a symmetrical grid pattern, but samples could be arranged in other patterns (such as in radially distributed lines, spiral lines, or ordered clusters).

- 5 Addressable arrays usually are computer readable, in that a computer can be programmed to correlate a particular address on the array with information about the sample at that position (such as hybridization or binding data, including for instance signal intensity). In some examples of computer readable formats, the individual features in the array are arranged regularly, for instance in a Cartesian grid pattern, which can be correlated to address information by a
10 computer.

- Binding or Stable Binding:** Physical methods of detecting the binding of complementary strands of nucleic acid molecules, include but are not limited to, such methods as DNase I or chemical footprinting, gel shift and affinity cleavage assays, Northern blotting, dot blotting and light absorption detection procedures. For example, one method involves observing
15 a change in light absorption of a solution containing an oligonucleotide (or an analog) and a target nucleic acid at 220 to 300 nm as the temperature is slowly increased. If the oligonucleotide or analog has bound to its target, there is an increase in absorption at a characteristic temperature as the oligonucleotide (or analog) and target disassociate from each other, or melt. In another example, the method involves detecting a signal, such as a detectable
20 label, present on one or both nucleic acid molecules. The binding between an oligomer and its target nucleic acid is frequently characterized by the temperature (T_m) at which 50% of the oligomer is melted from its target. A higher T_m indicates a stronger or more stable complex relative to a complex with a lower T_m .

- Buffer solution:** An aqueous solution consisting of a mixture of a weak acid and its
25 conjugate base or a weak base and its conjugate acid. It has the property that the pH of the solution changes very little when a small amount of acid or base is added to it. Buffer solutions can keep pH at a nearly constant value in a wide variety of chemical applications.

- Contacting:** Placement in direct physical association; includes solid, liquid, and gaseous associations. Contacting includes contact between one molecule and another molecule.
30 Contacting can occur in vitro with isolated cells or tissue or in vivo by administering to a subject. The concept of contacting may also be encompassed by adding a molecule to a solid, liquid, or gaseous mixture.

Control: A control may be any sample or standard used for comparison with an experimental sample. In some examples, the control is a sample obtained from a healthy patient or a non-tumor tissue sample obtained from a patient diagnosed with cancer (such as non-tumor tissue adjacent to the tumor). In some examples, the control is a historical control or standard reference value or range of values (such as a previously tested control sample, such as a group of cancer patients with poor prognosis, or group of samples that represent baseline or normal values, such as the level of one or more of the genes disclosed herein in non-tumor tissue). A control may also serve as a threshold level of expression of a biomarker that indicates a particular disease outcome.

Electrophoresis: The process of separating a mixture of charged molecules based on the different mobility of these charged molecules in response to an applied electric current. A particular type of electrophoresis is gel electrophoresis. The mobility of a molecule is generally related to the characteristics of the charged molecule, such as size, shape, and surface charge amongst others. The mobility of a molecule also is influenced by the electrophoretic medium, for example the composition of the electrophoresis gel. For example, when the electrophoretic medium is cross-linked acrylamide (polyacrylamide) increasing the percentage of acrylamide in the gel reduces the size of the resulting pores in the gel and retards the mobility of a molecule relative to a gel with a lower percentage of acrylamide (larger pore size). Gel electrophoresis can be performed for analytical purposes, but can be used as a preparative technique to partially purify molecules prior to use of other methods, such as mass spectrometry, PCR, cloning, DNA sequencing, array analysis, and immuno-blotting.

Label: A detectable compound or composition that is conjugated directly or indirectly to another molecule (such as an oligonucleotide) to facilitate detection of that molecule. Specific, non-limiting examples of labels include fluorescent molecules, enzymatic linkages and radioactive isotopes.

Nucleic acid molecules: A deoxyribonucleotide or ribonucleotide polymer including, without limitation, cDNA, mRNA, genomic DNA, methylated DNA, and synthetic (such as chemically synthesized) nucleic acids such as DNA, RNA, and/or methylated oligonucleotides. The nucleic acid molecule can be double-stranded or single-stranded. Where single-stranded, the nucleic acid molecule can be the sense strand or the antisense strand. In addition, nucleic acid molecule can be circular or linear. A nucleic acid molecule may also be termed a polynucleotide and the terms are used interchangeably.

Oligonucleotide: A plurality of joined nucleotides joined by native phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide.

Particular oligonucleotides and oligonucleotide analogs can include linear sequences up to about 200 nucleotides in length, for example a sequence (such as DNA or RNA) that is at least 6 nucleotides, for example at least 8, at least 10, at least 15, at least 20, at least 21, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 100 or even at least 200 nucleotides long, including 85-130 nucleotides long.

An oligonucleotide can be used to detect the presence of a complementary sequence by molecular hybridization. Such an oligonucleotide can also be termed a probe. In particular examples, such oligonucleotides include a label that permits detection of oligonucleotide probe:target sequence hybridization complexes. In a particular example, a probe includes at least one fluorophore, such as an acceptor fluorophore or donor fluorophore. For example, a fluorophore can be attached at the 5'- or 3'-end of the probe. In specific examples, the fluorophore is attached to the base at the 5'-end of the probe, the base at its 3'-end, the phosphate group at its 5'-end or a modified base, such as a T internal to the probe.

An oligonucleotide can be used to prime a nucleic acid amplification. Such an oligonucleotide may also be termed a primer. An oligonucleotide primer can be annealed to a complementary target nucleic acid molecule by nucleic acid hybridization to form a hybrid between the primer and the target nucleic acid strand. A primer can be extended along the target nucleic acid molecule by a polymerase enzyme.

The specificity of an oligonucleotide primer increases with its length. Thus, for example, a primer that includes 30 consecutive nucleotides will anneal to a target sequence with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, to obtain greater specificity, probes and primers can be selected that include at least 15, 20, 25, 30, 35, 40, 45, 50, 85, 100, or 120 or more consecutive nucleotides. In particular examples, a primer is at least 15 nucleotides in length, such as at least 15 contiguous nucleotides complementary to a target nucleic acid molecule.

Primer pairs can be used for amplification of a nucleic acid sequence, for example, by PCR, real-time PCR, or other nucleic-acid amplification methods known in the art. An “upstream” or “forward” primer is a primer 5' to a reference point on a nucleic acid sequence. A

“downstream” or “reverse” primer is a primer 3' to a reference point on a nucleic acid sequence. In general, at least one forward and one reverse primer are included in an amplification reaction.

Nucleic acid probes and/or primers can be readily prepared based on the nucleic acid molecules provided herein. PCR primer pairs and probes can be derived from a known sequence for example, by using any of a number of computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA) or PRIMER EXPRESS® Software (Applied Biosystems, Foster City, CA).

Methods for preparing and using oligonucleotide and other nucleic acid probes and primers and methods for labeling and guidance in the choice of labels appropriate for various purposes are described, for example, in Sambrook et al (In Molecular Cloning: A Laboratory Manual, CSHL, New York, 1989), Ausubel et al (ed.) (In Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1998), and Innis et al (PCR Protocols, A Guide to Methods and Applications, Academic Press, Inc., San Diego, CA, 1990).

Promoter: An array of nucleic acid control sequences, which directs transcription of a nucleic acid. Typically, a eukaryotic a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription, such as specific DNA sequences that are recognized by proteins known as transcription factors.

In prokaryotes, a promoter is recognized by RNA polymerase and an associated sigma factor, which in turn are brought to the promoter DNA by an activator protein binding to its own DNA sequence nearby.

Sample (or biological sample): A specimen containing genomic DNA, RNA (including mRNA), protein, or combinations thereof, obtained from a subject. As used herein, biological samples include cells, tissues, and bodily fluids, such as: blood; derivatives and fractions of blood, such as plasma or serum; extracted galls; biopsied or surgically removed tissue, including tissues that are, for example, unfixed, frozen, fixed in formalin and/or embedded in paraffin; tears; milk; skin scrapes; surface washings; urine; sputum; cerebrospinal fluid; prostate fluid; pus; or bone marrow aspirates. In a particular example, a sample includes a tumor biopsy (such as a prostate tumor biopsy). In another example, a sample includes circulating tumor cells, such as tumor cells present in blood of a subject with a tumor.

Obtaining a biological sample includes, but need not be limited to any method of collecting a particular sample known in the art. Obtaining a biological sample from a subject

also encompasses receiving a sample that was collected at a different location than where a method is performed; receiving a sample that was collected by a different individual than an individual that performs the method, receiving a sample that was collected at any time period prior to the performance of the method, receiving a sample that was collected using a different
5 instrument than the instrument that performs the method, or any combination of these. Obtaining a biological sample from a subject also encompasses situations in which the collection of the sample and performance of the method are performed at the same location, by the same individual, at the same time, using the same instrument, or any combination of these.

A biological sample encompasses any fraction of a biological sample or any component
10 of a biological sample that may be isolated and/or purified from the biological sample. For example: when cells are isolated from blood or tissue, including specific cell types sorted on the basis of biomarker expression; or when nucleic acid or protein is purified from a fluid or tissue; or when blood is separated into fractions such as plasma, serum, buffy coat PBMC's or other cellular and non-cellular fractions on the basis of centrifugation and/or filtration. A biological
15 sample further encompasses biological samples or fractions or components thereof that have undergone a transformation of matter or any other manipulation. For example, a cDNA molecule made from reverse transcription of mRNA purified from a biological sample may be termed a biological sample.

A biological sample may be obtained from any source including any organism (living,
20 dead, or extinct) or any cells derived from an organism or any artificial cells that may comprise nucleic acids and/or nucleic acid binding proteins. Examples include animals, plants, bacteria, archaea, fungi, prions, viruses, or any other source. Biological samples may also include strata that may contain nucleic acids or nucleic acid binding proteins including water, soil, air, fossilized material, including strata from both terrestrial and extraterrestrial origin.

Transcription factor: A protein that regulates transcription. In particular, transcription
25 factors regulate the binding of RNA polymerase and the initiation of transcription. A transcription factor binds upstream or downstream to either enhance or repress transcription of a gene by assisting or blocking RNA polymerase binding. The term transcription factor includes both inactive and activated transcription factors.

Transcription factors are typically modular proteins that affect regulation of gene
30 expression. Exemplary transcription factors include AAF, abl, ADA2, ADA-NF1, AF-1, AFP1, AhR, AIIN3, ALL-1, alpha-CBF, alpha-CP1, alpha-CP2a, alpha-CP2b, alphaHo, alphaH2-alphaH3, Alx-4, aMEF-2, AML1, AML1a, AML1b, AML1c, AML1DeltaN, AML2, AML3,

AML3a, AML3b, AMY-1L, A-Myb, ANF, AP-1, AP-2alphaA, AP-2alphaB, AP-2beta, AP-2gamma, AP-3 (1), AP-3 (2), AP-4, AP-5, APC, AR, AREB6, Arnt, Arnt (774 M form), ARP-1, ATBF1-A, ATBF1-B, ATF, ATF-1, ATF-2, ATF-3, ATF-3deltaZIP, ATF-a, ATF-adelta, ATPF1, Barhl1, Barhl2, Barx1, Barx2, Bcl-3, BCL-6, BD73, beta-catenin, Bin1, B-Myb, BP1, BP2, brahma, BRCA1, Brn-3a, Brn-3b, Brn-4, BTEB, BTEB2, B-TFIID, C/EBPalph, C/EBPbeta, C/EBPdelta, CACCBinding factor, Cart-1, CBF (4), CBF (5), CBP, CCAAT-binding factor, CCMT-binding factor, CCF, CCG1, CCK-1a, CCK-1b, CD28RC, cdk2, cdk9, Cdx-1, CDX2, Cdx-4, CFF, Chx10, CLIM1, CLIM2, CNBP, CoS, COUP, CP1, CP1A, CP1C, CP2, CPBP, CPE binding protein, CREB, CREB-2, CRE-BP1, CRE-BPa, CREMalpha, CRF, Crx, CSBP-1, CTCF, CTF, CTF-1, CTF-2, CTF-3, CTF-5, CTF-7, CUP, CUTL1, Cx, cyclin A, cyclin T1, cyclin T2, cyclin T2a, cyclin T2b, DAP, DAX1, DB1, DBF4, DBP, DbpA, DbpAv, DbpB, DDB, DDB-1, DDB-2, DEF, deltaCREB, deltaMax, DF-1, DF-2, DF-3, Dlx-1, Dlx-2, Dlx-3, Dlx4 (long isoform), Dlx-4 (short isoform), Dlx-5, Dlx-6, DP-1, DP-2, DSIF, DSIF-p14, DSIF-p160, DTF, DUX1, DUX2, DUX3, DUX4, E, E12, E2F, E2F+E4, E2F+p107, E2F-1, E2F-2, E2F-3, E2F-4, E2F-5, E2F-6, E47, E4BP4, E4F, E4F1, E4TF2, EAR2, EBP-80, EC2, EF1, EF-C, EGR1, EGR2, EGR3, EIIaE-A, EIIaE-B, EIIaE-Calpha, EIIaE-Cbeta, EivF, Elf-1, Elk-1, Emx-1, Emx-2, Emx-2, En-1, En-2, ENH-bind. prot., ENKTF-1, EPAS1, epsilonF1, ER, Erg-1, Erg-2, ERR1, ERR2, ETF, Ets-1, Ets-1 deltaVil, Ets-2, Evx-1, F2F, factor 2, Factor name, FBP, f-EBP, FKBP59, FKHL18, FKHRL1P2, Fli-1, Fos, FOXB1, FOXC1, FOXC2, FOXD1, FOXD2, FOXD3, FOXD4, FOXE1, FOXE3, FOXF1, FOXF2, FOXG1a, FOXG1b, FOXG1c, FOXH1, FOXI1, FOXJ1a, FOXJ1b, FOXJ2 (long isoform), FOXJ2 (short isoform), FOXJ3, FOXK1a, FOXK1b, FOXK1c, FOXL1, FOXM1a, FOXM1b, FOXM1c, FOXN1, FOXN2, FOXN3, FOX01a, FOX01b, FOXO2, FOXO3a, FOXO3b, FOXO4, FOXP1, FOXP3, Fra-1, Fra-2, FTF, FTS, G factor, G6 factor, GABP, GABP-alpha, GABP-beta1, GABP-beta2, GADD 153, GAF, gammaCMT, gammaCAC1, gammaCAC2, GATA-1, GATA-2, GATA-3, GATA-4, GATA-5, GATA-6, Gbx-1, Gbx-2, GCF, GCMA, GCN5, GF1, GLI, GLI3, GR alpha, GR beta, GRF-1, Gsc, Gscl, GT-IC, GT-IIA, GT-IIBalpha, GT-IIBbeta, H1TF1, H1TF2, H2RIIBP, H4TF-1, H4TF-2, HAND1, HAND2, HB9, HDAC1, HDAC2, HDAC3, hDaxx, heat-induced factor, HEB, HEB1-p67, HEB1-p94, HEF-1 B, HEF-1T, HEF-4C, HEN1, HEN2, Hesx1, Hex, HIF-1, HIF-1alpha, HIF-1beta, HiNF-A, HiNF-B, HiNF-C, HiNF-D, HiNF-D3, HiNF-E, HiNF-P, HIP1, HIV-EP2, Hlf, HLTF, HLTF (Met123), HLX, HMBP, HMG I, HMG I(Y), HMG Y, HMGI-C, HNF-1A, HNF-1B, HNF-1C, HNF-3, HNF-3alpha, HNF-3beta, HNF-3gamma, HNF4, HNF-4alpha, HNF4alpha1, HNF-4alpha2, HNF-4alpha3, HNF-4alpha4,

HNF4gamma, HNF-6alpha, hnRNP K, HOX11, HOXA1, HOXA10, HOXA10 PL2, HOXA11,
 HOXA13, HOXA2, HOXA3, HOXA4, HOXA5, HOXA6, HOXA7, HOXA9A, HOXA9B,
 HOXB-1, HOXB13, HOXB2, HOXB3, HOXB4, HOXB5, HOXB6, HOXA5, HOXB7,
 HOXB8, HOXB9, HOXC10, HOXC11, HOXC12, HOXC13, HOXC4, HOXC5, HOXC6,
 5 HOXC8, HOXC9, HOXD10, HOXD11, HOXD12, HOXD13, HOXD3, HOXD4, HOXD8,
 HOXD9, Hp55, Hp65, HPX42B, HrpF, HSF, HSF1 (long), HSF1 (short), HSF2, hsp56, Hsp90,
 IBP-1, ICER-II, ICER-ligamma, ICSBP, Id1, Id1 H', Id2, Id3, Id3/Heir-1, IF1, IgPE-1, IgPE-2,
 IgPE-3, IkappaB, IkappaB-alpha, IkappaB-beta, IkappaBR, II-1 RF, IL-6 RE-BP, II-6 RF,
 INSAF, IPF1, IRF-1, IRF-2, irlB, IRX2a, Irx-3, lrx-4, ISGF-1, ISGF-3, ISGF3alpha, ISGF-
 10 3gamma, Isl-1, ITF, ITF-1, ITF-2, JRF, Jun, JunB, JunD, kappay factor, KBP-1, KER1, KER-1,
 Kox1, KRF-1, Ku autoantigen, KUP, LBP-1, LBP-1a, LBX1, LCR-F1, LEF-1, LEF-1B, LF-A1,
 LHX1, LHX2, LHX3a, LHX3b, LHX5, LHX6.1a, LHX6.1b, LIT-1, Lmo1, Lmo2, LMX1A,
 LMX1B, L-My1 (long form), L-My1 (short form), L-My2, LSF, LXRalpha, LyF-1, LyI-1, M
 factor, Mad1, MASH-1, Max1, Max2, MAZ, MAZ1, MB67, MBF1, MBF2, MBF3, MBP-1 (1),
 15 MBP-1 (2), MBP-2, MDBP, MEF-2, MEF-2B, MEF-2C (433 AA form), MEF-2C (465 AA
 form), MEF-2C (473 M form), MEF-2C/delta32 (441 AA form), MEF-2D00, MEF-2D0B,
 MEF-2DA0, MEF-2DA'0, MEF-2DAB, MEF-2DA'B, Meis-1, Meis-2a, Meis-2b, Meis-2c,
 Meis-2d, Meis-2e, Meis3, Meox1, Meox1a, Meox2, MHox (K-2), Mi, MIF-1, Miz-1, MM-1,
 MOP3, MR, Msx-1, Msx-2, MTB-Zf, MTF-1, mtTF1, Mxi1, Myb, Myc, Myc 1, Myf-3, Myf-4,
 20 Myf-5, Myf-6, MyoD, MZF-1, NC1, NC2, NCX, NELF, NER1, Net, NF III-a, NF III-c, NF III-
 e, NF-1, NF-1A, NF-1B, NF-IX, NF-4FA, NF-4FB, NF-4FC, NF-A, NF-AB, NFAT-1, NF-AT3,
 NF-Atc, NF-Atp, NF-Atx, NfbetaA, NF-CLE0a, NF-CLE0b, NFdeltaE3A, NFdeltaE3B,
 NFdeltaE3C, NFdeltaE4A, NFdeltaE4B, NFdeltaE4C, Nfe, NF-E, NF-E2, NF-E2 p45, NF-E3,
 NFE-6, NF-Gma, NF-GMb, NF-IL-2A, NF-IL-2B, NF-jun, NF-kappaB, NF-kappaB(-like), NF-
 25 kappaB1, NF-kappaB1, precursor, NF-kappaB2, NF-kappaB2 (p49), NF-kappaB2 precursor,
 NF-kappaE1, NF-kappaE2, NF-kappaE3, NF-MHCIIA, NF-MHCIIB, NF-muE1, NF-muE2,
 NF-muE3, NF-S, NF-X, NF-X1, NF-X2, NF-X3, NF-Xc, NF-YA, NF-Zc, NF-Zz, NHP-1,
 NHP-2, NHP3, NHP4, NKX2-5, NKX2B, NKX2C, NKX2G, NKX3A, NKX3A v1, NKX3A v2,
 NKX3A v3, NKX3A v4, NKX3B, NKX6A, Nmi, N-Myc, N-Oct-2alpha, N-Oct-2beta, N-Oct-3,
 30 N-Oct-4, N-Oct-5a, N-Oct-5b, NP-TCII, NR2E3, NR4A2, Nrf1, Nrf-1, Nrf2, NRF-2betal, NRF-
 2gamma1, NRL, NRSF form 1, NRSF form 2, NTF, O2, OCA-B, Oct-1, Oct-2, Oct-2.1, Oct-
 2B, Oct-2C, Oct-4A, Oct4B, Oct-5, Oct-6, Octa-factor, octamer-binding factor, oct-B2, oct-B3,
 Otx1, Otx2, OZF, p107, p130, p28 modulator, p300, p38erg, p45, p49erg, p53, p55, p55erg,

p65delta, p67, Pax-1, Pax-2, Pax-3, Pax-3A, Pax-3B, Pax-4, Pax-5, Pax-6, Pax-6/Pd-5a, Pax-7, Pax-8, Pax-8a, Pax-8b, Pax-8c, Pax-8d, Pax-8e, Pax-8f, Pax-9, Pbx-1a, Pbx-1b, Pbx-2, Pbx-3a, Pbx-3b, PC2, PC4, PC5, PEA3, PEBP2alpha, PEBP2beta, Pit-1, PITX1, PITX2, PITX3, PKNOX1, PLZF, PO-B, Pontin52, PPARalpha, PPARbeta, PPARgamma1, PPARgamma2, PPUR, PR, PR A, pRb, PRD1-BF1, PRDI-BFc, Prop-1, PSE1, P-TEFb, PTF, PTFalpha, PTFbeta, PTFdelta, PTFgamma, Pu box binding factor, Pu box binding factor (BJA-B), PU.1, PuF, Pur factor, R1, R2, RAR-alpha1, RAR-beta, RAR-beta2, RAR-gamma, RAR-gamma1, RBP60, RBP-Jkappa, Rel, RelA, RelB, RFX, RFX1, RFX2, RFX3, RFX5, RF-Y, RORalpha1, RORalpha2, RORalpha3, RORbeta, RORgamma, Rox, RPF1, RPGalpha, RREB-1, RSRFC4, RSRFC9, RVF, RXR-alpha, RXR-beta, SAP-1a, SAP1b, SF-1, SHOX2a, SHOX2b, SHOXa, SHOXb, SHP, SIII-p110, SIII-p15, SIII-p18, SIM1, Six-1, Six-2, Six-3, Six-4, Six-5, Six-6, SMAD-1, SMAD-2, SMAD-3, SMAD-4, SMAD-5, SOX-11, SOX-12, Sox-4, Sox-5, SOX-9, Sp1, Sp2, Sp3, Sp4, Sph factor, Spi-B, SPIN, SRCAP, SREBP-1a, SREBP-1b, SREBP-1c, SREBP-2, SRE-ZBP, SRF, SRY, SRP1, Staf-50, STAT1alpha, STAT1beta, STAT2, STAT3, STAT4, STAT6, T3R, T3R-alpha1, T3R-alpha2, T3R-beta, TAF(I)110, TAF(I)48, TAF(I)63, TAF(II)100, TAF(II)125, TAF(II)135, TAF(II)170, TAF(II)18, TAF(II)20, TAF(II)250, TAF(II)250Delta, TAF(II)28, TAF(II)30, TAF(II)31, TAF(II)55, TAF(II)70-alpha, TAF(II)70-beta, TAF(II)70-gamma, TAF-I, TAF-II, TAF-L, Tal-1, Tal-1beta, Tal-2, TAR factor, TBP, TBX1A, TBX1 B, TBX2, TBX4, TBX5 (long isoform), TBX5 (short isoform), TCF, TCF-1, TCF-1A, TCF-1B, TCF-1C, TCF-1D, TCF-1E, TCF-1F, TCF-1G, TCF-2alpha, TCF-3, TCF-4, TCF-4(K), TCF-4B, TCF-4E, TCFbeta1, TEF-1, TEF-2, tel, TFE3, TFEB, TFIIA, TFIIA-alpha/beta precursor, TFIIA-alpha/beta precursor, TFIIA-gamma, TFIIB, TFIID, TFIIE, TFIIE-alpha, TFIIE-beta, TFIIF, TFIIF-alpha, TFIIF-beta, TFIIH, TFIIH*, TFIIH-CAK, TFIIH-cyclin H, TFIIH-ERCC2/CAK, TFIIH-MAT1, TFIIH-MO15, TFIIH-p34, TFIIH-p44, TFIIH-p62, TFIIH-p80, TFIIH-p90, TFII-I, Tf-LF1, Tf-LF2, TGIF, TGIF2, TGT3, THRA1, TIF2, TLE1, TLX3, TMF, TR2, TR2-11, TR2-9, TR3, TR4, TRAP, TREB-1, TREB-2, TREB-3, TREF1, TREF2, TRF (2), TTF-1, TXRE BP, TxREF, UBF, UBP-1, UEF-1, UEF-2, UEF-3, UEF-4, USF1, USF2, USF2b, Vav, Vax-2, VDR, vHNF-1A, vHNF-1B, vHNF-1C, VITF, WSTF, WT1, WT1I, WT1 I-KTS, WT1 I-del2, WT1 -KTS, WT1-del2, X2BP, XBP-1, XW-V, XX, YAF2, YB-1, YEBP, YY1, ZEB, ZF1, ZF2, ZFX, ZHX1, ZIC2, ZID, ZNF174, amongst others.

An activated transcription factor is a transcription factor that has been activated by a stimulus resulting in a measurable change in the state of the transcription factor, for example a post-translational modification, such as phosphorylation, methylation, and the like. Activation

of a transcription factor can result in a change in the affinity for a particular DNA sequence or of a particular protein, such as another transcription factor and/or cofactor.

II. Detailed Description of Several Embodiments

a. Synthetic Oligonucleotides

Disclosed herein are synthetic oligonucleotides useful in the detection of DNA binding proteins. The oligonucleotides comprise (a) a sequence that includes a nucleic acid binding protein binding site, (b) a sequence that facilitates the amplification of the primer such as a binding site for a PCR primer and/or an RNA polymerase promoter, and (c) a tag sequence. In some examples, the nucleic acid binding protein binding site is 20 nucleotides in length. At least the nucleic acid binding protein binding site is double stranded and the entire synthetic oligonucleotide can be double stranded. In still further examples, the oligonucleotide comprises SEQ ID NO: 195, SEQ ID NO: 196, and/or SEQ ID NOs: 3-98.

The nucleic acid binding protein binding site may be any binding site for a protein that can bind a specific nucleic acid sequence. Such binding sites include known binding sites, putative binding sites, mutant or otherwise altered binding sites, or even random sets of nucleotides that are used to seek out unknown nucleic acid binding proteins. In addition, the nucleic acid binding site may comprise methylated or otherwise altered nucleic acids.

A tag sequence is an artificial nucleic acid sequence that does not bind any known nucleic acid binding protein. In some examples, a tag sequence is a non-naturally occurring nucleic acid sequence that does not bind any known nucleic acid binding protein. One of skill in the art in light of this disclosure will understand how to make a tag sequence or set of tag sequences that can be incorporated into the probes described herein. Additionally, one of skill in the art in light of this disclosure would be able to determine without undue experimentation whether or not a particular sequence binds to a nucleic acid binding protein and would therefore not use that sequence as a tag sequence. Examples of tag sequences include SEQ ID NOs: 3-98 herein.

In some examples the oligonucleotide probe is between 85 and 130 nucleotides in length, between, 88 and 120 nucleotides, between 90 and 110 nucleotides, between 88 and 98 nucleotides in length, between 90 and 96 nucleotides in length, between 92 and 94 nucleotides in length, including 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 100, 101, 102, 103, 104, 105, 106, 107, 018, 019, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, or 130 nucleotides in length. An exemplary set of the synthetic

oligonucleotides described herein includes SEQ ID NOs: 99-194 and SEQ ID NOs: 197-214 herein.

In some examples, oligonucleotides described herein contain one or more modifications. Modified oligonucleotides include those comprising modified backbones or non-natural internucleoside linkages. As defined herein, oligonucleotides having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone.

Examples of modified oligonucleotide backbones include, but are not limited to, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of the nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, each of which is herein incorporated by reference.

Examples of modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above oligonucleosides include, but are not limited to, U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289;

5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference.

Modified oligonucleotides can also contain one or more substituted sugar moieties. In some examples, the oligonucleotides can comprise one of the following at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C₁ to C₁₀ alkyl or C₂ to C₁₀ alkenyl and alkynyl. Oligonucleotides can also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar. Representative United States patents that teach the preparation of modified sugar structures include, but are not limited to, U.S. Patent Nos. 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; and 5,700,920, each of which is herein incorporated by reference in its entirety.

Oligonucleotides can also include base modifications or substitutions. As used herein, “unmodified” or “natural” bases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified bases include other synthetic and natural bases, such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azouracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified bases have been described (see, for example, U.S. Patent No. 3,687,808; and Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993).

Certain of these modified bases are useful for increasing the binding affinity. For example, 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2°C. Representative U.S. patents that teach the preparation of modified bases include, but are not limited to, U.S. Patent Nos. 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540;

5,587,469; 5,594,121, 5,596,091; 5,614,617; 5,681,941; and 5,750,692, each of which is herein incorporated by reference.

Also disclosed is a synthetic oligonucleotide library that includes a plurality of the synthetic oligonucleotides described herein. The library can contain at least, 2, 16, 48, 96, 384,
5 1000, 10,000, or 100,000 or more different probes. Each synthetic oligonucleotide in the library can have a different nucleic acid binding protein binding site. Each probe can comprise a different nucleic acid protein binding site than all the other probes in the library. Additionally, each synthetic oligonucleotide in the library can comprise a different tag sequence than all the other tag sequences with each tag sequence in the library coupled with a particular nucleic acid
10 protein binding site.

b. Oligonucleotide synthesis

The synthetic oligonucleotides disclosed herein may be synthesized by any method now known in the art or yet to be disclosed. Oligonucleotide synthesis may be carried out by the addition of nucleotide residues to the 5'-terminus of a growing chain. Elements of
15 oligonucleotide synthesis include: de-blocking (detritylation): A DMT group is removed with a solution of an acid, such as TCA or Dichloroacetic acid (DCA), in an inert solvent (dichloromethane or toluene) and washed out, resulting in a free 5' hydroxyl group on the first base. Coupling: A nucleoside phosphoramidite (or a mixture of several phosphoramidites) is activated by an acidic azole catalyst, tetrazole, 2-ethylthiotetrazole, 2-benzylthiotetrazole, 4,5-
20 dicyanoimidazole, or a number of similar compounds. This mixture is brought in contact with the starting solid support (first coupling) or oligonucleotide precursor (following couplings) whose 5'-hydroxy group reacts with the activated phosphoramidite moiety of the incoming nucleoside phosphoramidite to form a phosphite triester linkage. The phosphoramidite coupling may be carried out in anhydrous acetonitrile. Unbound reagents and by-products may be
25 removed by washing.

A small percentage of the solid support-bound 5'-OH groups (0.1 to 1%) remain unreacted and should be permanently blocked from further chain elongation to prevent the formation of oligonucleotides with an internal base deletion commonly referred to as (n-1) shortmers. This is done by acetylation of the unreacted 5'-hydroxy groups using a mixture of
30 acetic anhydride and 1-methylimidazole as a catalyst. Excess reagents are removed by washing.

The newly formed tricoordinated phosphite triester linkage is of limited stability under the conditions of oligonucleotide synthesis. The treatment of the support-bound material with iodine and water in the presence of a weak base (pyridine, lutidine, or collidine) oxidizes the

phosphite triester into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleosidic linkage. This step can be substituted with a sulfurization step to obtain oligonucleotide phosphorothioates. In the latter case, the sulfurization step is carried out prior to capping. Upon the completion of the chain assembly, the product may be released from the solid phase to solution, deprotected, and collected. Products may be isolated by HPLC to obtain the desired oligonucleotides in high purity.

c. Oligonucleotide Labeling

The hybridized synthetic oligonucleotides can be detected by detecting one or more labels bonded to the sample nucleic acids. The labels can be incorporated by any of a number of methods. In one example, the label is simultaneously incorporated during nucleic acid amplification. Thus, for example, polymerase chain reaction (PCR) with labeled primers or labeled nucleotides will provide a labeled amplification product. Alternatively, transcription amplification using an RNA polymerase and a labeled nucleotide (such as fluorescein-labeled UTP and/or CTP) can be used to incorporate a label into the transcribed nucleic acids.

Alternatively, a label may be added directly to the original nucleic acid sample (such as mRNA, polyA mRNA, cDNA, etc.) or to the amplification product after the amplification is completed. Methods of attaching labels to nucleic acids are well known to those of skill in the art and include, for example, nick translation or end-labeling (e.g. with a labeled RNA) by phosphorylation of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore).

Detectable labels suitable for use include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical, chemical, or other detection systems. Useful labels include biotin for staining with labeled streptavidin conjugate, magnetic beads (for example DYNABEADSTM), fluorescent dyes (for example, fluorescein, Texas red, rhodamine, green fluorescent protein, and the like), radiolabels (for example, ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (for example, horseradish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (for example, polystyrene, polypropylene, latex, etc.) beads. Patents teaching the use of such labels include U.S. Patent No. 3,817,837; U.S. Patent No. 3,850,752; U.S. Patent No. 3,939,350; U.S. Patent No. 3,996,345; U.S. Patent No. 4,277,437; U.S. Patent No. 4,275,149; and U.S. Patent No. 4,366,241.

Methods of detecting such labels are also well known. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be

detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction product produced by the action of the enzyme on the substrate, and fluorescent and colorimetric labels are detected by visualizing the colored label.

5 The label may be added to the synthetic oligonucleotide prior to, or after amplification and/or hybridization to the array. Some detectable labels are directly attached to or incorporated into the synthetic oligonucleotide prior to hybridization. Other labels are joined to the synthetic oligonucleotide after hybridization to the array. Often, the indirect label is attached to a binding moiety that has been attached to the target nucleic acid prior to the hybridization. Thus, for
10 example, the target nucleic acid can be biotinylated before the hybridization. After hybridization, an avidin-conjugated fluorophore can bind the biotin hybrid duplexes comprising the biotin, thereby providing a label that is easily detected (see Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes, P. Tijssen, ed. Elsevier, N.Y., 1993).

15 *d. Kits*

Also disclosed is a kit comprising the synthetic oligonucleotide library that comprises the synthetic oligonucleotides disclosed herein and an array that comprises at least 2, 16, 48, 96, 384, 1000, 10,000 or 100,000 bound synthetic oligonucleotides with the bound synthetic oligonucleotides bound to a solid substrate. Each bound synthetic oligonucleotide comprises a
20 sequence that is complementary (binds via Watson-Crick base pairing - A-T, C-G) to the tag sequence on one of the synthetic oligonucleotides in the library. The bound synthetic oligonucleotides are addressable within the context of the microarray such that a known synthetic oligonucleotide can bind to a bound oligonucleotide comprising the complement to its tag sequence in a known location on the array. Detection of the binding of the synthetic
25 oligonucleotide to an addressed position on the array further can identify the nucleic acid binding protein binding site and, in turn indicate that the nucleic acid binding protein was originally present in a sample. Examples of complements to tag sequences in the bound synthetic oligonucleotides are the complementary sequences of SEQ ID NOs: 3-98.

e. Methods of Use

30 Also disclosed is a method of detecting the presence of a DNA binding protein in a sample. This method comprises, contacting a library of synthetic oligonucleotides described herein with a sample that may contain a nucleic acid binding protein. The synthetic oligonucleotides are allowed to bind to the nucleic acid binding protein thereby forming

complexes. The complexes are purified by electrophoresis and a label is incorporated into the synthetic oligonucleotide. The synthetic oligonucleotide is hybridized to an array comprising bound synthetic oligonucleotides as described above and the presence of the label is detected on the array. Detection of the label on a specific address on the array is an indication that a nucleic acid binding protein that can specifically bind the nucleic acid binding site on the synthetic oligonucleotide that is addressed to the specific address on the array via its tag sequence was originally present in the sample.

The label may be any label including a fluorescent label. Incorporation of the label may be by any method including any method that amplifies or transcribes the nucleic acid such as by performing PCR amplification or contacting the complex with an RNA polymerase. The electrophoresis may be any type of electrophoresis including electrophoresis performed using the method and device described in US Patent Application Publication number 2012/0160683 (which is incorporated by reference herein in its entirety.)

For example, the disclosed oligonucleotides may be used with electrophoresis which utilizes a cassette comprising a proximal end and a distal end opposite the proximal end, a front face and a back face substantially parallel to the front face, and at least one chamber defined between the front face and the back face, the chamber having an upper portion and a lower portion, wherein the upper portion includes an upper opening at or near the proximal end of the cassette and the lower portion includes a lower opening at or near the distal end of the cassette, and wherein the front face comprises at least one window opening into the upper portion of the chamber. In some embodiments, the cassette can further comprise a semi-permeable membrane covering at least a portion of the at least one window opening. The semi-permeable membrane can be removably secured to the front face of the cassette. The semi-permeable membrane can be configured to allow passage of a buffer solution surrounding at least a portion of the cassette, and to at least partially block passage of a biomolecule contained within the upper portion of the at least one chamber. In some embodiments, a semi-permeable membrane can substantially block passage of a biomolecule of interest, thus substantially preventing it from exiting the chamber through the window opening. Some embodiments may utilize a cassette comprising acrylic. Some embodiments of a cassette comprise a unitary body. Cassettes can be integrated with one or more other components to form an electrophoresis device.

One embodiment of an electrophoresis device which can be utilized in the disclosed methods with the disclosed oligonucleotides comprises a cassette, wherein the cassette comprises a proximal end and a distal end opposite the proximal end, a front face and a back

face substantially parallel to the front face, and at least one chamber defined between the front face and the back face, the chamber having an upper portion and a lower portion, wherein the upper portion includes an upper opening at or near the proximal end of the cassette and the lower portion includes a lower opening at or near the distal end of the cassette, and wherein the front face comprises at least one window opening into the upper portion of the chamber. The electrophoresis device can further comprise a first buffer solution in fluid contact with the lower opening of the lower portion, a second buffer solution in fluid contact with the window opening of the upper portion, and at least one electrode electrically coupled to each of the first and second buffer solutions. Fluid contact can include both direct and indirect fluid contact. For example, fluid contact can include fluid contact through a membrane, such as a semi-permeable membrane.

In some embodiments of an electrophoresis device, the cassette further comprises a semi-permeable membrane covering at least a portion of the window opening of the upper portion, such as a membrane comprising cellulose or cellophane. In one particular embodiment, the electrophoresis device, comprises a first buffer container; and a cassette capable of being coupled to the first buffer container to form a first sidewall of the first buffer container, wherein the cassette comprises a proximal end and a distal end opposite the proximal end, a front face and a back face and at least one chamber defined between the front face and the back face, the chamber having an upper portion and a lower portion, wherein the upper portion includes a proximal opening at or near the proximal end of the cassette and the lower portion includes a distal opening at or near the distal end of the cassette wherein the proximal opening and distal opening are aligned to form a straight passageway there through, and wherein the front face comprises at least one window opening into the upper portion of the chamber.

The disclosed oligonucleotides can be used with reversible current electrophoresis, such that when current is run in a first direction, free molecules elute into the buffer at the distal end of the device. Then the cassette body can be placed into a new bath, the current reversed, and the biomolecules of interest can then move in the opposite direction, out of the gel, and back into the buffer solution at the proximal end of the cassette body, where they can then be easily collected from the device.

For example, one method of separating a biomolecule of interest from a sample comprises providing a sample, wherein the sample contains a biomolecule of interest and free probes, loading the sample onto a gel in a lower portion of a cassette body, electrophoresing the sample by applying a current for a time period sufficient for substantially all of the free probes

(such as one or more of the disclosed oligonucleotides) to elute out of a distal end of the lower portion, into a first buffer solution in a first buffer container, leaving the biomolecule of interest within the gel, removing the first buffer solution, providing a new buffer solution in fluid contact with the lower portion of the cassette, reversing the current with respect to the gel, and
5 electrophoresing the sample for a time period sufficient for substantially all of the biomolecule of interest to elute out of a proximal end of the lower portion, into an upper portion of the cassette body.

Disclosed methods can further comprise collecting biomolecule of interest. Collecting the biomolecule of interest can comprise withdrawing a volume of buffer solution containing the
10 biomolecule of interest from the upper portion of the cassette body. Collecting the biomolecule of interest can comprise securing a semi-permeable membrane over at least part of the upper portion, so as to prevent substantially all of the biomolecule of interest from passing through the semi-permeable membrane into a second buffer solution in fluid contact with the upper portion of the cassette.

15 In some methods, reversing the current with respect to the gel comprises switching the polarity of a first and second electrode in electrical contact with the first and second buffer solutions, respectively. In some methods, reversing the current with respect to the gel comprises changing the orientation of the gel with respect to a first and second electrode in electrical contact with the first and second buffer solutions, respectively.

20 Removing the first buffer solution can comprise draining the first buffer solution from a buffer container. Providing a new buffer solution in fluid contact with the lower portion of the cassette can comprise refilling the buffer container with the new buffer solution. In other methods, removing the first buffer solution comprises removing the cassette from a first buffer container containing the first buffer solution, and providing a new buffer solution in fluid
25 contact with the lower portion of the cassette comprises placing the cassette in a second buffer container containing the new buffer solution.

EXAMPLES

The following examples are illustrative of disclosed methods. In light of this disclosure,
30 those of skill in the art will recognize that variations of these examples and other examples of the disclosed method would be possible without undue experimentation.

Example 1

Sample Preparation

Nuclear protein extracts were prepared according to the QIAGEN® protocol using a Qproteome Nuclear Protein Kit (Catalog number 37582, QIAGEN®, Valencia, CA) from 20 to 5 25 mg tissue or 10 million cells as appropriate. The nuclear protein extracts were aliquotted in 5 µg volumes and stored at -80 °C prior to the reaction.

A blocking reaction was then performed. The blocking reaction included Poly(dI-dC) a double stranded alternating copolymer that is commonly used as a blocking agent in EMSA assays at a concentration of 1 µg/ml per 1 µg/ml of nuclear protein as well as a mixture of 10 polynucleotides consisting of equal proportions of SEQ ID NOs 3-98. This mixture was added at 0.4 µg/ml per 1 µg/ml of nuclear protein. Exact concentrations and volumes used are as follows:

	Nuclear protein [1.0µg/µl]	5 µl
	Poly(IC) [0.5µg/µl]	2 µl
	Probe mix [200ng/µl]	2 µl
15	Add water to	15 µl

The blocking reaction lasted for 15 minutes at room temperature.

The blocked reaction mixture was then mixed with a set of oligonucleotide probes consisting of polynucleotides of SEQ ID NOs: 197-214 in the following mixture:

	10x buffer	3.0 µl
20	50% glycerol	1.5 µl
	25 mM DTT/2.5% Tween®-20	3.0 µl
	Probe set (38 ng/µl)	1.5 µl
	Water	6.0 µl

This mixture was incubated for 25 minutes at room temperature.

25 Probes were then run on a gel electrophoresis system configured to separate and collect complexes comprising transcription factor and bound probe from free probe. This system is described in US Patent Application 2012/0160683, which is incorporated by reference herein.

A 6% acrylamide gel was prepared as follows:

	30% acrylamide	4 ml
30	7.5x TBE	1.33 ml
	Water	14.6 ml
	10% APS	100 µl
	TEMED	10 µl

The running buffer was a 0.5x Tris/borate/EDTA (TBE) buffer.

The gel was run for 135 minutes, a time previously determined to be sufficient to clear all unbound probe out of the distal end of the gel. The gel cassette was rinsed, a semipermeable membrane installed in the collection chamber as described in the referenced patent application, and fresh TBE buffer added. At this point, the electrode polarity is reversed and the gel run with electrode polarity reversed for 210 minutes. After the 210 minute electrophoresis, the transcription factor protein/probe complexes are collected in about a 1.5 ml volume. The 1.5 ml volume is concentrated to 50 μ l using an Amicon 10k filter unit and the samples then purified using a QIAamp nucleotide removal kit. The volume of the eluate with the purified probes is 50 μ l.

Amplification and labeling was performed using a T7 High Yield RNA® Synthesis Kit (New England Biolabs). Biotinylated UTP is added to the ribonucleotide solution from the kit to a level where the biotinylated UTP makes up 33% of the UTP in the solution. The amplification and labeling mixture is as follows:

15	10x buffer (from kit)	2 μ l
	Ribonucleotide mix	7.5 μ l
	(100mM each ATP, CTP, GTP, 50mM UTP and 25 mM biotinylated UTP)	
	Purified probes from electrophoresis	6.0 μ l
	T7 RNA polymerase (from kit)	2.0 μ l
20	Water	2.5 μ l

Amplification/labeling was performed for 2 hours at 37 °C, following which the samples are placed on ice.

Labeled RNA was purified using a QIAGEN® RNeasy Mini-elute Cleanup kit.

Microarray slides were prepared by Microarrays Inc (Huntsville, AL). Four subarrays of 25 nucleotide-long tag sequences were printed per slide. Complements of ninety-six tag sequences (SEQ ID NOs: 3-98) were used for each subarray. Each tag sequence was printed 4 times – once in each each subarray.

Labeled RNA was mixed with hybridization buffers A and B (Hyb & Frag Buffers, CodeLink®, Applied Microarrays) and hybridized to each microarray under the following conditions:

Samples were heated at 68°C for 3 minutes and then placed on ice for 2 minutes. Then 55 μ l of sample was injected into a Tecan 400 pro hybridization station at 37 °C for 18 hours.

Samples were washed with 0.75 TNT fast wash at room temperature, 2 times, then incubated in 0.75 TNT at 37°C for 1 hour.

A Cy5-Streptavidin working solution was prepared by mixing 1µl Cy5 with 499µl TNB buffer. A 55 µl volume of Cy5-Streptavidin working solution was injected into the sample through a positive pressure pipette and incubate at room temperature for 30 min. Slides were then washed with 1 x TNT wash 4 times at 5 minutes per wash. Slides were then washed with 0.1× SSC/ 0.05% Tween 20 at room temperature for 30 seconds. Slides were then dried by nitrogen flow.

A GenePix 4000B Microarray scanner was used to scan the slide using a single wavelength scan at 635 nm. The PMT gain ranged from 450-650.

Normalization of signal was performed as follows:

Average signal of the subarray = total signal of the subarray/(96x4)

Average signal of the probe = total signal of the probe/4

Relative strength of a probe = Average signal of the probe/ Average signal of the subarray

Example 2

Quantifiable Detection of NF-κB in liver cells

Liver nuclear extract was used as the control (group 1), 10 ng (group 2), 50 ng (group 3) and 100 ng (group 4) of purified NF-κb protein were added to same amount of liver nuclear extract and then bind to a group of mixed probes, which included NF-κb probes, ER-alpha probes, SP-1 probes, mutated NF-κb probes, mutated ER-alpha probes, and mutated SP-1 probes. The probes that bind to nuclear protein in the samples were separated with our gel device and then labeled and hybridized to individual subarray of microarray slide. The relative strength of one probes were calculated as:

Relative strength = average signal of one probe/average signal of the subarray

Results are shown in FIG. 1. Relative strength is calculated on the Y axis. The relative strength of three NF-κb probes (SEQ ID NO: 197, SEQ ID NO: 198, and SEQ ID NO: 99) increased in relation in nuclear extracts with added NF-κb. A mutant probe for NF-κB (SEQ ID NO: 202) did not increase and did decrease in relative strength with the addition of NF-κB.

In view of the many possible embodiments to which the principles of the disclosed invention may be applied, it should be recognized that the illustrated embodiments are only

preferred examples of the invention and should not be taken as limiting the scope of the invention. Rather, the scope of the invention is defined by the following claims. We therefore claim as our invention all that comes within the scope and spirit of these claims.

We claim:

1. A synthetic oligonucleotide comprising:
a first sequence comprising a nucleic acid binding protein binding site;
5 a second sequence comprising a binding site for a first PCR primer or an RNA
polymerase promoter;
a third sequence comprising a tag sequence, wherein the tag sequence binds no known
nucleic acid binding protein;
wherein at least the first sequence is double stranded and wherein the total length of the
10 oligonucleotide is between 85 and 130 nucleotides.
2. The synthetic oligonucleotide of claim 1, wherein the second sequence is SEQ ID NO:
195 or SEQ ID NO: 196.
- 15 3. The synthetic oligonucleotide of claim 2, further comprising a fourth sequence, the
fourth sequence comprising a binding site for a second PCR primer.
4. The synthetic oligonucleotide of claim 1, wherein the entire oligonucleotide is double-
stranded.
20
5. The synthetic oligonucleotide of claim 1, comprising one or more copies of a single
nucleic acid binding protein binding site.
6. The synthetic oligonucleotide of claim 1, wherein the oligonucleotide is between 90 and
25 98 nucleotides in length.
7. The synthetic oligonucleotide of claim 1, comprising SEQ ID NO: 1 and/or SEQ ID NO:
2.
- 30 8. The synthetic oligonucleotide of claim 1, wherein the tag sequence is 25 nucleotides in
length.
9. The synthetic oligonucleotide of claim 8, comprising any of SEQ ID NOs: 3-98.

10. A method of detecting the presence of a nucleic acid binding protein in a sample, the method comprising:

contacting a synthetic oligonucleotide of any of claims 1-8 with a sample, the sample
5 comprising a nucleic acid binding protein;

subjecting the sample to conditions that allow binding of the DNA binding protein to the synthetic oligonucleotide to form a nucleic acid binding protein/oligonucleotide complex;

purifying the nucleic acid binding protein/oligonucleotide complex by electrophoresis;

denaturing the nucleic acid binding protein/oligonucleotide complex;

10 incorporating a label into the synthetic oligonucleotide;

hybridizing the synthetic oligonucleotide to an array, wherein the array comprises a bound oligonucleotide, the bound oligonucleotide comprising the complement of the tag sequence and

detecting the presence of the label;

15 wherein the hybridizing the synthetic oligonucleotide to the array is performed after the purifying the nucleic acid binding protein/oligonucleotide complex by electrophoresis.

11. The method of claim 10, wherein the label is a fluorescent label.

20 12. The method of claim 10, wherein incorporating the label into the synthetic oligonucleotide comprises performing PCR amplification with primers complementary to a universal PCR forward primer sequence and a universal PCR reverse primer sequence.

13. The method of claim 10, wherein incorporating the label into the synthetic
25 oligonucleotide comprises contacting the synthetic oligonucleotide with an RNA polymerase.

14. The method of claim 10, wherein the electrophoresis is performed using an electrophoresis device comprising:

a cassette, wherein the cassette comprises a proximal end and a distal end opposite the
30 proximal end, a front face and a back face and at least one chamber defined between the front face and the back face, the chamber having an upper portion and a lower portion, wherein the upper portion includes an upper opening at or near the proximal end of the cassette and the

lower portion includes a lower opening at or near the distal end of the cassette, and wherein the front face comprises at least one window opening into the upper portion of the chamber;

a first buffer solution in fluid contact with the lower opening of the lower portion;

a second buffer solution in fluid contact with the window opening of the upper portion;

5 and

at least one electrode electrically coupled to each of the first and second buffer solutions.

15. A synthetic oligonucleotide library comprising:

10 a first synthetic oligonucleotide of any of claims 1-8, the first synthetic oligonucleotide comprising a first nucleic acid binding protein binding site and a first tag sequence; and

a second synthetic oligonucleotide of any of claims 1-8, the second synthetic oligonucleotide comprising a second nucleic acid binding protein binding site and a second tag sequence;

15 wherein the first nucleic acid binding protein binding site and the second nucleic acid binding protein binding site are different sequences.

16. The synthetic oligonucleotide library of claim 15, wherein the first tag sequence is different from the second tag sequence.

20 17. The synthetic oligonucleotide library of claim 15, comprising two or more of the probes of SEQ ID NOs: 99-195 or SEQ ID NOs: 197-214.

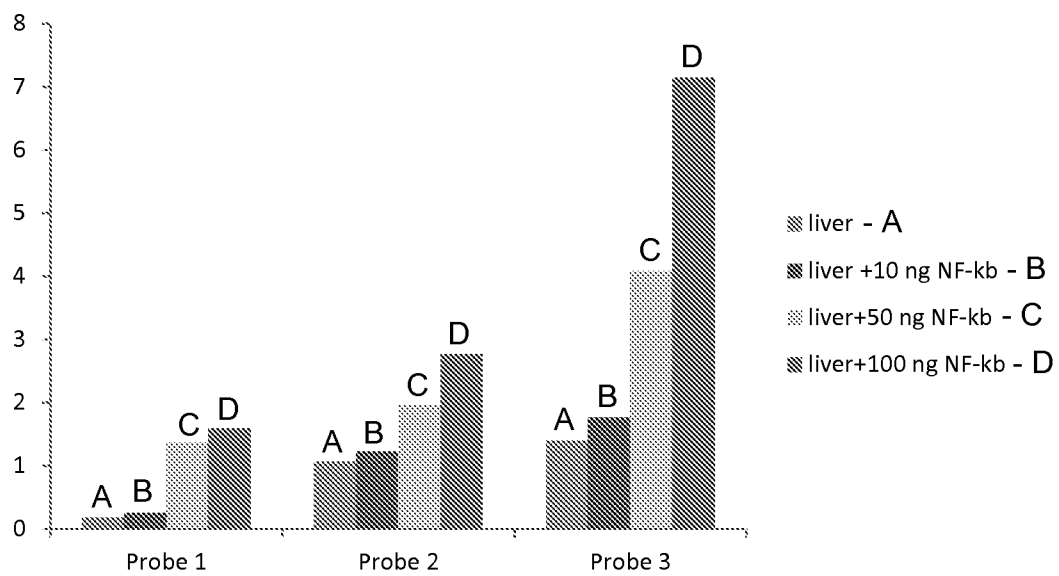
18. A kit comprising:

the synthetic oligonucleotide library of any of claims 14-16; and

25 an array, the array comprising a third synthetic oligonucleotide comprising the complement of the first tag sequence and a fourth synthetic oligonucleotide comprising the complement of the second tag sequence; wherein the third synthetic oligonucleotide and the fourth synthetic oligonucleotide are each addressable on the microarray.

30 19. The kit of claim 18 wherein the third synthetic oligonucleotide and the fourth synthetic oligonucleotide comprise the complements of any of SEQ ID NOs: 3-98.

FIG. 1



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2014/053183**A. CLASSIFICATION OF SUBJECT MATTER****C12N 15/11(2006.01)i, C12Q 1/68(2006.01)i**

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N 15/11; C12N 15/63; C12P 19/34; C12Q 1/68; C12M 1/34

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS(KIPO internal) & keywords: nucleic acid binding protein, primer, RNA polymerase promoter, tag, binding site

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2008-147899 A1 (OREGON HEALTH & SCIENCE UNIVERSITY) 4 December 2008 See abstract; claims 1-29 and 34-46; figures 1A-4C.	1-17
A	US 2003-0215854 A1 (CLAUSEN, P. A. et al.) 20 November 2003 See abstract; claims 1-20; figures 1-6.	1-17
A	US 2003-0104368 A1 (ZHOU, K.) 5 June 2003 See abstract; claims 1-32.	1-17
A	GUIEYSSE, A. L. et al., 'Identification of a triplex DNA-binding protein from human cells', J. Mol. Biol., 28 March 1997, Vol. 267, No. 2, pp. 289-298. See abstract; table 1; figure 1.	1-17
A	US 6066452 A (WEISSMAN, S. M. et al.) 23 May 2000 See abstract; claims 1-14.	1-17



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

15 December 2014 (15.12.2014)

Date of mailing of the international search report

15 December 2014 (15.12.2014)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

HEO, Joo Hyung

Telephone No. +82-42-481-8150



INTERNATIONAL SEARCH REPORTInternational application No.
PCT/US2014/053183**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 19
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
Claim 19 is unclear since it is referring to the multiple dependent claims which do not comply with PCT Rule 6.4(a).
3. ☒ Claims Nos.: 18
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of any additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/053183

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008-147899 A1	04/12/2008	AU 2008-256851 A1 CA 2687804 A1 EP 2162548 A1 IL 202264 D0 US 2010-0184614 A1	04/12/2008 04/12/2008 17/03/2010 16/06/2010 22/07/2010
US 2003-0215854 A1	20/11/2003	AT 377753 T AU 2003-226116 A1 AU 2003-226116 B2 CA 2480646 A1 CA 2480646 C DE 60317315 D1 DE 60317315 T2 DK 1490682 T3 EP 1490682 A1 EP 1490682 A4 EP 1490682 B1 JP 04425640 B2 JP 2005-521411 A US 2009-0208927 A1 US 7122317 B2 US 7771940 B2 WO 03-083476 A1	15/11/2007 13/10/2003 01/04/2010 09/10/2003 03/08/2010 20/12/2007 21/08/2008 25/03/2008 29/12/2004 24/05/2006 07/11/2007 03/03/2010 21/07/2005 20/08/2009 17/10/2006 10/08/2010 09/10/2003
US 2003-0104368 A1	05/06/2003	None	
US 06066452 A	23/05/2000	None	