



(86) **Date de dépôt PCT/PCT Filing Date:** 2002/12/17
(87) **Date publication PCT/PCT Publication Date:** 2003/06/26
(45) **Date de délivrance/Issue Date:** 2015/10/27
(85) **Entrée phase nationale/National Entry:** 2004/06/17
(86) **N° demande PCT/PCT Application No.:** CA 2002/001941
(87) **N° publication PCT/PCT Publication No.:** 2003/052099
(30) **Priorité/Priority:** 2001/12/17 (US60/340,009)

(51) **Cl.Int./Int.Cl.** **C40B 40/06** (2006.01),
C07H 21/00 (2006.01), **C07H 21/04** (2006.01),
C12N 15/10 (2006.01), **C12N 15/11** (2006.01),
C12P 19/34 (2006.01), **C12Q 1/68** (2006.01),
C40B 30/00 (2006.01), **C40B 40/08** (2006.01),
C40B 40/10 (2006.01), **C40B 50/06** (2006.01),
G01N 33/53 (2006.01), **G01N 33/68** (2006.01)

(72) **Inventeurs/Inventors:**
CHEN, TAO, CA;
LI, JINGHAN, CA;
CHEN, TE-MING (DECEASED), ZZ

(73) **Propriétaires/Owners:**
CHEN, TAO, CA; ...

(54) **Titre : APPROCHE AXEE SUR LA BIOLOGIE DES SYSTEMES : PLATEFORMES DE CRIBLAGE A HAUT RENDEMENT, A
DIMENSIONS MULTIPLES**
(54) **Title: SYSTEM BIOLOGY APPROACH: HIGH THROUGHPUT SCREENING (HTS) PLATFORMS WITH MULTIPLE DIMENSIONS**

(57) **Abrésumé/Abstract:**

A systems biology approach of multiple-dimensional High Through-put Screening platforms for identification of both known and unknown sequences. Those platforms could be used alone or and in combinations for different purposes. The probe designs are codon-based with distinguishable orientations. The identification are mainly focused on 5'UTR, ORF, 3'UTR sequences and their corresponding transcriptional and translational products. For example, sequences originated from start codon, stop codon and two consecutive codons recognized by a restriction enzyme. The targeting sequences contain genetic variation(s), such as SNPs. Genetic algorithms for probe design and library construction were provided as well. Oligonucleotides are further classified according to their GC content, which provide uniformed conditions for hybridization. Other applications include the processes and performance of PCR, cloning, dot-blot hybridization, ELISA, sequence determination, signature sequence and signature signal identification, library construction, DNA and protein microarrays constructions.



(73) **Propriétaires(suite)/Owners(continued):**LI, JINGHAN, CA

(74) **Agent:** NA

ABSTRACT

A systems biology approach of multiple-dimensional High Through-put Screening platforms for identification of both known and unknown sequences. Those platforms could be used alone or and in combinations for different purposes. The probe designs are codon-based with distinguishable orientations. The identification are mainly focused on 5'UTR, ORF, 3'UTR sequences and their corresponding transcriptional and translational products. For example, sequences originated from start codon, stop codon and two consecutive codons recognized by a restriction enzyme. The targeting sequences contain genetic variation(s), such as SNPs. Genetic algorithms for probe design and library construction were provided as well. Oligonucleotides are further classified according to their GC content, which provide uniformed conditions for hybridization. Other applications include the processes and performance of PCR, cloning, dot-blot hybridization, ELISA, sequence determination, signature sequence and signature signal identification, library construction, DNA and protein microarrays constructions.

SYSTEMS BIOLOGY APPROACH: HIGH THROUGHPUT SCREENING (HTS) PLATFORMS WITH MULTIPLE DIMENSIONS

Pursuant to Section 28.4 of the Patent Act, this application claims priority to U.S. Provisional Patent Application Ser. No. 60/340,009 filed on December 17, 2001.

COPYRIGHT NOTICE

The applicants notify that a portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction of the patent disclosure, as it appears in Canadian Intellectual Property Office patent file or records, but otherwise reserves all copyright rights whatsoever.

BACKGROUND OF THE INVENTION

While nucleic acids consist of four nucleotides with four distinct bases: Adenine (A), Thymine (T) / Uracil (U), Guanine (G) and Cytosine (C) respectively, the coding sequences of genes are organized in codons that in turn code for specific amino acids. Codons are arranged in an oriented, sequential, and linear manner with a unique starting point and end point.

The codons (genetic code) consist of 64 nucleotide triplets: 61 codons encode the 20 essential amino acids; three codons are stop codons. Although there are some exceptions wherein 5'-GTG, 5'-ATA, 5'-TTG, 5'-ACG and 5'-CTG may function as start codons (5'-ATA is the start codon of mammalian mitochondria), 5'-ATG / (5'-AUG) is the prominent start codon. Interestingly, the situation is similar for stop codons: Except for some exceptions, such as 5'-AGA and 5'-AGG, which are stop codons instead of coding for arginine in mammalian mitochondrial, there are three dominant stop codons: 5'-TAA(5'-UAA), 5'-TGA(5'-UGA) and 5'-TAG(5'-UAG).

Although a specific coding region consists of a specific combination of a set of specific codons at a specific length, any sequence of Open Reading Frame (ORF) of a given gene of a given length could be identified among a group of linear DNA sequences, which consist of all possible combinations of 61 codons with a start codon at its 5'- and / or a stop codon at its 3'-end with the same length as a given ORF. Thus, any and all sequences of ORF of any given length could be deduced from either its 5'-end or 3'-end in accordance with the genetic algorithm of 61.sup. (n-1) under the conditions: n=1, n>1, n<infinity. n is not equal to zero. n is an integer. n represents any codon. n is the unit of measurement of the length of oligonucleotide. One start codon of 5'-towards 3' orientation is at 5'-end or one stop codon of 5'-towards 3' orientation is at 3'-end of each

oligonucleotide of n -codon length long. For example, if $n = 3$ and one 5'-ATG is at 5'-end, there are 3,721 distinct oligonucleotide sequences in 5'-orientation, which are three codons length long. Three codons equal nine nucleotides long.

Any translation of 5'-ORF of a given gene of a given length could be identified among a group of peptides of linear amino acids sequences, which consist of all possible combinations of 20 Essential L-Amino Acids (EAA) with a L-amino acid encoded by a start codon such as Methionine is at its N-terminal with the same unit number(s) of length as the corresponding 5'-ORF. Thus, any and all peptide sequences of any given length could be deduced from its N-terminal(s) in accordance with the genetic algorithm of $20.\sup.(n-1)$ under the conditions: $n=1$, $n>1$, $n< \text{infinity}$. n is not equal to zero. n is an integer. n represents L-Amino Acid. n is the unit of measurement of the length of peptide. One L-amino acid encoded by a start codon is at N-terminal of each peptide of n amino acids length long. For example, if $n = 4$ and one Methionine is at N-terminal, there are 8,000 distinct peptide sequences of N-terminal orientation that are four amino acids long.

Any translation of 3'-ORF of a given gene of a given length could be identified among a group of peptides of linear amino acids sequences, which consist of all possible combinations of 20 Essential L-Amino Acids (EAA) with any one of the 20 EAA is at its C-terminal with the same unit number(s) of the length as the corresponding 3'-ORF. Thus, any and all sequences of peptides of any given length could be deduced from its C-terminal(s) in accordance with the genetic algorithm of $20.\sup.(n)$ under the conditions: $n=1$, $n>1$, $n< \text{infinity}$. n is not equal to zero. n is an integer. n represents L-Amino Acid. n is the unit of measurement of the length of peptide. One of the 20 EAA is at its C-terminal of each peptide of n amino acids length long. For example, if $n = 4$ and one of the 20 EAA is at C-terminal, there are 160,000 distinct peptide sequences of C-terminal orientation that are four amino acids long.

Any 5'-Untranslated Region (5'-UTR) of a given gene of a given length could be identified among a group of linear DNA sequences, which consist of all possible combinations of 64 codons with a start codon adding at its 3'-end with the same length as the given 5'-UTR. Thus, any and all sequences of 5'-UTR of any given length could be deduced from its 3'-end including a start codon in accordance with the genetic algorithm of $64.\sup.(n-1)$ under the conditions: When $n=1$, the position of codon is $(1-n)+1$. When $n>1$ and $n< \text{infinity}$, the position of codon is $(1-n)$. n is not equal to zero. n is an integer. n represents codon. n is the unit of measurement of the length of oligonucleotide. One start codon of 5' towards 3' orientation is at 3'-end of each oligonucleotide of n codons length

long. The negative sign in front of n indicates that the codon position is at 5'-UTR. For example, if $n = 3$ and one 5'-ATG of 5' towards 3' orientation is at 3'-end, there are 4,096 distinct oligonucleotide sequences of 3'-orientation with three codons length long. The length of three codons equals the length of nine nucleotides.

Any given 3'-Untranslated Region (3'-UTR) of a given gene of a given length could be identified among a group of linear DNA sequences, which consist of all possible combinations of 64 codons with a stop codon adding at its 5'-end with the same length as the given 3'-UTR. Thus, any and all sequences of 3'-UTR of any given length could be deduced from its 5'-end including a stop codon in accordance with the genetic algorithm of $64 \cdot \sup(n-1)$ under the conditions: $n=1$, $n>1$, $n< \text{infinity}$. n is not equal to zero. n is an integer. n represents codon. n is the unit of measurement of the length of oligonucleotide. One stop codon of 5' towards 3' orientation is at 5'-end of each oligonucleotide of n codon length long. For example, if $n = 3$ and one 5'-TGA of 5' towards 3' orientation is at 5'-end, there are 4,096 distinct oligonucleotide sequences of 5'-orientation being three codons length long. The length of three codons equal nine nucleotides long.

Exceptions are existed. For example, 5'-TGA, which is usually codes for the termination of the synthesis of a peptide chain, sometimes codes for selenocysteine, an amino acid which is not among the 20 essential amino acids. Other exceptions such as 5'-AGA and 5'-ATA are not usable in *Micrococcus Luteus* while 5'-CGG is not usable in *Mycoplasmas* and *Spiroplasmas* (Kanoi et al., J. Mol. Bio. 230: 51-56, 1993) (Oba et al., Proc. Natl. Acad. Sci. U.S.A. 88: 921-925, 1991). Both 5'-TAA and 5'-TAG encode Glutamine in *Tetrahymena*, *Paramecium* and *Acetabularia* of Ciliates and Algae while 5'-CTG encodes Serine in *Candida cylindrica* of Fungi (Tourancheau et al., EMBO J. 14: 3262-3267, 1995).

There is often more than one 5'-ATG codon in a single gene. For example the full length of Glyceraldehyde-3-phosphate Dehydrogenase (GenBank Accession: NM_002046.2) has ten 5'-ATG codons in its ORF. Usually, the first suitable 5'-ATG is the start codon. Therefore, identifying every 5'-ATG / 5'-AUG of a given single gene will facilitate the identification of the start codon and the start site of ORF of the given gene.

It is of note that although point mutations, deletions, insertion and single nucleotide polymorphisms (SNPs) in the coding region are a change of nucleotide(s), these genetic variation(s) are actually a change of codon(s) or / and ORF(s). In view of this, it is clear that in comparison with the nucleotide based methods, the codon-based methodologies and tools are more effective and substantial. Particularly, the codon-based methods could

unify and consolidate the divergent nucleotide-based gene analytical methods dealing with diverse subjects such as SNPs, mutations, deletions and insertions.

In the reviewing of some current prevalent methods of gene cloning and analysis, considerable criticisms were cited in the art. For example, Subtractive Hybridization is a method of identifying and cloning cDNA of interest by hybridizing two different but relevant cDNA libraries. This method is not only time consuming and labour intensive, it is often impractical (Wieland et al., Proc. Natl. Acad. Sci. U.S.A. 87: 2729-2724); (Hubank et al., Nucleic Acids Res. 22: 5640-5648, 1994). Subtractive Hybridization is based on the assumption that the two different but relevant cDNA libraries were complete and fully representative of all expressed genes, but it does not provide any technical measurement or vehicle to ensure that the assumption is correct. Using two poorly defined cDNA libraries to deduce unknown genes would thus be both difficult in theory and practice.

The method of Differential Display Reverse Transcription Polymerase Chain Reaction (DD-RT-PCR) is based on the idea of using a set of arbitrary oligonucleotides as primers in PCR to display differentially expressed cDNA (Liang et al., Science 257: 967-971, 1992); (Pardee et al., U.S. Pat. No. 5, 262,311, and 1993); (Welsh et al., Nucleic Acids Res. 18: 7213-7218, 1990); (Welsh et al., Nucleic Acids Res. 20: 4965-4970, 1992). A set of arbitrarily designed and selected PCR primers might be a system of bias itself. The primers were selected based on single nucleotide or base instead of on codons. TGA, the termination codon, was identified among the designed PCR primers at the first reading frame, such as 5'-AAAAGCTTCTGAAGTTGTGGGGT (Liang et al., Science 257: 969, 1992) and OPA18 (AGGTGACCGT) (Liang et al., Nucleic Acids Res. 21: 3273, 1993). The results were often less than satisfying (Li et al., Nucleic Acids Res. 22: 1764-1765, 1994); (Bertioli et al., Nucleic Acids Res. 23: 4520-4523, 1995); (Shoham et al., Biotechniques 20: 182-183, 1996); (Graf et al., Nucleic Acids Res. 25: 2239-2240, 1997); (MacLeod et al., U.S. Pat. No.6, 221, 600, 1999). Obviously, DD-RT-PCR might provide partial cDNA fragments analysis at the best (Kinzler et al., U.S. Pat. No. 5,866,330, 1995). Another disadvantage is its adoption of polyacrylamide gel of electrophoresis as the display system as Welsh did before (Welsh et al., Nucleic Acids Res. 18: 7213-7218, 1990); (Welsh et al., Nucleic Acids Res. 20: 4965-4970, 1992). In most of the cases, the size of the cDNA fragments which could be efficiently identified and isolated from 6% polyacrylamide gel after electrophoresis is usually less than 500 to 600 base pairs. This set a technical limitation for displaying, identifying and cloning a full-length of cDNA molecules directly since the size of many genes are well beyond 1kb. Moreover, a

considerable portion of those displayed sequences of 500 to 600 base pairs might contain sequences of 3'-UTR.

Rapid Amplification of cDNA Ends (RACE) targets 5'-end of mRNA (Frohman et al., Proc. Natl. Acad. Sci. U.S.A. 85: 8998-9002, 1988); (Maruyama et al., Gene 138: 171-174, 1994). RACE uses Calf Intestinal Phosphatase (CIP) to remove 5'-end phosphates of uncapped mRNA molecules while leaving the 5'-capped mRNA intact. Subsequently, Tobacco Acid Pyrophosphatase (TAP) is added to the reaction to remove 5'-cap of the 5'-capped mRNA molecules. After the removing, 5'-phosphate of the uncapped mRNA is exposed to the environment. Then, the oligonucleotide designed as the PCR primer targeting sequence and T4 RNA ligase are added to the reaction. The 3'-hydroxyl group of the oligonucleotide will be ligated to the 5'-phosphate group of mRNA catalyzed by T4 RNA ligase. Thus, RACE eliminates uncapped mRNAs and selects the ones with a 5'-cap for further PCR aided cloning. The disadvantage is that mRNA molecules with full-length codon sequence but without a 5'-cap may be eliminated from the sample. It is of note that it is not unusual for several hundred base pair long sequence existing between 5'-cap and start codon of mRNA in vertebrates. Furthermore, the GC rich content of 5'-UTR suggests that a high degree of secondary structure may exist (Kozak, J. Cell Biol. 115: 887-903, 1991), which may have a negative impact on PCR priming from 5'-UTR of mRNA which is adjacent to 5'-cap. It is also noted that non-template nucleotides could be added to 3'-ends of cDNAs during RACE process (Chen et al., Biotechniq. 30: 574-582, 2001). Chen et al. recommended that RACE should be used carefully in determining the terminal sequences of nucleic acids.

Representational Difference Analysis (RDA) is another approach for cDNA cloning (Lisitsyn et al., Science 259: 946-951, 1993). RDA and DD-RT-PCR both rely on PCR to amplify the target fragments to a visible level on the gel. RDA is a process of subtraction of PCR amplicons from two different but relevant cDNA resources that eliminates the fragments present in both cDNA sources whereas DD-RT-PCR displays all the expressed cDNA fragments of both sources. Additionally, RDA requires restriction enzyme digestion and ligation. For a favourable enzymatic reaction, several influential factors, such as substrates, ΔG and the conformation of active sites, need to be determined, coordinated and optimized according to enzyme dynamic mechanisms. In general, more manipulation of cDNA may cause more artifacts and errors. The methods of Subtract Hybridization and RDA both assume their cDNA libraries used for subtractions are unbiased and fully representative. In fact, such assumption has neither been tested nor proven.

Serial Analysis of Gene Expression (SAGE) consists of a series of procedures including immobilizing cDNA on streptavidin beads prior to restriction enzyme digestion, ligating to nucleotide tags of 9 to 10 base pairs and ligating those tags into multiple tags within a single clone (Velculescu et al., Science 270: 484-487, 1995); (Kinzler et al., U.S. Pat. No. 5,866,330, 1995); (Zhang et al., Science 276:1270-1271, 1997); (MacLeod et al., U.S. Pat. No. 6,221,600, 1999). The amplified Restriction Fragment Length Polymorphism based mRNA fingerprinting had a higher percentage of reproducibility than DD-RT-PCR (Habu et al., Biochemical and Biophysical Research Communications 234: 519, 1997). However, SAGE has the drawback that its gene screening spectrum is limited. To detect gene expression, SAGE uses a method of cleaving a cDNA sample with a first restriction endonuclease which has four base pair (4-bp) recognition site (Kinzler et al., U.S. Pat. No. 5,866,330, 1999). It is known in the art that any one of the 4-bp recognition restriction sites could be identified among the pre-determined 4-bp sequence in accordance with the algorithm of $4^{sup.n}$. All possible combinatorial sequences of 4-bp sequence are 256. The number of the current available restriction endonucleases with 4-bp recognition sites is far less than 256. For example, only 20 restriction endonucleases with 4-bp recognition sites were identified in the 2002-03 Catalogue of New England BioLabs Inc. Accordingly, SAGE would have a full-range screening spectrum for the detection of gene expression until all of the 256 distinct restriction endonucleases with 4-bp recognition sites become available.

The expressed sequence tag (EST) method is another widely used tool for gene discovery (Adams et al., Science 252: 1651-1656, 1991). However, it evaluates only a limited number of genes at a time (Velculescu et al., Science 270: 484-487, 1995). The spectrum of sequences derived either from the SAGE or EST has the limitation.

To achieve the goal of full coverage and representation in gene identification, an oligonucleotide library constructed by all possible combinations of A.T.G.C. in accordance with the algorithm of $4^{sup.n}$ was proposed (Studier, Proc. Natl. Acad. Sci. U.S.A. 86: 6917-6921, 1989); (Szybalski, Gene 90: 177-178, 1990). In the algorithm of 4^n , n is the unit of measurement of the length of oligonucleotide. n represents nucleotide. Huse introduced the concept of random tuplets in the method of oligonucleotides synthesis. A tuple can be a dinucleotide, a trinucleotide or can also be four or more nucleotides (Huse, U.S. Pat. No. 5,523,388, 1996 and U.S. Pat. No. 5,808,022, 1998). To synthesize a diverse population of expressible oligonucleotides having a desirable bias of random codon sequences encoding a desirable bias of amino acids was proposed by Huse (Huse, U.S. Pat. Appl.No.2001/0024782, 2001), (Huse, U.S. Pat. No.6,258,530, 2001). Huse

proposed an algorithm of $20^{\text{sup}.n}$ (20^n , n is the unit of measurement of the length of oligonucleotide). If n represents nucleotide (Studier, Proc. Natl. Acad. Sci. U.S.A. 86: 6917-6921, 1989; Huse, U.S. Pat. No. 6,258,530, 2001), neither oligonucleotide produced according to algorithm $4^{\text{sup}.n}$ nor $20^{\text{sup}.n}$ has orientation capacity. Then, none of the oligonucleotides of the oligonucleotide library constructed in accordance with algorithm of $4^{\text{sup}.n}$ or algorithm of $20^{\text{sup}.n}$ would be able to discriminate the template strand (anti-sense) from non-template strand (sense) of a DNA double helix and *vice versa* in hybridization. Another disadvantage is that they inevitably include huge amounts of non-sense codons in the sequences of oligonucleotides, which virtually do not exist in ORF.

For example, for 6mer oligonucleotides (six nucleotides in length), there are 4,096 ($4^{\text{sup}.6}$) oligonucleotides by Studier and Szybalski's method; there are 64,000,000 ($20^{\text{sup}.6}$) oligonucleotides by Huse's method. The present invention proposed the algorithm of $61^{\text{sup}.(n-1)}$. In algorithm of $61^{(n-1)}$, n is the unit of measurement of the length of the oligonucleotide. n represents codon / triplet of nucleotides. For 6mer oligonucleotide (two codon in length), there are 61 5'-end start codon (5'-ATG) oriented oligonucleotides or 61 3'-end stop codon (3'-AAT or 3'-GAT or 3'-AGT) oriented oligonucleotides by the inventive methods.

DNA Microarrays is a systematic approach of detecting gene expression patterns in a quantitative, parallel, simultaneous and massive manner (Fodor et al., Science 251: 767-773, 1991); (Schena et al., Science 270: 467-470, 1995); (Fodor et al., U.S. Pat No. 5510270, 1996 and U.S. Pat. No. 5,800,992, 1998); (Southern et al., U.S. Pat. No. 5700637, 1997); (Chu et al., Trends in Biotechnology, 17, 217-218, 1999). It usually consists of thousands to hundred of thousands of known DNA sequences immobilized on a miniaturized solid surface as the probes. Each distinctive DNA sequence immobilized has its own well defined positions on the substrate. Through hybridization, DNA Microarrays could identify and demonstrate the responsive sequences, dynamics and patterns of expressed genes of a given sample. It can visualize the results of the hybridization of thousands of cDNA molecules in one single experiment. The nucleic acids in a given test(s) and control sample(s) were usually previously labelled with fluorescent molecules, such as Cy3 and Cy5 respectively. There are also cases wherein the nucleic acids in a given sample(s) was/were radioactively labelled, such as ^{33}P , ^{32}P and ^{35}S . Oligonucleotide and cDNA Microarrays are the two major formats of DNA Microarrays. The oligonucleotides were either pre-synthesized or synthesized in situ. However, a

specific RNA or cDNA sample of a specific cell line, tissue, organ and organism under defined circumstances usually limits its application of both formats of DNA Microarrays. Affymetrix, Inc. attempted to circumvent these problems by designing generic oligonucleotide arrays. However, the design was based on all possible combinations of four nucleotides or bases (A.T.G.C.). 4^n is the algorithm. n is the length of polydeoxynucleotides measured by nucleotide. n represents nucleotide (Lipshutz et al., Nat. Genet. 21: 20-24, 1999). It is actually the same model and system proposed by Studier. The disadvantages as discussed above can hardly to be ignored.

First, the oligonucleotides of a set or library constructed by all possible combinations of four nucleotides can not discriminate target sequences among non-coding, coding and regulatory regions on a single strand of nucleic acids as probes.

Second, even within a targeting coding region, the template strand (anti-sense) and non-template strand (sense) would be targeted indifferently by those generic oligonucleotides in hybridization.

Third, even withing a target sequence of a single strand, such as the template strand, those mentioned nucleotide-based oligonucleotide probes are incapable to discriminate the orientations of the sequence in hybridization.

Fourth, it is known in the art that the analysis of gene functionality is mainly relied on the analysis of coding regions. But the algorithm of 4^n is not a codon-based approach. The redundancy is phenomenal and hinders the accuracy of hybridization. It increases the cost of production and complicates the operation dramatically. For example, for 24mer oligonucleotides, the number of all possible combinations of oligonucleotides based on algorithm of $61^{(n-1)}$ is 3,142,742,836,021 [$61^{(8-1)}$] while the number of all possible combinations of oligonucleotides based on algorithm of 4^n is 281,474,976,710,656 [4^{24}]. The relationship between codon and nucleotide regarding the length of an oligonucleotide is as follows: n codon-length long oligonucleotide equals 3 times n nucleotide-length long oligonucleotide. n represents codon while 3 multiply n represents nucleotide. The redundancy is 89.6 times more than the virtual ORF sequences. Furthermore, producing a 24mer oligonucleotide library for oligonucleotide arrays by the present invention is 89.6 times more cost-effective in production than its counterpart by algorithm of 4^n . Moreover, the efficiency will increase further with the elongation of the length of oligonucleotide following algorithm of $4^{3 \times n}$ divided by $61^{(n-1)}$. The redundancy of oligonucleotide sequence with specified length could be calculated in accordance with algorithm of $4^{3n} - 61^{(n-1)}$.

Fifth, since the Affymetrix's generic oligonucleotide arrays were constructed in accordance with algorithm of $4^{\text{sup.n}}$, the GC contents among oligonucleotide probes vary from 0% to 100%. However, GC content of oligonucleotides belongs to one of the factors that can not be ignored when determining the conditions for hybridization. Once thousands of oligonucleotides with variable GC content are immobilized on one piece of solid support, all of them will be exposed to a unique hybridization environment. Thus, a considerable number of the oligonucleotides have to hybridize under less than optimized stringency conditions, either too high or too low as a result. Consequently, it inevitably results in false positives or and negative hybridization results. Wood et al. and Fodor et al. have made their efforts in attempting to address this issue (Wood et al., Proc. Natl. Acad.Sci. U.S.A. 82: 1585-1588, 1985, Fodor et al., U.S. Pat. No. 6,197,506, 2001). However, the application of 2.4 to 3.0 M tetramethyl ammonium or tetraethyl ammonium (Wood et al., Proc. Natl. Acad.Sci. U.S.A. 82: 1585-1588, 1985 and Fodor et al., U.S. Pat. No. 6,197,506, 2001) as the buffer may minimize the effect of the GC bias in the hybridization to a certain degree.

Clearly, a novel method and system for gene cloning and analysis is needed. Ideally, it should be more accurate and precise. It should be capable to differentiate the two separate strands of a nucleic acid molecule with orientation systematically when targeting. It should have a full-coverage screening capacity. Its probe or and primer libraries should have the capacity of reflection of a full representation of the targets while eliminating all unnecessary redundancy.

SUMMARY OF THE INVENTION

The present invention provides the universal genetic algorithm of $Y=X^{\text{sup.}(n-m)}$. It is applicable to both sense and anti-sense strands of a gene. It is applicable to all the corresponding gene products, such as mRNA, cDNA, anti-sense RNA, anti-sense cDNA, peptide and protein. This algorithm is applicable at various evolutionary stages as well. In accordance with the universal genetic algorithm of $Y=X^{\text{sup.}(n-m)}$ and its derivatives, any possible sequences of gene and peptide of a given length can be predicted and deduced. The methods could be used to construct a series of novel and unique libraries for use in cloning genes of interest. They are particularly useful to clone novel genes with either known or unknown sequences.

In other embodiments, the libraries are advantageously used alone or in

combination to clone genes of interest. For example, 5'-start codon such as 5'-ATG oriented ORF library and 3'-start codon such as 3'-GTA (5'-ATG) oriented 5'-UTR library are used in tandem in hybridization to determine potential start codons within a gene of interest. Using the example above, 5'-start codon oriented ORF library will indicate all possible potential start codon sites through hybridization with the targeting DNA, cDNA and mRNA samples; and 3'-start codon oriented 5'-UTR library will indicate all possible potential start codon sites as well. The identity of the specific start codon can then be determined by any of a variety of means known in the art, for example, when 9mers are utilized, a 15mer oligonucleotide corresponding to the start codon of the gene, centered on 5'-ATG, in other words, XXXXXXATG(NN), wherein "N" is deduced from a codon encoding an amino acid, could be deduced experimentally. This oligonucleotide can then be used in combination with an oligo-d(T)₁₅₋₁₈ as a pair of PCR primers to clone a full-length gene of interest. Alternatively, 3'-termination codon, such as 3'-AGT (5'-TGA), orients ORF library. 5'-termination codon, such as 5'-TGA, orients 3'-UTR library. They may be used in tandem in hybridization as discussed above and the resulting oligonucleotide will correspond to 3'- termination codon of a gene of interest. The mentioned hybridization could be performed by any of a variety of means for example, dot blotting and oligonucleotide microarray technologies as well as other means known in the art. This is the "tandem" tactic in sequence deducing strategy.

The sequencing deducing power, one of the major characteristics of this invention, has orientation. For example, in another preferred embodiment, each of 3'-ends of 9mer 5'-ATG oriented oligonucleotide primer(s) which were identified as the upstream primer in an original positive PCR amplicon(s) was/were covalently linked with three consecutive 5'-nitroindoles, referred to hereafter as "the first adaptor", per oligonucleotide. This elongates the 9mer oligonucleotides to 12mers. The sequence(s) of the 9mer oligonucleotide(s) is / are known. 5'-nitroindole is a universal base. Thus, "the first adaptor" should be capable to hybridize with each of the 61 distinctive codons evenly and respectively. Subsequently, each one of the 61 distinctive codons for the 20 essential amino acids (3mer oligonucleotide in length) is used as "the second adaptor". Each one of the 61 "second adaptors" will be covalently linked to 3'-end of one of the 12mer oligonucleotides that has a 3'-end "the first adaptor" at each time respectively. The ligation elongates 12mer oligonucleotide to 15mer 5'-ATG oriented oligonucleotide. There are 61 distinct second adaptors. Thus, there are totally 61 distinctive 15mer oligonucleotides derived per original positive PCR amplicon. 61 individual testing PCR reactions were set up with each of the

61 distinctive 15mer oligonucleotides as upstream primer per reaction. The downstream primer for each reaction is oligo-d(T)₁₅₋₁₈. Each testing PCR has the same template from the testing cDNA sample whereas each testing PCR's control has the same template from the control cDNA sample. The second round of PCR will indicate which one(s) is/are the responsible 15mer oligonucleotide(s) among the 61 distinctive 15mer upstream primers. Once the responsible 15mer oligonucleotide(s) were determined by PCR, the sequence(s) of 5'-ATG oriented oligonucleotide(s) was/were deduced. To repeat this cycle, the corresponding sequence(s) of 18mer 5'-ATG oriented oligonucleotide(s) could be deduced as well. Either the length of 15mers or 18mers of the deduced oligonucleotide is sufficient as the upstream primer. With oligo-d(T)₁₅₋₁₈ as the downstream primer, a full-length of gene of interest could be identified and cloned using means known in the art. This is the tactic of "two adaptors" in sequence deducing strategy.

A peptide sequence could be deduced by applying the inventive strategy as well. For example, the N-terminal NH₂-EF (EcoR I sequence) oriented peptide library and the C-terminal EF-COOH (EcoR I sequence) oriented peptide library are used in tandem to determine potential EcoR I sites within a peptide of interest. When 6-peptides are utilized, 10-peptides corresponding to the EcoR I site of the peptide, centered on the EF, in other words, (NNNN) EF (NNNN), wherein "N" is any amino acid, could be deduced experimentally by any of a variety of means, such as ELISA and protein microarray technologies as well as other means known in the art. This is the tactic of "tandem" in peptide sequence deducing strategy.

The orientation deducing power includes ORF sequence being deduced from the peptide. For example, the above determined 10-peptide sequence could be reverse translated to a group of corresponding 30mer oligonucleotides, which in turn could be used in combination with an oligo-d(T) as pairs of PCR primers to clone the partial ORF of the gene of interest. The cloned partial ORF sequence could be used as a probe to identify and clone the full-length of the gene of interest using means known in the art.

Compared to the limitations of spectrums of sequences derived from the SAGE and EST, the present invention provides full-range spectrums for gene screening. For example, those inventive libraries of short oligonucleotides, such as 9mers, 12mers, 15mers, and 18mers of 5'-ATG orientated ORF libraries provide sufficient resources for selecting and determining the signature sequences for each expressed gene in an entire genome, such as humans. Once the signature sequence for each expressed gene of the entire genome of any biological species such as humans was identified, determined and

registered, the complete sequence analysis may not be required. Once the signature sequences of genetic related diseases were identified, determined and registered, the clinical genetic diagnosis tests could be more accurate and rapid. Those signature sequences could also be used as the genetic markers and drug targeting markers in prognosis tests, drug design and development. They can be used in forensic studies as well. Codon consists of triplet nucleotides namely A, T, G, C, and U. Each of A, T, G, C and U possesses a characteristic maximum energy absorption spectrum at 259 nm, 267 nm, 252 nm, 271 nm and 258 nm respectively. Therefore, each signature sequence possesses a detectable unique energy emission pattern. The existing extrinsic label substances, such as light emitting compounds, quenching compounds and radioactive compounds could be coupled with the signature sequence by means known in the art. The couplings of extrinsic labels to the signature sequence(s) of both nucleic acids and peptide could enable a more sophisticated signature signal for distinction and characterization. Those signature signal(s) could be the chemical and/or physical signals in different forms such as energy emitting, absorbing, transferring and quenching. The signature sequence and the corresponding signature signal could be detectable and measurable either chemically, or magnetically, or electromagnetically, or electronically or optically or combinatorially (Chan, U.S. Pat. No. 6,210,896, 2001).

Additionally, the massive parallel signature sequence (MPSS) tags (Brenner et al., Nature Biotech 18, 630-634, 2000) could be selected among those short oligonucleotides as well.

Another notable advantage of present invention is its capacity for streamlining the nucleotide redundancy in current DNA microarray technology, particularly of oligonucleotide microarrays. For example, producing a 24mer oligonucleotide library for oligonucleotide arrays by present invention is 89.6 times cost-effective in production than the design based on the algorithm of $4^{\text{sup.}n}$ (Table 16). Moreover, the efficiency will increase further with the elongation of the length of oligonucleotide following algorithm of $4^{\text{sup.}3 \times n}$ divided by $61^{\text{sup.}(n-1)}$ as $[4^{3n} / 61^{(n-1)}]$. The redundancy could be calculated in accordance with the algorithm of $4^{3n} - 61^{(n-1)}$.

The inventive oligonucleotide libraries constructed by all possible combinations of 61 codons in accordance with algorithm of $61^{\text{sup.}(n-1)}$ has a unique characteristic: the GC content of the most of the oligonucleotide libraries have a statistical Poisson distribution. For example, 3,721 9mer 5'-ATG orientation codon-based universal ORF sense oligonucleotide library could be classified into seven categories according to GC

content. (1) 64 of 77.8% GC content, (2) 384 of 66.7% GC content, (3) 928 of 55.6% GC content, (4) 1,168 of 44.4% GC content, (5) 820 of 33.3% GC content, (6) 308 of 22.2% GC content and (7) 49 of 11.1% GC content. The 3,721 3'-AGT/5'-TGA orientation codon-based universal 9mer ORF sense oligonucleotide library and the 3,721 3'-GAT/5'-TAG orientation codon-based universal 9mer ORF sense oligonucleotide library are the same. 5'-ATG oriented of both sense and antisense 5'-UTR oligonucleotides library possesses 4,096 distinctive 9mer oligonucleotides respectively. Those oligonucleotides were divided and classified into seven categories according to GC content: (1) 64 of 4,096 with 77.8% GC content, (2) 384 of 4,096 with 66.7% GC content, (3) 960 of 4,096 with 55.6% GC content, (4) 1,280 of 4,096 with 44.4% GC content, (5) 960 of 4,096 with 33.3% GC content, (6) 384 of 4,096 with 22.2% GC content and (7) 64 of 4,096 with 11.1% GC content. 9mer 5'-TGA or 5'-TAG oriented of both sense and antisense 3'-UTR oligonucleotides library respectively possesses 4,096 distinctive 9mer oligonucleotides. Those oligonucleotides were divided and classified into seven categories according to GC content: (1) 64 of 4,096 with 77.8% GC content, (2) 384 of 4,096 with 66.7% GC content, (3) 960 of 4,096 with 55.6% GC content, (4) 1,280 of 4,096 with 44.4% GC content, (5) 960 of 4,096 with 33.3% GC content, (6) 384 of 4,096 with 22.2% GC content and (7) 64 of 4,096 with 11.1% GC content. In some embodiments, oligonucleotides with 77.8% or greater GC content are grouped together while oligonucleotides with 11.1% or less GC content are grouped together. The oligonucleotides of a given identical length with an identical orientation are being classified by GC content (%) in order to regroup into a sub-library or panels. This is one of the most important features of these inventive panels. This eliminates or minimizes the extreme instances such as 0% GC content or 100% GC content in most of the libraries whereas this is not possible with the oligonucleotide designed based on algorithm of 4.sup.n. This property makes the universal orientation oligonucleotide library of codon-based design more suitable as probes in oligonucleotide arrays and PCR technologies. In accordance with the present invention, the synthetic oligonucleotides could be classified and organized into different groups by GC content. One group of the synthetic oligonucleotides with identical GC content could be immobilized on one piece of solid support composed of a suitable material, for example, glass, plastic, silicon, nylon, nitrocellulose or other suitable supports known in the art. The hybridization conditions, particularly the T_m could be tailored in accordance with the GC content. Each group of oligonucleotides could therefore interact with their targeting sequences either on the surface of the solid phase (DNA Microarray and Bot-Dot) or in

aqueous phase (PCR) under the optimized and well-defined hybridization conditions. As a result of this arrangement, false positives and negative results in the hybridization would be reduced while the fidelity and reliability is increased.

Single Nucleotide Polymorphism (SNP) is one of the important factors and phenomena in identifying and determining the genetic basis for disease and drug response. Although SNPs can occur within the coding region, 5'-UTR, 3'-UTR and non-coding regions, regardless the functionality, SNPs occurred in coding regions is the change of amino acid(s) of peptide chain(s). In nature, SNPs in coding regions are a change of codon(s) or and ORF(s). The present invention of the construction of oligonucleotide array is codon-based methodology. Therefore, it covers all the possible SNPs and point mutations in the coding regions and non-coding regions as well. One of the great benefits in the practice of the present invention is the simplicity and low cost of production. With the inventive universal oligonucleotide arrays, one can do it all. The present invention has made the oligonucleotide arrays, dot-blotting hybridization and PCR primers sets capable of multi-functioning. Additionally, every distinct oligonucleotide of the oligonucleotide libraries except hexamer oligonucleotide libraries designed by the codon-based methods posses orientation power while the corresponding one based on four nucleotides or bases does not.

Another advantage of the present invention is that each oligonucleotide of the oligonucleotide arrays could be used in conjunction with oligo-d(T) as a pair of PCR primers. Each specified PCR amplifies a corresponding cDNA molecule(s) which in turn is used as a cDNA probe(s) in the construction of cDNA arrays. The obtained cDNA(s) could be duplicated and propagated routinely as permanent and convenient resources for all suitable purposes. The cDNAs could be expressed as the corresponding peptide or protein which in turn could be used as antigen to produce the corresponding monoclonal or multiclonal antibody using means apparent to one skilled in the art. Those derived specific antigens and/or corresponding antibodies could be used to construct protein microarrays or the capture layer of ELISA and/or in other immunoassays. Therefore the inventive oligonucleotide libraries have the unique capacity to derive the corresponding cDNA and peptide libraries.

The benefit of the inventive oligonucleotide library derived cDNA library is enormous. For example, the inventive oligonucleotide libraries and methods will enable cDNA microarrays and cDNA Dot-blotting to employ unknown cDNA probes, which were derived from the known inventive oligonucleotides. Using those "unknown" cDNA, one

skilled in the art can identify the gene expression patterns and the responsible sequences of interest through hybridization. Then using the known corresponding oligonucleotide(s) as the upstream primer(s) with oligo-d(T) as the downstream primer, the full length of the gene of interest could be eventually identified and cloned in accordance with means known in the art. The strategy of using "unknown cDNA derived from the known oligonucleotide" is much more efficient, rapid, economical and time saving than the current DNA microarrays technologies. In contrast, current DNA microarrays require all cDNA sequences to be pre-determined prior to depositing and immobilizing them on the solid surface as the probe.

Overall, the codon-based methods are also a systematic approach which have the unique capacity to simplify and standardize various existing methods based on four nucleotides and used for different purposes into a set of unique, universal, high-throughput and all-purpose methods. Nowadays, deoxyoligonucleotide and peptides can be synthesized respectively and routinely by a DNA or protein synthesizer. Such method(s) and working platform(s) is/are particular useful in the design and production of various types of high density arrays of oligonucleotides including the Dot-Blotting hybridization and PCR Cloning working kits at massive scale level with production cost effectiveness. The counterparts of universal anti-sense oligonucleotide libraries were also provided by this present invention in order to identify and select the efficient anti-sense oligonucleotides and reagents for research, diagnosis, treatment and drug development.

The inventive methods provide a powerful platform to combine and integrate PCR cloning technology and DNA microarray technology together to identify and clone the responsible genes having unknown sequences. They provide a different dimension for exploring gene expression patterns and functionalities of the genome. They are particularly useful in the fields of genomics, proteomics and bioinformatics regarding high-throughput data analyzing, mining and integrating. It is an attempt to standardize a variety of formats of biological data and tools. These algorithms, chemical composition formulae and the biological models deduced and established from these methods has a great potential to contribute to the development of DNA computing, standardization and the automation of laboratory reagents, procedures, equipment, computer software, and other relevant facilities.

According to a first aspect of the invention, there is provided a kit for identifying genetic signatures within a sample comprising at least one of the following:

a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein

each of the oligonucleotides is represented by the formula $5'-I_S(C_S)_{n1}-3'$, wherein I_S represents an initiation codon in sense orientation, C_S represents an amino acid coding codon in sense orientation and $n1$ is an integer;

a 5' start codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n2}I_A-3'$, wherein I_A represents an initiation codon in antisense orientation, C_A represents an amino acid coding codon in antisense orientation and $n2$ is an integer;

a 5' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(V_S)_{n3}I_S-3'$, wherein I_S represents an initiation codon in sense orientation, V_S represents a codon in sense orientation, and $n3$ is an integer;

a 5' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_A(V_A)_{n4}-3'$, wherein I_A represents an initiation codon in antisense orientation, V represents a codon in antisense orientation, and $n4$ is an integer;

a 3' stop codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n5}T_S-3'$, wherein C_S represents an amino acid coding codon in sense orientation, T_S represents a termination codon in sense orientation, and $n5$ is an integer;

a 3' stop codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_A(C_A)_{n6}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation, T_A represents a termination codon in antisense orientation and $n6$ is an integer;

a 3' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_S(V_S)_{n7}-3'$, wherein T_S represents a termination codon in sense orientation, V_S represents a codon in sense orientation and $n7$ is an integer;

a 3' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(V_A)_{n8}T_A-3'$, wherein T_A represents a termination codon in antisense orientation, V_A represents a codon in antisense orientation and $n8$ is an integer;

a 5' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-R_S(C_S)_{n9}-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense

orientation with the proviso that neither of the two codons is a termination codon, C_9 represents an amino acid coding codon in sense orientation and n_9 is an integer;

a 5' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n_{10}}R_A-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and n_{10} is an integer;

a 3' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n_{11}}R_S-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and n_{11} is an integer;

a 3' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-R_A(C_A)_{n_{12}}-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and n_{12} is an integer;

a between 5' and 3' (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n_{13}}-3'$, wherein C_S represents an amino acid coding codon in sense orientation and n_{13} is an integer;

a between 5' and 3' (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n_{14}}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation and n_{14} is an integer;

a N-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $R_E(N)_{n_{15}}$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition site in sense orientation, and n_{15} is an integer;

a C-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $(N)_{n_{16}}R_E$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition site in sense orientation, and n_{16} is an integer;

a N-terminal peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $M(N)_{n_{17}}$ -C-terminal, wherein M represents an amino acid encoded by an initiation codon, N represents an amino acid and

n_{17} is an integer; and

a C-terminal peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $(N)_{n_{18}}$ -C-terminal, wherein N represents an amino acid and n_{18} is an integer.

According to a second aspect of the invention, there is provided a method of preparing a cDNA library comprising:

providing a quantity of mRNA,

admixing the mRNA with either a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_s(C_s)_{n_1}-3'$, wherein I_s represents an initiation codon in sense orientation, C_s represents an amino acid coding codon in sense orientation and n_1 is an integer; or a between 5' and 3' (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_s)_{n_{13}}-3'$, wherein C_s represents an amino acid coding codon in sense orientation and n_{13} is an integer and an oligo-d(T) $_N$ panel, wherein N is an integer; or an oligo-d(T) $_N$ panel; and

incubating the mRNA and the panel in the presence of reagents needed for DNA synthesis and under conditions suitable for DNA synthesis, thereby producing first strands of a cDNA library.

According to a third aspect of the invention, there is provided a method of amplifying a cDNA library comprising:

admixing a quantity of cDNA with a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_s(C_s)_{n_1}-3'$, wherein I_s represents an initiation codon in sense orientation, C_s represents an amino acid coding codon in sense orientation and n_1 is an integer; and

incubating the cDNA and the panel in the presence of reagents needed for DNA amplification and under conditions suitable for DNA amplification, thereby producing a cDNA library.

According to a fourth aspect of the invention, there is provided a method of preparing an ordered cDNA library comprising:

providing a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_s(C_s)_{n_1}-3'$, wherein I_s represents an initiation codon in sense orientation, C_s represents an amino acid coding codon in sense orientation and n_1 is an integer, each said oligonucleotide being mounted at a specific site on said panel; and

admixing a quantity of mRNA or cDNA or genomic DNA with the panel under conditions suitable for nucleic acid hybridization; and

identifying positive oligonucleotide(s) binding to at least one mRNA or cDNA molecule; and

mounting each of the positive oligonucleotides to a specific site on a solid support; and

rescreening the quantity of mRNA or cDNA or genomic DNA with the positive oligonucleotides.

According to a fifth aspect of the invention, there is provided an oligonucleotide primer or probe comprising at least one oligonucleotide selected from the group consisting of:

an oligonucleotide represented by the formula $5'-I_S(C_S)_{n1}-3'$, wherein I_S represents an initiation codon in sense orientation, C_S represents an amino acid coding codon in sense orientation and $n1$ is an integer;

an oligonucleotide represented by the formula $5'-(C_A)_{n2}I_A-3'$, wherein I_A represents an initiation codon in antisense orientation, C_A represents an amino acid coding codon in antisense orientation and $n2$ is an integer;

an oligonucleotide represented by the formula $5'-(V_S)_{n3}I_S-3'$, wherein I_S represents an initiation codon in sense orientation, V_S represents a codon in sense orientation, and $n3$ is an integer;

an oligonucleotide represented by the formula $5'-I_A(V_A)_{n4}-3'$, wherein I_A represents an initiation codon in antisense orientation, V_A represents a codon in antisense orientation, and $n4$ is an integer;

an oligonucleotide represented by the formula $5'-(C_S)_{n5}T_S-3'$, wherein C_S represents an amino acid coding codon in sense orientation, T_S represents a termination codon in sense orientation, and $n5$ is an integer;

an oligonucleotide represented by the formula $5'-T_A(C_A)_{n6}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation, T_A represents a termination codon in antisense orientation and $n6$ is an integer;

an oligonucleotide represented by the formula $5'-T_S(V_S)_{n7}-3'$, wherein T_S represents a termination codon in sense orientation, V_S represents a codon in sense orientation and $n7$ is an integer;

an oligonucleotide represented by the formula $5'-(V_A)_{n8}T_A-3'$, wherein T_A represents a termination codon in antisense orientation, V_A represents a codon in

antisense orientation and $n8$ is an integer;

an oligonucleotide represented by the formula $5'-R_S(C_S)_{n9}-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and $n9$ is an integer;

an oligonucleotide represented by the formula $5'-(C_A)_{n10}R_A-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and $n10$ is an integer;

an oligonucleotide represented by the formula $5'-(C_S)_{n11}R_S-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and $n11$ is an integer;

an oligonucleotide represented by the formula $5'-R_A(C_A)_{n12}-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and $n12$ is an integer;

an oligonucleotide represented by the formula $5'-(C_S)_{n13}-3'$, wherein C_S represents an amino acid coding codon in sense orientation and $n13$ is an integer;

an oligonucleotide represented by the formula $5'-(C_A)_{n14}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation and $n14$ is an integer; and

combinations thereof.

According to a seventh aspect of the invention, there is provided a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_S(C_S)_{n1}-3'$, wherein I_S represents an initiation codon in sense orientation, C_S represents an amino acid coding codon in sense orientation and $n1$ is an integer.

According to an eighth aspect of the invention, there is provided a 5' start codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n2}I_A-3'$, wherein I_A represents an initiation codon in antisense orientation, C_A represents an amino acid coding codon in antisense orientation and $n2$ is an integer.

According to a ninth aspect of the invention, there is provided a 5' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is

represented by the formula $5'-(V_S)_{n3}I_S-3'$, wherein I_S represents an initiation codon in sense orientation, V_S represents a codon in sense orientation, and $n3$ is an integer.

According to a tenth aspect of the invention, there is provided a 5' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-I_A(V_A)_{n4}-3'$, wherein I_A represents an initiation codon in antisense orientation, V represents a codon in antisense orientation, and $n4$ is an integer.

According to an eleventh aspect of the invention, there is provided a 3' stop codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n5}T_S-3'$, wherein C_S represents an amino acid coding codon in sense orientation, T_S represents a termination codon in sense orientation, and $n5$ is an integer.

According to a twelfth aspect of the invention, there is provided a 3' stop codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_A(C_A)_{n6}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation, T_A represents a termination codon in antisense orientation and $n6$ is an integer.

According to a thirteenth aspect of the invention, there is provided a 3' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_S(V_S)_{n7}-3'$, wherein T_S represents a termination codon in sense orientation, V_S represents a codon in sense orientation and $n7$ is an integer.

According to a fourteenth aspect of the invention, there is provided a 3' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(V_A)_{n8}T_A-3'$, wherein T_A represents a termination codon in antisense orientation, V_A represents a codon in antisense orientation and $n8$ is an integer.

According to a fifteenth aspect of the invention, there is provided a 5' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-R_S(C_S)_{n9}-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and $n9$ is an integer.

According to a sixteenth aspect of the invention, there is provided a 5' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n10}R_A-3'$, wherein R_A represents a

two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and $n10$ is an integer.

According to a seventeenth aspect of the invention, there is provided a 3' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n11}R_S-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and $n11$ is an integer.

According to an eighteenth aspect of the invention, there is provided a 3' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-R_A(C_A)_{n12}-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and $n12$ is an integer.

According to a nineteenth aspect of the invention, there is provided a between 5' and 3' (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n13}-3'$, wherein C_S represents an amino acid coding codon in sense orientation and $n13$ is an integer.

According to twentieth aspect of the invention, there is provided a between 5' and 3' (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n14}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation and $n14$ is an integer.

According to a twenty-first aspect of the invention, there is provided a N-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $R_E(N)_{n15}$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition site in sense orientation, and $n15$ is an integer.

According to a twenty-second aspect of the invention, there is provided a C-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $(N)_{n16}R_E$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition site in sense orientation, and $n16$ is an integer.

According to a twenty-third aspect of the invention, there is provided a N-terminal

peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal-M(N)_{n17}-C-terminal, wherein M represents an amino acid encoded by an initiation codon, N represents an amino acid and n17 is an integer.

According to a twenty-fourth aspect of the invention, there is provided a C-terminal peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal-(N)_{n18}-C-terminal, wherein N represents an amino acid and n18 is an integer.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skilled in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described.

I. DEFINITIONS, ALGORITHM AND KITS

“Panel” refers to a plurality of reagents, for example, oligonucleotides or peptides, as discussed below. The panel may be mounted to a support or the entire panel or individual members thereof may be in a substantially aqueous phase. Suitable supports include but are by no means limited to nitrocellulose, nylon, glass, silicon, beads, for example, streptavidin beads, nanoparticles and the like. It is of note that while the singular form is used, it is to be understood that the panel may in fact comprise a single library mounted to, that is, spanning, several separate supports. As will be appreciated by one of skilled in the art, when the panel includes a solid support, the oligonucleotides or peptides may be mounted directly to the support or the surface of the support or may be linked by a suitable linker as known in the art.

“Genetic signature” refers to a biological characteristic of, for example, a gene, peptide, a nucleic acid sequence, a peptide sequence, antibody, cell, cell line or organism. Examples of genetic signatures include but are by no means limited to locations of start and stop codons within a gene, location of restriction enzyme sites within a gene, presence of antibodies recognizing a specific amino acid sequence in a sample,

expression patterns of a cell or a cell line and the like.

“Substantially all” refers to the fact that a sufficient number of a group are present that the desired result can be obtained or determined.

“Universal base” refers to compounds capable of substituting for binding to any one of A, C, G, T and U in nucleic acids without significant structure destabilization. The oligonucleotide incorporated with the universal base analogues is able to function as a probe in for example hybridization, and as a primer in for example PCR reactions and DNA sequencing. Examples of universal bases include but are by no means limited to 5'-nitroindole-2'-deoxyriboside, 3-nitropyrrole, inosine and pypoxanthine.

“Consisting essentially of” means that the described compounds consist of those nucleotides or peptides as described in the formula listed as well as other components, which do not materially affect the invention. In the case of the oligonucleotides, these might include for example one or more universal bases, for example, a nitroindole, as described herein.

“any codon” refers to any one of the 64 nucleotide triplets of the genetic code.

“sense” orientation or strand refers to the coding strand or complementary strand of non-coding strand or antisense strand of a double stranded DNA molecule.

“antisense” orientation or strand refers to the non-coding strand or complementary strand of a coding strand or sense strand of a double-stranded DNA molecule.

“amino acid coding codon” refers to a codon of a given genetic code encoding for an amino acid. In most cases, 61 codons code for the 20 essential amino acids. As an example, in one genetic code, in sense orientation, 5'-AGG codes for arginine. The antisense of this codon would be 5'-CCT.

“initiation codon” refers to a codon that may function as a start codon. In most cases, the initiation codon (in sense orientation) is 5'-ATG. The antisense orientation of this codon is 5'-CAT. As discussed herein, other initiation codons may be used, for example, 5'-ATA, which is the start codon in mammalian mitochondria. Other exceptions include 5'-GTG, 5'-ATA, 5'-TTG, 5'-ACG and 5'-CTG, which may function as start codons in some cases.

“termination codon” refers to a codon that may function as a stop codon. In most cases, there are three dominant stop codons: 5'-TAA, 5'-TGA and 5'-TAG. In antisense orientation, those corresponding codons are 5'-TTA, 5'-TCA and 5'-CTA.

“plurality” refers to 2 or more.

The invention envisions a gene as a linear polymer selected from a group

consisting of all possible combinations of 61 codons with a start codon at its 5'-end and a stop codon at its 3'-end. This is different from the traditional one which views a gene as a DNA sequence consisting of the combinations of four distinct nucleotide of A, T, G and C whether coding region, 5'-UTR or 3'-UTR. With the invention, any coding region is selected from a group consisting of all possible combinations of 61 codons with a start codon at its 5' end and a stop codon at its 3-end. Any 5'-Untranslated Region (5'-UTR) is selected from a group consisting of all possible combinations of 64 codons with a start codon at its 3'-end. Any 3'-Untranslated Region (3'-UTR) is selected from a group consisting of all possible combinations of 64 codons with a stop codon adding at its 5'-end. Any coding region was envisioned as one selected from a group of linear DNA sequences which consists of all possible combinations of 61 codons. Another advantage of this innovative concept is its capacity to differentiate the genes of mammalian genomic DNA origin from mitochondrial genes. The genes of mammalian mitochondria possess unique characteristics: for example, 5'-ATA replaces 5'-ATG for Met; 5'-TGA encodes Trp instead of termination. Therefore, any gene of mammalian mitochondria could be envisioned as one selected from a group of linear DNA sequences which consists of all the possible combinations of 59 codons, in which 5'-ATA substitutes 5'-ATG and 5'-TGA substitutes for 5'-AGA and 5'-AGG of the group of 61 codons. Such a linear DNA sequence has 5'-ATA at its 5'-end as the start codon and one of 5'-AGA, 5'-AGG and 5'-TAA at its 3'-end as the stop codon. The invention also envisions a gene product, such as peptide or polypeptide as a linear polymer selected from a group consisting of all possible combinations of 20 essential amino acids (EAA) encoded by 61 codons with an amino acid encoded by a 5'-start codon, such as Methionine at its N-terminal. The 20 essential amino acids (EAA) are encoded by the 61 distinct codons.

The present invention provides an inventive universal genetic algorithm. This algorithm provides a universal calculation formula for the total number of sequences of a given length measured by either single codon as the unit or L-amino acid encoded by codon as the expressed-codon unit when the orientation has been decided. The orientation for gene is either 5'- or 3'- orientation. The orientation for a peptide, the product of a gene, is either N-terminal or C-terminal orientation. The length measured by codon can convert to the length measured by single nucleotide by multiplying 3.

This algorithm is applicable to sense and anti-sense strands of a gene and all the corresponding gene products, such as mRNA, cDNA, anti-sense RNA, anti-sense cDNA, peptide and protein. This algorithm is applicable at various evolutionary stages.

The inventive universal genetic algorithm is presented herein:

$$Y = X^{\sup.(n-m)}$$

The annotation of the algorithm is presented herein:

1. Definition of X

(1) Nucleic Acids:

X: The number of all distinct codons. X is a variable. X is an integer. X is not equal zero. X is from 1 to infinity. At the current evolutionary stage, for all distinct codons, X=64. For all distinct codons that encode L-amino acid, X=61.

(2) Peptides:

X: The number of all distinct L-amino acids encoded by at least one codon. X is a variable. X is an integer. X is not equal zero. X is from 1 to infinity. At the current evolutionary stage, there are 20 distinct essential L-amino acids that are encoded by their 61 distinct corresponding codons. X=20.

2. Definition of n

(1) Nucleic Acids:

n: The number of all codons arranged linearly without overlapping per sequence. n is a variable. n is an integer. n is not equal zero. n is from 1 to infinity. n is the length of the sequence measured by codon (triplet of nucleotides). n is the number of codons for either 5'- or 3'-orientated ORF sequences.

(2) Peptides:

n: The number of all L-amino acids arranged linearly without overlapping per sequence. n is a variable. n is an integer. n is not equal zero. n is from 1 to infinity. n is the length of the sequence measured by L-amino acids encoded by the codons. n is the number of amino acids for either N-terminal or C-terminal orientated peptide or protein sequence.

3. Definition of m

(1) Nucleic Acids:

m: The number of a single or all codons arranged linearly without overlapping in a unique sequence at the beginning per entire sequence of either 5'- or 3'- orientation. m is pre-determined orientation sequence. For example, if there are no codons arranged linearly without overlapping in a unique sequence at the beginning per entire sequence of either 5'- or 3'- orientation, m=0; if each sequence started from 5'-ATG in a 5'-orientation, m=1; if each sequence started from 3'-AGT (5'-TGA) in a 3'-orientation, m=1; if each sequence started from 5'-GAATTC (*EcoR* I recognition sequence) in a 5'-orientation, m=2;

if each sequence started from 5'-CACACAGGAGAAAAGCCA (The conservative motif of six amino acids of a zinc finger gene family) in a 5'-orientation, $m=6$. m is a variable. m is an integer. m is from 0 to n . $m < n$.

(2) Peptides:

m : The number of a single or all amino acids arranged linearly without overlapping in a unique sequence at the beginning per entire sequence of either N-terminal or C-terminal orientation. m is the pre-determined orientation sequence. For example, if there are no amino acids encoded by codons arranged linearly without overlapping in a unique sequence at the beginning per entire sequence of either N-terminal or C-terminal orientation, $m=0$; if each sequence started from an amino acid encoded by a start codon, such as Methionine encoded by 5'-ATG, in N-terminal orientation, $m=1$; if each sequence started from one amino acid of the amino acids encoded by their 61 corresponding codons in a C-terminal orientation, $m=1$; if each sequence started from N-EF (amino acids encoded by *EcoR* I recognition sequence) in a N-terminal orientation, $m=2$; if each sequence started from NH_2 -HTGFP (The conservative motif of six amino acids of zinc finger gene family) in a N'-terminal orientation, $m=6$. m is a variable. m is an integer. m is from 0 to n . $m < n$.

4. Definition of the Negative Sign

A negative sign in front of codon(s) means that the position of the codon(s) is (are) in 5'-Untranslated Region (5'-UTR).

5. Definition of Y

(1) Nucleic Acids:

Y : The total number of sequences of n codon length long

(2) Peptides:

Y : The total number of sequences of n amino acids length long. The amino acids are encoded by the corresponding codons

6. Application

(1) Nucleic Acids:

(a) If $X=61$ and $m=0$, $Y=61.\sup.n$

Applicable to ORF sequences between 5'- and 3'-ends

(b) If $X=61$ and $m=1$, $Y=61.\sup.(n-1)$

Applicable to ORF sequences oriented by one 5' start codon or one 3' stop codon (The first position of codon at 5'-terminal of ORF sequences is occupied by one of the start codons per sequence, or the first position of codon at 3'-terminal of ORF sequences is

occupied by one of the stop codons per sequence.)

(c) If $X=61$ and $m=2$, $Y=61.\sup.(n-2)$

Applicable to ORF sequences oriented by one 5' two-codon restriction enzyme recognition sequence or one 3' two-codon restriction enzyme recognition sequence (The first and second consecutive positions of codon at 5'-terminal of ORF sequences are occupied by one 5' two-codon restriction enzyme recognition sequence per sequence, or the first and second consecutive codon positions at 3'-terminal of ORF sequences are occupied by one of 3'-two-codon restriction enzyme recognition sequence per sequence.)

(d) If $X=61$ and $(n-m)>1$, $Y=61.\sup.(n-m)$

Applicable to ORF sequences oriented by one 5' pre-determined sequence which consists of m codons or one 3' pre-determined sequence which consists of m codons (m consecutive codon positions started from 5'-terminal of ORF sequences are occupied by one of 5' pre-determined sequences which consists of m codons per sequence, or m consecutive codon positions started from 3'-terminal of ORF sequences are occupied by one of 3' pre-determined sequence which consists of m codons per sequence.)

(e) If $X=64$ and $(n-m)>1$, $Y=64.\sup.(1-n)$

Applicable to 5'-UTR sequences oriented by one 3' start codon (3' of the last position of codon at 3'-end of 5'-UTR is covalently linked with a 5' start codon, such as 5'-ATG, per sequence.)

(f) If $X=64$ and $(n-m)=1$, $Y=64.\sup.(1-n)+1$

Applicable to 5'-UTR sequences oriented by one 3' start codon (3' of the last position of codon at 3'-end of 5'-UTR is covalently linked with a 5' start codon, such as 5'-ATG, per sequence.)

(g) If $X=64$ and $(n-m)>1$, $Y=64.\sup.(n-1)$

Applicable to 3'-UTR sequences oriented by one 5' stop codon (5' of the last position of codon at 5'-end of 3'-UTR is covalently linked with a 5' stop codon, such as 5'-TGA, per sequence.)

(2) Peptides:

(a) If $X=20$ and $m=0$, $Y=20.\sup.n$

Applicable to sequences oriented by one C-terminal amino acid encoded by one of the 61 corresponding codons (The first position of amino acid started from C-terminal is occupied by one amino acid from the group of amino acids encoded by 61 corresponding codons per sequence.) It is applicable to the sequences between N-terminal and C-terminal.

(b) If $X=20$ and $m=1$, $Y=20.\sup.(n-1)$

Applicable to sequences oriented by one N-terminal amino acid encoded by a start codon (The first position of amino acid started from N-terminal is occupied by one amino acid encoded by start codon, such as Methionine encoded by 5'-ATG per sequence.)

(c) If $X=20$ and $m=2$, $Y=20.\sup.(n-2)$

Applicable to sequences oriented by one N-terminal two-amino acid encoded by the corresponding two-codon restriction enzyme recognition sequence or one C-terminal two-amino-acids encoded by corresponding two-codon restriction enzyme recognition sequence (The first and second consecutive amino acid positions at N-terminal of peptide sequences are occupied by one of N-terminal two-amino-acids encoded by two-codon restriction enzyme recognition sequence per sequence, or the first and second consecutive amino acid positions at C-terminal of peptide sequences are occupied by one of C-terminal two-amino-acids encoded by two-codon restriction enzyme recognition sequence per sequence.)

(d) If $X=20$ and $(n-m)>1$, $Y=20.\sup.(n-m)$

Applicable to sequences oriented by m N-terminal amino acid(s) encoded by corresponding codon(s) or sequence(s) oriented by m C-terminal amino acid(s) encoded by corresponding codon(s) (m amino acid position(s) of amino acid(s) started from N-terminal of peptide sequence(s) is/are occupied by one of N-terminal pre-determined amino acid sequence(s) which consists of m amino acid(s) per sequence or m amino acid position(s) of amino acid(s) started from C-terminal of peptide sequence(s) is/ are occupied by one of C-terminal pre-determined sequence(s) which consists of m amino acid(s) per sequence.)

The inventive universal genetic algorithm of $Y=X.\sup.(n-m)$ provides a quantitative vehicle to deduce any and all the possible sequence(s) of either nucleic acid or peptide of a given length. Thus, the series working platforms addressing different biological and medical problems could be established systematically. The inventive methodologies are codon-based systems biology methodologies. The methods exclude nonsense codons, which are not present in coding regions such as ORF of a gene, and include all codons in untranslated regions in the probe or and primer designing. In terms of technology, this is enormously beneficial. Accompanied with the elimination of oligonucleotide redundancy, the fidelity, reproducibility and reliability of results obtained by PCR, sequencing, cloning and DNA microarrays analysis is increased.

7. Orientation

There are two orientations of each single strand of nucleic acids: 5'-towards-3' and 3'-towards-5'. There are two orientations of each peptide sequence: N-terminal towards C-terminal and C-terminal towards N-terminal.

8. Strands of the Double Helix of Nucleic Acids

There are two strands of each helix of a nucleic acid molecule: Sense strand (non-template strand) and anti-sense strand (template strand).

9. Hybridization

Hybridization is an interaction between two strands of nucleic acids by hydrogen bonds in accordance with the rules of Watson-Crick DNA complementarity. Hybridization can be performed under different stringent conditions known in the art.

10. A Number of Novel and Unique Libraries for Use in Cloning Genes of Interest

(1) There is provided a library with 5'-end start codon orientation. The library of oligonucleotides consists of all possible combinations of 61 codons (TABLE 1) with a start codon, such as 5'-ATG, as 5'-end terminal codon for each oligonucleotide at a given length and a peptide library corresponding to amino acids deduced from the nucleotide sequence. As will be appreciated by one of skilled in the art, the result of this arrangement is that those oligonucleotides would preferentially hybridize to regions of template strand (antisense) of genomic DNA or 1st single strand of cDNA. Which is from an antisense start codon, such as 5'-CAT to its upstream sequence within the antisense coding region of 5' towards 3' orientation due to the fact that sequences corresponding to termination codons are specifically excluded.

(2) There is also provided a library with 3'-end antisense start codon orientation. The library of antisense oligonucleotides consists of all possible combinations of 61 antisense codons (TABLE 2) with an antisense start codon, such as 5'-CAT, as 3'-end terminal antisense codon for each antisense oligonucleotide at a given length. As will be appreciated by one of skilled in the art, the result of this arrangement is that those antisense oligonucleotides would preferentially hybridize to regions of non-template strand (sense) of genomic DNA or mRNA or 2nd single strand of cDNA. Which is from a start codon, such as 5'-ATG to its downstream sequence within the coding region of 5' towards 3' orientation due to the fact that sequences corresponding to antisense termination codons are specifically excluded.

(3) There is also provided a library with 3'-end stop codon orientation. The library of oligonucleotides consists of all possible combinations of 61 codons (TABLE 1) with a stop codon, such as 5'-TGA, as 3'-end terminal codon for each oligonucleotide at a

given length and a peptide library corresponding to amino acids deduced from the nucleotide sequence. As will be appreciated by one of skilled in the art, the result of this arrangement is that those oligonucleotides would preferentially hybridize to regions of template strand (antisense) of genomic DNA or 1st single strand of cDNA. Which is from an antisense stop codon, such as 5'-TCA to its downstream sequence within the antisense coding region of 5' towards 3' orientation due to the fact that sequences corresponding to termination codons are specifically excluded.

(4) Similarly, there is also provided a library with 5'-end antisense stop codon orientation. The library of antisense oligonucleotides consists of all possible combinations of 61 antisense codons (TABLE 2) with an antisense stop codon, such as 5'-TCA, as 5'-end terminal antisense codon for each antisense oligonucleotide at a given length. As will be appreciated by one of skilled in the art, the result of this arrangement is that those antisense oligonucleotides would preferentially hybridize to regions of non-template strand (sense) of genomic DNA or mRNA or 2nd single strand of cDNA. Which is from a stop codon, such as 5'-TGA to its upstream sequence within the coding region of 5' towards 3' orientation due to the fact that sequences corresponding to antisense termination codons are specifically excluded.

(5) There is also provided a library with orientations of either 5'-end two-codon restriction enzyme recognition sequence or 3'-end two-codon restriction enzyme recognition sequence (TABLE 3). For example, the library of oligonucleotides consists of all possible combinations of 61 codons (TABLE 1) with a two-codon recognition sequence, such as 5'-GACGTC (Aat II), as 5'-end terminal of two consecutive codons for each oligonucleotide at a given length and a peptide library corresponding to amino acids deduced from the nucleotide sequence. As will be appreciated to one of skilled in the art, the sequence of restriction enzymes is on the codon basis. The termination codons within their recognition sequence are omitted from the library. The result of this arrangement is that those oligonucleotides would preferentially hybridize to regions of template strand (antisense) of genomic DNA or 1st single strand of cDNA. Which is from an antisense two-codon recognition sequence, such as 5'-GACGTC (Aat II) to its upstream sequence within the antisense coding region of 5' towards 3' orientation.

(6) Accordingly, there is also provided a library with orientations of either 3'-end antisense two-codon restriction enzyme recognition sequence or 5'-end antisense two-codon restriction enzyme recognition sequence (TABLE 3). For example, the library of antisense oligonucleotides consists of all possible combinations of 61 antisense codons

(TABLE 2) with an antisense two-codon recognition sequence, such as 5'-GACGTC (Aat II), as 3'-end terminal of two consecutive antisense codons for each antisense oligonucleotide at a given length. As will be appreciated to one of skilled in the art, the sequence of restriction enzymes is on the antisense codon basis. The antisense termination codons within their recognition sequence are omitted from the library. The result of this arrangement is that those antisense oligonucleotides would preferentially hybridize to regions of non-template (sense) strand of genomic DNA or mRNA or 2nd single strand of cDNA. Which is from a two-codon recognition sequence, such as 5'-GACGTC (Aat II) to its downstream sequence within the coding region of 5' towards 3' orientation.

(7) There is also provided a library with 3'-end start codon orientation. The library of oligonucleotides consists of all possible combinations of 64 codons (TABLE 4) with a start codon, such as 5'-ATG, as 3'-end terminal codon for each oligonucleotide of a given length. As will be appreciated by one of skilled in the art, the result of this arrangement is that those oligonucleotides would preferentially hybridize to Antisense 5'-Untranslated Region (Antisense 5'-UTR) of template strand (antisense) of genomic DNA or 1st single strand of cDNA. Which is from an antisense start codon, such as 5'-CAT of antisense ORF to its downstream sequence of antisense 5'-UTR of 5' towards 3' orientation due to the fact that sequences corresponding to termination codons are specifically included.

(8) Accordingly, there is also provided a library with 5'-end antisense start codon orientation. The library of antisense oligonucleotides consists of all possible combinations of 64 antisense codons (TABLE 5) with an antisense start codon, such as 5'-CAT, as 5'-end terminal antisense codon for each antisense oligonucleotide of a given length. As will be appreciated by one of skilled in the art, the result of this arrangement is that those antisense oligonucleotides would preferentially hybridize to 5'-Untranslated Region (5'-UTR) of non-template strand (sense) of genomic DNA or mRNA or 2nd single strand of cDNA. Which is from a start codon, such as 5'-ATG of ORF to its upstream sequence of 5'-UTR of 5' towards 3' orientation due to the fact that sequences corresponding to antisense termination codons are specifically included.

(9) There is also provided a library with 5'-end stop codon orientation. The library of oligonucleotides consists of all possible combinations of 64 codons (TABLE 4) with a stop codon, such as 5'-TGA, as 5'-end terminal codon for each oligonucleotide of a given length. As will be appreciated by one of skilled in the art, the result of this

arrangement is that those oligonucleotides would preferentially hybridize to Antisense 3'-Untranslated Region (Antisense 3'-UTR) of template strand (antisense) of genomic DNA or 1st single strand of cDNA. Which is from an antisense stop codon, such as 5'-TCA of antisense ORF to its upstream sequence of antisense 3'-UTR of 5' towards 3' orientation due to the fact that sequences corresponding to termination codons are specifically included.

(10) Accordingly, there is also provided a library with 3'-end antisense stop codon orientation. The library of antisense oligonucleotides consists of all possible combinations of 64 antisense codons (TABLE 5) with an antisense stop codon, such as 5'-TCA, as 3'-end terminal antisense codon for each antisense oligonucleotide of a given length. As will be appreciated by one of skilled in the art, the result of this arrangement is that those antisense oligonucleotides would preferentially hybridize to 3'-Untranslated Region (3'-UTR) of non-template strand (sense) of genomic DNA or mRNA or 2nd single strand of cDNA. Which is from a stop codon, such as 5'-TGA of ORF to its downstream sequence of 3'-UTR of 5' towards 3' orientation due to the fact that sequences corresponding to termination codons are specifically included.

The instant invention provides a plurality of reagents for identifying genetic signatures within a sample. In one embodiment, there is provided a kit for identifying genetic signatures within a sample comprising at least one of the following:

a 5' start codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula 5'-I_S(C_S)_{n1}-3', wherein I_S represents an initiation codon in sense orientation, C_S represents an amino acid coding codon in sense orientation and n1 is an integer;

a 5' start codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula 5'-(C_A)_{n2}I_A-3', wherein I_A represents an initiation codon in antisense orientation, C_A represents an amino acid coding codon in antisense orientation and n2 is an integer;

a 5' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula 5'-(V_S)_{n3}I_S-3', wherein I_S represents an initiation codon in sense orientation, V_S represents a codon in sense orientation, and n3 is an integer;

a 5' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula 5'-I_A(V_A)_{n4}-3', wherein I_A represents an initiation codon in antisense orientation, V_A represents a codon in antisense

orientation, and n4 is an integer;

a 3' stop codon (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n5}T_S-3'$, wherein C_S represents an amino acid coding codon in sense orientation, T_S represents a termination codon in sense orientation, and n5 is an integer;

a 3' stop codon (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_A(C_A)_{n6}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation, T_A represents a termination codon in antisense orientation and n6 is an integer;

a 3' UTR (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-T_S(V_S)_{n7}-3'$, wherein T_S represents a termination codon in sense orientation, V_S represents a codon in sense orientation and n7 is an integer;

a 3' UTR (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(V_A)_{n8}T_A-3'$, wherein T_A represents a termination codon in antisense orientation, V_A represents a codon in antisense orientation and n8 is an integer;

a 5' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-R_S(C_S)_{n9}-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and n9 is an integer;

a 5' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n10}R_A-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and n10 is an integer;

a 3' restriction enzyme (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n11}R_S-3'$, wherein R_S represents a two codon (six nucleotide) restriction enzyme recognition site in sense orientation with the proviso that neither of the two codons is a termination codon, C_S represents an amino acid coding codon in sense orientation and n11 is an integer;

a 3' restriction enzyme (antisense) panel comprising a plurality of oligonucleotides,

wherein each of the oligonucleotides is represented by the formula $5'-R_A(C_A)_{n12}-3'$, wherein R_A represents a two codon (six nucleotide) restriction enzyme recognition site in antisense orientation with the proviso that neither of the two codons is a termination codon, C_A represents an amino acid coding codon in antisense orientation and $n12$ is an integer;

a between 5' and 3' (sense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_S)_{n13}-3'$, wherein C_S represents an amino acid coding codon in sense orientation and $n13$ is an integer;

a between 5' and 3' (antisense) panel comprising a plurality of oligonucleotides, wherein each of the oligonucleotides is represented by the formula $5'-(C_A)_{n14}-3'$, wherein C_A represents an amino acid coding codon in antisense orientation and $n14$ is an integer;

a N-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $R_E(N)_{n15}$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition sites in sense orientation, and $n15$ is an integer;

a C-terminal restriction enzyme peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $(N)_{n16}R_E$ -C-terminal, wherein N represents an amino acid, R_E is one of the amino acid sequences of two codon restriction enzyme recognition sites in sense orientation, and $n16$ is an integer;

a N-terminal peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $M(N)_{n17}$ -C-terminal, wherein M represents an amino acid encoded by an initiation codon, N represents an amino acid and $n17$ is an integer; and

a C-terminal peptide panel comprising a plurality of peptides, wherein each of the peptides is represented by the formula N-terminal- $(N)_{n18}$ -C-terminal, wherein N represents an amino acid and $n18$ is an integer.

As will be appreciated by one of skilled in the art, the panels may be used alone or in combination, as discussed below and as will be apparent to one of skilled in the art.

In some embodiments, R_A represents sequences of two codon restriction enzyme recognition sites in antisense orientation. In other embodiments, R_S represents sequences of two codon restriction enzyme recognition sites in sense orientation.

As will be appreciated by one of skilled in the art, $n1$ to $n16$ individually may be any positive, non-zero integer. That is, within a given kit, $n1$ may be 3 and $n2$ may be 2; alternatively, for example, both $n1$ and $n2$ may be 2. In other embodiments, $n1$ to $n16$ may individually be an integer from 1-8, from 1-7, from 1-6, from 1-5 or from 1-4.

As will be appreciated by one of skilled in the art, a single panel may consist of 2 or more oligonucleotides or peptides of one of the above-described formulae; 5 or more oligonucleotides or peptides of one of the above-described formulae; 10 or more oligonucleotide or peptides of one of the above-described formulae; 15 or more oligonucleotides or peptides of one of the above-described formulae; 20 or more oligonucleotides or peptides of one of the above-described formulae; 25 or more oligonucleotides or peptides of one of the above-described formulae; or 50 or more oligonucleotides or peptides of one of the above-described formulae.

In yet other embodiments, a panel may comprise substantially all of the oligonucleotides or peptides of one of the above-described formulae.

In other embodiments of the invention, each oligonucleotide or peptide of the panel may consist essentially of an oligonucleotide or peptide according to the specific formula for the respective panel, as discussed herein.

II. PARALLEL SEQUENCE IDENTIFYING, DEDUCING ISOLATING AND CLONING AIDED BY POLYMERASE CHAIN REACTION (PCR)

The method of Polymerase Chain Reaction (PCR) was invented by Mullis et al., US Pat. No. 4683202; U.S. Pat. No. 4683195; US Pat. No. 4800159; US Pat. No. 4965188; Taq and AmpliTaq® DNA polymerases were covered by US Pat. No. 4889818; US Pat. No. 5075216; US Pat. No. 5079352; the PCR protocols were performed as described by Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., 1990; Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, 1989.

1. The Isolation of Transcripts and the Steady-state Transcripts (Post Transcription)

(1) RNA Isolation

In one preferred embodiment, the cells in culture treated and untreated (the control) or at different time points after stimulation or at different stages and times of development or differentiated or diseased compared to the normal controls are synchronized before nuclei isolation. Nuclei of cells were isolated and RNA or and poly(A) mRNA were isolated respectively in accordance with the resources as described by Ausuble et al., Current Protocols in Molecular Biology, 1987; Chen et al., Biochemical and Biophysical Research Communications 200: 1130-1135, 1994; Chomczynski et al., Anal. Biochem. 162: 156-159, 1987.

In one preferred embodiment, the cells mentioned above are synchronized. The isolation of steady-state RNA from cells was processed respectively in accordance with the resources as described by Chomczynski et al., *Anal. Biochem.* 162: 156-159, 1987; Ausuble et al., *Current Protocols in Molecular Biology*, 1987.

(2) cDNA Synthesis

In one embodiment, the isolated total RNA from each sample was tested by the inventive RNA Quality Control Kit. 100 ng tested RNA from each sample was dissolved as the template RNA in 10ul reaction solution in an Eppendorf® tube and treated at 70°C for 5 minutes. Then it was put on ice immediately. 2ul of the First-strand cDNA Synthesis Master Solution [5x first-strand cDNA synthesis buffer (BRL), 0.1MDTT, 500uM dNTP mix, 40 U/ul RNasin® (Promega), H₂O DEPC] and 2ul of 25 uM oligo(dT)₁₈ were added subsequently. The incubation was at 37°C for 3 minutes. 40 U of M-MuLV reverse transcriptase was added and incubated at 37°C for 60 minutes. The reaction was inactivated by heating at 95°C for 5 minutes followed by chilling on ice immediately. In another preferred embodiment, 0.1 to 5 ug of total RNA isolated from each resource was reverse synthesized into cDNA as described by Ausuble et al., *Current Protocols in Molecular Biology*, 1987; Sambrook et al. *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989.

In one preferred embodiment, the reverse transcriptase employed in the experiment was Avian Myeloblastosis Virus (AMV) reverse transcriptase (Promega). In one preferred embodiment, the reverse transcriptase employed in the experiment was Moloney Murine Leukemia Virus (M-MuLV) reverse transcriptase (BioLabs). In one preferred embodiment, Oligo-(dT)₁₂ was used for priming in the reverse transcription reaction. In one preferred embodiment, Oligo-(dT)₁₅ was used for priming in the reverse transcription reaction. In one preferred embodiment, Oligo-(dT)₁₈ was used for priming in the reverse transcription reaction. In one preferred embodiment, 95% of Oligo-(dT)₁₅ was mixed with 5% hexamer primers consisting of equal amount of each distinctive 3,721 sense hexamers (two codons) oligonucleotides of the present invention in accordance with the algorithm of 61.sup.2. The mixed primers were used for priming in the reverse transcription reaction. In one preferred embodiment, the hexamer primers consist of equal amount of each distinctive 3,721 sense hexamers (two codons) oligonucleotides of the present invention used for priming in the reverse transcription reaction. Each pool of the cDNA synthesized was analyzed by the inventive cDNA quality control kit to ensure the

quality. The size of newly synthesized cDNA was analyzed on agarose gel after electrophoresis. The selection range was set at 300 bp to 6,000 bp for subsequent processes. Each of the selected cDNA pools was made into aliquots and stored at -20°C.

2. PCR

The inventive series universal oligonucleotide libraries, such as 9mer sense ORF oligonucleotide 5'-ATG orientation library which possesses 3,721 distinct oligonucleotides that include all the possible 9mer ORF sequences started from 5'-ATG, could be used directly as PCR upstream primers in conjunction with oligo-d(T)₁₂ as the downstream primer in PCR to identify, amplify, isolate and clone the gene of interest. In one preferred embodiment, the final volume of each PCR reaction is 20 ul. It contains 1 to 25 ng of cDNA, 1.5 mM MgCl₂, 50 mM KCl, 20 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 0.1 mM DTT, 150 uM each of dNTPs (dATP, dCTP, dGTP and dTTP), 0.05% Tween® -20, 10 to 25 pM of each primer and 1 to 2 units of Taq DNA polymerase in a PCR reaction tube. The thermostable DNA polymerase used in conjunction with the present invention may be selected from the group of polymerases which includes, without limiting the generality of the foregoing, Taq DNA polymerase, AmpliTaq® Gold DNA polymerase, Pfu DNA polymerase, Tfi DNA polymerase, Tli DNA polymerase, Tth DNA polymerase, Vent_R (exo⁻) DNA polymerase and Deep Vent_R (exo⁻) DNA polymerase. The analogues and modified dNTPs may be used in conjunction with the present invention which include, without limiting the generality of the foregoing, 5'-nitroindole, 3'-nitropyrrole, inosine, hypoxanthine, biotin-11-dUTP, biotin-16-dUTP, 5'-bromo-dUTP, dUTP, dig-11-dUTP and 7-deaza dGTP. In one preferred embodiment, the threshold cycle consists of denaturing temperature of 45 second at 94°C, annealing temperature of 90 second at 41 °C and extension temperature of 60 second at 72°C. The number of cycles for PCR amplification was 30, each of which consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 41°C and an extension step of 60 seconds at 72°C. The end cycle consists of 5 minutes at 72°C following by 4°C. The upstream primers were chosen from the group of 1,168 distinct 9mer oligonucleotides having 44.4% GC content, such as 5'-ATGTGCAGA and 5'-ATGTGCACA respectively. The downstream primer is oligo-d(T)₁₂. It is apparent to skilled artisans that the optimization of annealing temperatures and the concentrations of PCR primers are often determined on a case by case basis and need to be finalized experimentally and/or empirically. Generally, the annealing temperature is based on T_m (melting temperature) of the specified oligonucleotide selected as the PCR primer. T_m is related to the GC content of oligonucleotide primer. The present invention has further

classified and consolidated 3,721 distinct 9mer oligonucleotides into five categories for PCR in accordance with the GC content to facilitate determining a PCR temperature profile. The basic reference is provided herein:

(1) Denaturation:

94°C, 30sec for all.

(2) Annealing:

41°C or 41°C plus 1-4°C or 41°C minus 1-4°C for 60sec:

11.1%GC Content of 49 upstream primers, such as 5'-ATGATAATA

22.2%GC Content of the 308 upstream primers, such as 5'-ATGGAAATA

42°C or 42°C plus 1-4°C or 42°C minus 1-4°C for 60sec:

33.3%GC Content of the 820 upstream primers, such as 5'-ATGGCAATA

43°C or 43°C plus 1-4°C or 43°C minus 1-4°C for 60sec:

44.4%GC Content of the 1,168 upstream primers, such as 5'-ATGGCAGAA

44°C or 44°C plus 1-4°C or 44°C minus 1-4°C for 60sec:

55.6%GC Content of the 928 upstream primers, such as 5'-ATGGCAGCA

45°C or 45°C plus 1-4°C or 45°C minus 1-4°C for 60sec:

66.7%GC Content of the 384 upstream primers, such as 5'-ATGGCAGCC

77.8%GC Content of the 64 upstream primers, such as 5'-ATGGCCGCC

(3) Extension: 72°C 60sec for all.

(4) Cycle Number: 30

(5) Final Extension: 72°C 5min for all.

If no bands on an Agarose gel are observed, the annealing temperature might be adjusted in the range of 1°C to 4°C below the original annealing temperature and, if unwanted bands and/or several bands appeared, the annealing temperature might be adjusted in the range of 1°C to 4°C above the original annealing temperature in each subsequent optimization step. It is recommended that if the inventive 9mer, 12mer, 15mer, 18mer, 21mer, and 24mer oligonucleotides are used as the PCR primers, the range of annealing temperatures is often 37°C to 56°C. The higher the annealing temperature was increased, the more specific the PCR results obtained. Therefore, the annealing temperature has the potential to be increased as high as the extension temperature in some cases under certain conditions. The Touchdown PCR protocol is designed to identify the optimized annealing temperature for specific PCR conditions.

The Touchdown PCR protocol starts with an annealing temperature above the primer's ideal temperature. At each cycle, the annealing temperature is programmed to

decrease 1°C until reaching the targeting annealing temperature. In one preferred embodiment, 9mer 5'-ATG oriented ORF sense oligonucleotides of present invention such as 5'-ATGGCAGCA had three consecutive universal bases such as 5'-nitroindoles covalently added at each of their 5'-ends to form 12mer oligonucleotides. The 12mer oligonucleotides were then used as PCR primers. For example, the threshold cycle consists of a denaturing step of 45 seconds at 94°C. The second cycle consists of denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 50°C and an extension step of 60 seconds at 72°C. The third cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 49°C and an extension step of 60 seconds at 72°C. The fourth cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 48°C and an extension step of 60 seconds at 72°C. The fifth cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 47°C and an extension step of 60 seconds at 72°C. The sixth cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 46°C and an extension step of 60 seconds at 72°C. The seventh cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 45°C and an extension step of 60 seconds at 72°C. The eighth cycle consists of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 44°C and an extension step of 60 seconds at 72°C. The number of cycles for subsequent PCR amplification was 30, with each cycle consisting of a denaturing step of 30 seconds at 94°C, an annealing step of 90 seconds at 44°C and an extension step of 60 seconds at 72°C. The final cycle consists of 5 minutes at 72°C following by 4°C. The Touchdown PCR protocol was employed to determine the corresponding optimized annealing temperature in the above-mentioned embodiment as described by Don et al., *Nucleic Acids Res.* 19: 4008, 1991.

The present invention uses Agarose gel as the principal display system. 0.7% to 1.4% Agarose gels display and separate PCR products and cDNA fragments from 0.1KB to 23 KB. Using 3% Agarose 1000 in 1X TAE buffer, one could possibly display and separate cDNA fragments with as little as 10 bp difference in size (Invitrogen Catalogue, 2002). Furthermore, cDNAs and PCR products can be more readily isolated from Agarose gels than Polyacrylamide gels and are also much easier to amplify and clone in subsequent procedures than their counterparts isolated from Polyacrylmide. The present invention emphasizes that each RNA/cDNA sample must have the corresponding normal RNA/cDNA control samples which include positive and negative controls respectively at

the same time in the PCR reaction. Furthermore, at least one control for each testing sample is suggested. Each distinctive PCR should include all the necessary controls known in the art. Usually, one PCR testing product was loaded in parallel with the corresponding PCR control product(s) on two or more adjacent lanes on an agarose gel prior to electrophoresis. As a result of this arrangement, the results clearly demonstrate which of the PCR amplicons was unique for which samples.

3. Sequence Deducing Aided by PCR

If one positive PCR amplicon was identified after the primary PCR aided screening by using a pair of PCR primers such as, 5'-ATGGCAGCA and oligo-d(T)₁₂, the sequence such as 5'-ATGGCAGCA, was identified without further sequence analysis. Then, three consecutive universal bases such as 5'-nitroindoles were added covalently and sequentially at the 3'-end of 5'-ATGGCAGCA to elongate to 5'-ATGGCAGCA555, a 12mer oligonucleotide (5 represents 5'-nitroindole). Each of 61 identical 5'-ATGGCAGCA555 were then joined covalently with a distinct codon from the group of 61 codons such as 5'-GCA respectively to form 61 distinct 15mer oligonucleotides such as 5'-ATGGCAGCA555GCA. Using the 61 distinct 15mer oligonucleotides as the upstream PCR primers in conjunction with oligo-d(T)₁₈ as the common downstream PCR primer, 61 distinct PCRs with corresponding controls were run on a PCR thermal cycler. If one positive PCR amplicon was identified after the secondary PCR aided screening from these 61 distinctive sets of PCR with 61 pairs of distinctive PCR primers, the positive 15mer oligonucleotide started from 5'-ATG with three consecutive 5'-nitroindoles within the sequence would consequently be identified and determined. Since three consecutive 5'-nitroindoles occupied the positions within that 15mer oligonucleotide from position 10 to 12 in 5' towards 3' orientation; further determination of positions 10, 11 and 12 of that 15mer oligonucleotide is needed. Using each one of the 61 codons to replace the "codon" consisting of three consecutive 5'-nitroindoles, another 61 distinctive 15mer oligonucleotides without any 5'-nitroindole within the sequences were formed. Using the 61 newly formed distinctive 15mer oligonucleotides as the upstream PCR primers in conjunction with oligo-d(T)₁₈ as the common downstream PCR primer, another 61 distinctive set of PCRs with corresponding controls were run again on a PCR thermal cycler. If one positive PCR amplicon was identified after the third PCR aided screening, the entire sequence of the positive 15mer sense ORF sequence started from 5'-ATG was identified and determined. One ordinary skilled in the art could either use the deduced 15mer oligonucleotide in conjunction with oligo-d(T)₁₈ to clone the gene of interest or

further deduce out an 18mer oligonucleotide to increase the specificity of the primer and subsequent PCR reaction by repeating the cycle once more. The present invention presents a strategy of sequence deducing by PCR. The inventive design of PCR primer is based on a non-bias system constructed in accordance with the algorithm of $61.\text{sup.}(n-1)$. PCR aided sequence analysis and deducing is simple, straightforward and time saving.

4. Real-time PCR

Real-time PCR monitors the energy emission of excited fluorescence during PCR amplification whereas conventional PCR is focused on endpoint detection. One of the major advantages of Real-time PCR is the wider dynamic spectrum which is $10.\text{sup.4}$ times than the conventional PCR. However, Real-time PCR has its own criteria for the selection of primers. The inventive oligonucleotide library is universal and covers any and all possible sense ORF sequences downstream 5'-ATG regardless of the biological species. Among the 3,721 distinctive 9mer oligonucleotides, there is no G at the 5'-end and 86.9% are without runs of three or more Gs or Cs or GCs at the 3'-end. Additionally, 3,364 of the 3,721 distinctive 9mer oligonucleotides have GC content in the ideal range of 30-80%. Obviously, the elongated derivative oligonucleotides from the inventive 9mer oligonucleotides with length of 15mers and above would also be the ideal resource for primer selection for Real-time PCR.

5. DD-RT-PCR

The study of the probabilities of priming site in DNA of 45,000 base pair indicated that $P(0)$, the probability of no priming site of 12mer oligonucleotides, is 0.995. $P(1)$, the probability of exactly one priming site of 12mer oligonucleotides, is 0.005. $P(>1)$, the probability of more than one priming site of 12mer oligonucleotides, is $<10^{-4}$ ($<10.\text{sup.-4}$) (Studier, Proc. Natl. Acad. Sci. U.S.A. 86: 6917-6921, 1989). Theoretically, an oligonucleotide with the length of 15mers to 18mers and/or above could be able to detect a single copy gene from the human genomic DNA. In practice, a 12mer oligonucleotide is capable of detecting an mRNA molecule. Long oligonucleotide (> 10 mers) may decrease the specificity if its binding affinity is high (Herschlag et al., Proc. Natl. Acad. Sci. U.S.A. 88: 6921-6925, 1991). Due to the increasing probability of forming secondary structure(s) with increasing length of an oligonucleotide, a short oligonucleotide has distinct advantages over a longer one. Particularly, a 9mer oligonucleotide has been proven as an effective primer in PCR (Williams et al., Nucleic Acids Res. 18: 6531-6535, 1990). The inventive 9mer PCR primer was designed in accordance with the algorithm of $61.\text{sup.}(n-1)$. The codon-based PCR primer design has eliminated the nonsense codon completely in the

primer sequences whereas the nucleotide-based PCR primer design, such as DD-RT-PCR does not. The present invention took a different approach than conventional DD-RT-PCR by targeting the 5'-end of ORF. It is known in the art that targeting 3'-end regions of genes often contributes to the noise-level in the subsequent cloning process. Another advantage of the present invention over DD-RT-PCR is the use of agarose gels as the display system instead of polyacrylamide gels. The isolated cDNAs and PCR products from Agarose gel are much easier to amplify than their counterparts from polyacrylamide in subsequent PCR and cloning procedures. Overall, the inventive PCR primers were selected systematically from relevant oligonucleotide libraries instead of by random and arbitrary selection. The present invention presents a precise strategy for both PCR and cloning instead of shotgun strategy as per DD-RT-PCR.

Some of the embodiments and examples mentioned above are generally known in the art and the artisan with the ordinary skill will recognize that the scope of the present invention is not limited to and by those embodiments and examples.

III. CONSTRUCTIONS OF THE NORMALIZED COMPREHENSIVE cDNA LIBRARY

In a typical mammalian cell, mRNA molecules can be classified in three major categories: abundant, intermediate and scarce. For the abundant class, each mRNA molecule may have 12,000 copies per cell. There are about 300 copies for each average intermediate mRNA molecule while about 15 copies for each scarce mRNA molecule. However, the scarce category has the most biological significance. It contains about 11,000 or more different mRNA molecules while the intermediate class has about 500 and the abundant class has more than 4 (Alberts et al. Molecular Biology of The Cell, 528, 1989). To screen and clone mRNA molecules of interest, cDNA synthesis and cDNA library construction is often the first priority. In practice, it is usually the first crucial step in the entire experimental procedure. However, some of the cDNA libraries have been found to contain certain distortions regarding gene expression, gene numbers and kinds. This is a problem in current cDNA library construction. The distortion is mainly reflected as the lack of 5'-termini of cDNA molecules, under representing scarce mRNA molecules and over representing abundant mRNA molecules. Without resolving this issue, the fidelity of cDNA library could be in question. Using such cDNA libraries may mislead or fail the subsequent screening and cloning. The results may be undesirable in unforeseen ways.

Efforts have been made to reduce the over representation of the abundant class of

mRNA in cDNA library construction. They are known in the art such as Ko, M.S.H. (Ko, Nucleic Acids Res. 18: 5705-5711, 1990); Patanjali et al. (Patanjali et al., Proc. Natl. Acad. Sci. U.S.A. 88: 1943-1947, 1991); Sasaki et al. (Sasaki et al., Nucleic Acids Res. 22: 987-992, 1994); Soares et al. (Soares et al., Proc. Natl. Acad. Sci. U.S.A. 91: 9228-9232, 1994); Pouzyrev et al. (Pouzyrev et al., U.S. Pat. No. 6,372,436, 2002) and Li et al. (Li et al., U.S. Pat. No. 6,399,334, 2002). Reducing the relative amount of abundant mRNA and subsequently increasing the ratio of the scarce mRNA is one aspect of the normalization of cDNA library. Other aspects of cDNA library construction remain unexamined, for example, how does one represent mRNA molecules at equal and equivalent amount under different conditions and how does one ensure that most of cDNA molecules in the cDNA libraries have a start codon at their 5'-terminus.

The present invention has resolved these problems. The present invention uses 3,721 distinctive individual 9mer oligonucleotides with 5'-ATG at each 5'-terminal as the PCR upstream primers and oligo-d(T)₁₅ as the downstream primers to synthesize cDNA in 3,721 individual vials to set 3,721 distinctive PCRs which could be performed in GeneAmp® PCR System 9600. The cDNA synthesis could be completed within one week. The synthesized and amplified cDNA fragments of each PCR could be detected, selected and isolated precisely from 1% Agarose gel stained by Ethidium Bromide. The preferred selection standard of cDNA molecules for subsequent cloning procedure is set at a minimum size of 500 bp. Another preferred selection standard is set at the minimum size of 400 bp. Another preferred selection standard is set at the minimum size of 300 bp. Another preferred selection standard is set at the minimum size of 200 bp. The selected cDNA fragments are then excised from the gel and extracted. Each of the selected cDNA fragments could be either re-amplified by its original set of PCR primers for subsequent cloning or cloned directly by TA cloning technology known in the art. The vector which contained the insert could be a plasmid, cosmid, phagemid, phage DNA or any other DNA molecules capable of propagating autonomously with selection marker(s), such as tetracycline or ampicillin resistance in a host cell as known in the art. For example, Lambda gt11 or Lambda ZapII could be used for this purpose. Those inventive cDNA libraries could be further used as probes in the construction of cDNA microarrays and Dot-blots. Alternatively, cDNA could be synthesized through reverse transcription by using 3,721 distinctive 9mer 5'-ATG oriented oligonucleotides as primers with or without oligo-d(T) to reverse transcribe mRNA into double stranded cDNA. The trace amounts of each synthesized cDNA molecule are further amplified by PCR to make the amount of each

cDNA sufficient for subsequent insertion and ligation into a suitable vector.

IV. PARALLEL GENE SEQUENCES DEDUCING AND CLONING AIDED BY DOT-BLOTTING AND PCR

Dot-blotting has been widely used for a long time. The greatest advantage of Dot-blotting screening over DNA microarrays is that it is manually manageable and economical. Furthermore, radioisotope labelling is at least 10 times more sensitive than fluorescence labelling. Dot-blots (nylon and nitrocellulose) may be reused after stripping the hybridization signals. Dot-blotting is a well-established standard laboratory method with simple and straightforward procedures. The performance does not require expensive equipment such as a laser scanner. The operation is simpler and the cost is much lower than DNA microarray technology whereas being a miniaturized DNA microarrays. Lennon et al. have revealed how Dot-Blotting Hybridization could be employed to exploit the data of genome programs on a large-scale manner (Lennon et al., Trends Genet.7: 314-317, 1991). To design and develop a new generation of Dot-Blotting Hybridization protocols with codon-based methods is one of the aspects of the present invention.

1. Oligonucleotide Dot-blotting

The present invention of oligonucleotide libraries such as, 9mer sense ORF oligonucleotide library with 5'-ATG orientation provides a unique miniaturized systematic working platform on the surface of nylon or nitrocellulose filter or glass alone to identify, deduce and determine the sequence of the two codons immediately downstream and adjacent to the 5'-ATG of the gene of interest. The inventive 3,721 distinct 9mer oligonucleotides include all the possible 9mer sequences downstream and including 5'-ATG for all species. It provides a non-bias and full-range molecular screening system. Every one of the 3,721 distinct 9mer ORF oligonucleotides with 5'-ATG was immobilized at a distinct, precise position on the surface of nylon or nitrocellulose filter or glass. Each oligonucleotide can be deposited on the pre-moistened surface of nylon or nitrocellulose filter or glass by using a Pipetman® manually and/or by a robot such as Array Spotter (Amersham Pharmacia Biotech, 2001). In some embodiments, in accordance with GC content of the oligonucleotide, Dot-blot was classified into several sub dot-blots. In these embodiments, each sub dot-blot has all its oligonucleotides with the same GC content to provide uniformity in subsequent hybridization. In one embodiment, 1,344 spots (32 by 42) were occupied by 928 distinct immobilized 9mer ORF oligonucleotides (55.6%GC content) and occupied by 1,168 distinct immobilized 9mer ORF oligonucleotides (44.4%GC content)

separately on an 8.5 inch by 11 inch nylon or nitrocellulose filter. Each spot has a diameter of 2 mm. The total of 3,721 distinct 9mer oligonucleotides plus control spots could be immobilized on seven sheets of nylon or nitrocellulose grouped according to 77.8%, 66.7%, 55.6%, 44.4%, 33.3%, 22.2% and 11.1% GC content respectively. In another embodiment, the 928 distinct 9mer ORF oligonucleotides (55.6% GC content) plus control spots were spotted on five pre-wetted 5 inch by 10 inch sheets of nylon or nitrocellulose and then immobilized by UV cross link in a moisture chamber or baked at 80°C for 90 minutes. The filters were sealed in clean transparent plastic bags and stored in -20°C.

2. cDNA Dot-blotting

The present invention uses 3,721 distinctive individual 9mer oligonucleotides with 5'-ATG at each 5'-terminal as the PCR upstream primers and oligo-d(T)₁₅ as the downstream primers to synthesize 3,721 distinct sets of cDNA in 3,721 distinct individual PCRs as described above. cDNA was isolated and purified in accordance with the standard protocol (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, 1989). The isolated and purified cDNA were aliquoted, recorded and stored. Each distinct cDNA is deposited and immobilized at the same position as its corresponding oligonucleotide was on the pre-moistened surface of nylon or nitrocellulose filter or glass. The dot-blot with immobilized cDNA arrays were prehybridized in the hybridization solution (50% formamide, 4.times.SSC, 5XDenhardt's reagent, 0.2%SDS, 0.1NaPPi and 100 ug /ml Salmon sperm DNA) 4 hours at 42° C. The ³²P or ³³P-labeled RNA was denatured at 65° C. The labeled RNA with 2-4X10⁶ cpm per ml hybridized with the filter at 42°C for 12 hours or overnight. The concentration of labeled RNA should be maintained at least in the range of 10 to 100 nM. Another preferred embodiment is to convert RNA isolated from nuclei into cDNA by reverse transcription. The cDNA probes were used for labeling and then hybridization with the Dot-blot. The radioactive intensity of labeled cDNA probes was at least 2-4X10⁶ cpm per ml in the hybridization solution. The Dot-blot was washed by 0.1XSSC and 0.5% SDS for 5 -10 minutes at 50°C to 65°C followed by further washing with a washing solution of 2XSSC and 0.2% SDS for 10-30 minutes at 50°C. In one preferred embodiment, the stringency increased by using 0.2XSSC and 0.2% SDS for 5 minutes at 50°C. The Dot-blot was then placed and sealed in separate clean transparent plastic bags. The Dot-blot was exposed to Kodak® film with intensifying screens at -70°C. After the exposure to Kodak® film, all the black spots with density higher than 2 to 4 times background were identified and their corresponding sequences were determined by DNA sequence analysis using the stored

cDNA samples. In another embodiment, the Dot-blot samples were exposed to a storage phosphor screen such as Fuji® STIII at 4° C for 12 hours wrapped with alumni foil and placed in the dark. The screens were scanned by PhosphorImager® 400A. The digital images were transferred to, recoded and analyzed on a Sun-4® workstation.

With the inventive oligonucleotide libraries, cDNA Dot-blotting as well as DNA microarrays are capable of employing the unknown cDNA probes which were derived from the known inventive oligonucleotides. Using those "unknown" cDNA, one skilled in the art can identify the gene expression patterns and the responsible sequences of interest through hybridization. Then using the known corresponding oligonucleotide(s) as the upstream primer(s) with oligo-d(T) as the downstream primer, the full length of gene of interest could be identified and cloned in accordance with means known in the art. Without the inventive oligonucleotide libraries and methods, neither cDNA Dot-blotting nor current DNA microarrays can accomplish this.

3. Transcription Detection

The test cells and control in culture are synchronized before nuclei isolation. Nuclei of cells were isolated as described by Ausuble et al., Current Protocols in Molecular Biology, 1987; Chen et al., Biochemical and Biophysical Research Communications 200: 1130-1135, 1994.

Radioisotope, such as ^{32}P or ^{33}P -labeled, nascent RNA was extracted by phenol-chloroform from both testing group and the controls separately. The labeled RNA was precipitated in ethanol and purified by passage through a Sephadex® G-50 fine spin column.

The dot-blot samples with immobilized 9mer oligonucleotide arrays were prehybridized by incubation with the hybridization solution (0.4M Na_2HPO_4 , 0.5 % sodium lauroyl sarcosine) for 10 minutes at 25°C. The ^{32}P or ^{33}P -labeled RNA was denatured at 65°C. The radioactive intensity of the labeled RNA is at least 2×10^6 cpm per ml in the hybridization solution in which the filter was immersed. The hybridization was performed at 12°C - 15°C for 4 hours or under 12°C overnight. The concentration of labeled RNA was in the range of 10 to 50 nM. In another preferred embodiment, RNA isolated was converted into cDNA by reverse transcription. The cDNA probes were labeled and hybridized with the Dot-blot samples. The radioactive intensity of the cDNA probes was at least 2×10^6 cpm per ml in the hybridization solution. The Dot-blot samples were washed with 6XSSC three times for 10 minutes at 8°C to 14°C and then washed 4 times with SSC at room temperature for 15 minutes.

The Dot-blot were sealed in separate clean transparent plastic bags. The Dot-blot were exposed to Kodak® film with intensifying screens at -70°C. In another embodiment, the Dot-blot were exposed to a storage phosphor screen such as Fuji® STIII at 4°C for 12 hours wrapped with alumni foil and placed in the dark. The screens were scanned by a PhosphorImager® 400A. The digital images were transferred to, recoded and analyzed on a Sun-4® workstation.

4. Steady-state Transcription (Post Transcription) Detection

The test cells and control cells in culture are synchronized before the isolation of steady-state RNA. Steady-state RNA was extracted from the cells as described by Chomczynski et al., Anal. Biochem. 162: 156-159, 1987. The hybridization and detection process is as described above.

5. Parallel Sequence Deducing for Designing PCR Primers

Since each sequence of the 3,721 distinct 9mer ORF oligonucleotides with 5'-ATG was known and its precise position on the surface of nylon or nitrocellulose filter or glass was known as well, after exposure to Kodak® film, any black spots with density higher than 2 to 4 times background could be identified. Therefore, the targeting sequences were easily identified from the record. In one embodiment, using one of those identified 9mer oligonucleotides as upstream primer at each time per PCR and oligo-d(T)₁₂ as the downstream primer, the full-length ORF sequences of the gene(s) of interest could be identified and cloned by means known in the art. In another preferred embodiment, three consecutive 5-nitroindoles were covalently added as an adaptor to each of the 5'-ATG oriented upstream oligonucleotide primers in order to elongate the 9mers to 12mers. In another preferred embodiment, covalently adding four consecutive 5-nitroindole as an adaptor to each of the 5'-ATG oriented upstream oligonucleotide primers elongated the 9mers to 13mers. In another preferred embodiment, covalently adding five consecutive 5-nitroindoles as an adaptor to each of the 5'-ATG oriented upstream oligonucleotide primers elongated the 9mers to 14mers. Together with oligo-d(T)₁₄ as the downstream oligonucleotide primer, those pairs of PCR primers could be used in the subsequent PCRs. In another preferred embodiment, covalently adding six consecutive 5-nitroindoles as an adaptor to each of the 5'-ATG oriented upstream oligonucleotide primers elongated the 9mers to 15mers. Together with oligo-d(T)₁₆ as the downstream oligonucleotide primer, those pairs of PCR primers could be used in the PCRs.

In other embodiments, the libraries are advantageously used alone or in

combination to clone genes of interest. For example, the 5'-start codon such as 5'-ATG oriented ORF library and 3'-start codon such as 3'-GTA (5'-ATG) oriented 5'-UTR library are used in tandem in hybridization to determine potential start codons within a gene of interest. Using the example above, the 5'-start codon oriented ORF library will indicate all possible potential start codon sites through hybridization with the target DNA, cDNA and mRNA samples, and the 3'-start codon oriented 5'-UTR library will indicate all possible potential start codon sites as well. The identity of the specific start codon can then be determined by any of a variety of means known in the art, for example, when 9mers are utilized, a 15mer oligonucleotide corresponding to the start codon of the gene, centered on the 5'-ATG, in other words, XXXXXXATG(NN), wherein "N" is deduced from a codon encoding an amino acid could be deduced experimentally. This oligonucleotide can then be used in combination with an oligo-d(T)₁₅₋₁₈ as a pair of PCR primers to clone the full-length gene of interest. The PCR amplified cDNAs with their corresponding controls were loaded on 1 to 1.4% Agarose gel paralleled for electrophoresis. The results of electrophoresis were recorded by photograph under UV light. The cDNA fragments of interest, which were usually >500 bp in size, were marked and excised from the Agarose gel for cDNA extraction. If there were multiple bands of interest in a lane compared with its paralleled control sample, the cDNA fragments with the largest size were often given preferentially selected as the candidate clones used for subsequent cloning and sequencing analysis.

V. PARALLEL SEQUENCE DEDUCING AND CLONING AIDED BY DNA MICROARRAYS AND PCR

Regarding the construction of high-density oligonucleotide microarrays, there are two major formats. One is microarrays made by pre-synthesized oligonucleotides or oligonucleotide analogues; the other is microarrays made by in-situ synthesized oligonucleotides or oligonucleotide analogues.

1. Oligonucleotide Pre-synthesis and Immobilization

In one preferred embodiment, the pre-synthesis of oligonucleotides was processed by the phosphoramidite methods such as Caruthers et al., *Nucleic Acids Res. Symp. Ser.* 7: 215-223, 1980; Beaucage et al., *Tetrahedron Lett.* 22: 1859-1862, 1981; McBride et al., *Tetrahedron Lett.* 24: 245-248, 1983; and Beaucage et al., *Tetrahedron Lett.* 48: 2223-2311, 1992. In one preferred embodiment, the pre-synthesis of oligonucleotides was

processed by the H-phosphonate methods, such as Garegg et al., Chem. Scripta 25: 280-282, 1985; Garegg et al., Chem. Scripta 26: 59-62, 1986; Garegg et al., Tetrahedron Lett. 27: 4051-4054, 1986; Froehler et al., Nucleic Acids Res., 14: 5399-5407, 1986; Froehler et al., Tetrahedron Lett. 27: 469-4472, 1986; Froehler et al., Tetrahedron Lett. 27: 5575-5578, 1986. In another preferred embodiment, the pre-synthesis of oligonucleotides was processed by an automated nucleic acid synthesizer, such as ABI 381-A, ABI 391, ABI 392, ABI 3900 and Expedite® 8909 Nucleic Acid Synthesizer of PE Applied Biosystems at a 0.2µm scale using standard protocols in accordance with the manual of the manufacturer. Prior to the coupling step on the solid phase, the pre-synthesized oligonucleotides were purified, desalted and lyophilized at different grades of purity in accordance with different purposes such as PCR®-grade (ethanol precipitation to remove the salt), Probe-grade (purified by HPLC) or and Gene-synthesis-grade (purified by polyacrylamide gel electrophoresis). Such purification is known to those of skilled in the art. In one embodiment, the pre-synthesized oligonucleotide was without universal bases such as 5'-nitroindole at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had one universal base such as 5'-nitroindole at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had two consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had three consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had four consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had five consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the pre-synthesized oligonucleotide had six consecutive universal bases such as 5'-nitroindoles at its 5'-end.

Generally, immobilizing the pre-synthesized oligonucleotides requires the pre-treatment of the surfaces of glass and polystyrene with poly L-lysine in order to be ready for the immobilization. The immobilization was generally processed in a humid chamber at 37°C for 12 hours or overnight. The pre-synthesized oligonucleotides for immobilization require pre-treatment to chemically modify either amino groups or sulfhydryl groups of the pre-synthesized oligonucleotide before the attachment to the solid phase. In one preferred embodiment, the pre-synthesized oligonucleotides were covalently tethered to the surface of the solid support such as pre-synthesized oligonucleotides with either a 5'-end or 3'-end amine modification covalently tethered to epoxysilane monolayer of the glass surface by

forming amine linkage. The glass slides were then washed in distilled water and stored at 4°C to be ready for hybridization as described by Beattie et al., *Molecular Biotech.* 4: 213-225, 1995.

In another preferred embodiment, the pre-synthesized oligonucleotides with 5'-phosphate modification were reacted with imidazole to form 5'-phosphormidazolidine which was in turn coupled covalently with the surface through a phosphormidate bond as described by Chu et al., *Nucleic Acids Res.* 11: 6513-6529, 1983.

In one preferred embodiment, the pre-synthesized oligonucleotide possesses a 5' amino group. The said 5' amino group was obtained by using the reagent N-trifluoroacetyl-6-aminohexyl-2-cyanoethyl N', N'-diisopropyl-phosphoramidite (PE Applied Biosystems). The clean microscope glass slides were treated with 1% 3-aminopropyl trimethoxysilane with 95% acetone/distilled H₂O, subsequently washed by acetone and dried. The pre-treated glass slides were further treated with 0.2% 1,4-phenylene diisothiocyanate (PDC) in 10% pyridine/dimethyl formamide and washed by methanol and acetone as described by Guo et al., *Nucleic Acids Res.* 22: 5456-5465, 1994.

In one preferred embodiment, biotinylated pre-synthesized oligonucleotides are immobilized to streptavidin/avidin covalently coated surface by non-covalent attachment as described by Holmstron et al., *Anal. Biochem.* 209: 278-283, 1993.

2. Oligonucleotide In-situ Synthesis and Immobilization

In one preferred embodiment, a solid support such as the surface of silicon was coated with photo-removable groups, and in-situ synthesis of oligonucleotides was processed by photolithographic methods. First, a specific defined discrete position was exposed to light for activation. Then, an activated nucleotide monomer was flooded to the surface for specific chemical coupling at that defined position. The array was made by successive cycles of deprotecting defined positions of the array by photolithography such as Fodor et al., *Science* 251: 767-773, 1991; Pease et al. *Proc. Natl. Acad. Sci. U.S.A.* 91: 5022-5026, 1994; Lockhart et al., *Nature Biotech.* 14: 1675, 1996; Pirrung et al., *US Pat. No. 5,143, 854*, 1992; Fodor et al., *US Pat. No. 5,445, 934*, 1995; Fodor et al., *US Pat. No. 5,510, 270*, 1996; Fodor et al., *US Pat. No. 5,800, 992*, 1998.

In another preferred embodiment, a flexible linker with a hydroxyl group was coupled to the surface of glass plate via glycidoxypyrpyl silane and in-situ synthesis of oligonucleotides was initiated from the bound hydroxyl group. All the procedures were processed in accordance with methods as described by Southern et al., *Genomic* 13:

1008-1017, 1992; Maskos et al., Nucleic Acids Res. 20: 1679-1684, 1992; Southern et al., Nucleic Acids Res. 22: 1368-1373, 1994.

In another embodiment, in-situ synthesis of oligonucleotides and deposition on the perfluorinated hydrophobic surface of silicon dioxide was processed by a robotic liquid dispenser such as ink-jet printer heads controlled by a piezoelectric droplet generator as described by Blanchard et al., Biosensors & Bioelectronics 11: 687-690; Wallace et al., U.S. Pat. No.4, 812,856; Hayes et al., U.S. Pat. No.5, 053,100.

In one embodiment, the in-situ synthesized oligonucleotide was without a universal base such as 5'-nitroindole at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide had one universal base such as 5'-nitroindole at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide had two consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide had three consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide had four consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide had five consecutive universal bases such as 5'-nitroindoles at its 5'-end. In another embodiment, the in-situ synthesized oligonucleotide was with six consecutive universal bases such as 5'-nitroindoles at its 5'-end.

3. 5'-terminal Sense Sequence Deducing Oligonucleotide Arrays and PCR Aided Cloning

The present invention of oligonucleotide libraries such as, 9mer or 12mer or 15mer sense ORF oligonucleotide libraries with 5'-ATG orientation provide a unique probing system to identify, deduce and determine the sequence of the first three consecutive codons from 5'-ATG with 9mers or the first four consecutive codons from 5'-ATG with 12mers or the first five consecutive codons from 5'-ATG with 15mers of the gene of the interest respectively. The inventive 9mer or 12mer or 15mer sense ORF oligonucleotide libraries of 5'-ATG orientation possesses 3,721 or 226,981 or 13,845,841 distinct oligonucleotides respectively which include any and all the 9mers or 12mers or 15mers possible starting from 5'-ATG in almost any cell, tissues, organs and individual organism under most of conditions on the earth. It is a non-biased and full-range 5'-terminal sequence screening system.

In accordance with the GC content of the oligonucleotide, high density oligonucleotide arrays were classified into several sub arrays. Each sub array has all its oligonucleotides with the same GC content to allow for uniform hybridization conditions. In one preferred embodiment, 928 distinct 9mer ORF oligonucleotides (55.6%GC content) were immobilized in a sub oligonucleotide array. 1,168 distinct immobilized 9mer ORF oligonucleotides (44.4% GC content) consist of another sub oligonucleotide array. The total of the 3,721 distinct 9mer oligonucleotides plus control sequences could be immobilized and distributed on seven sub oligonucleotide arrays sorted according to 77.8%, 66.7%, 55.6%, 44.4%, 33.3%, 22.2% and 11.1% GC content respectively. In another preferred embodiment, before immobilization on the arrays, each of the above 3,721 distinct 9mer oligonucleotides was covalently coupled to an adaptor consisting of three consecutive universal bases such as 5'-nitroindoles at its 5'-end to form 3,721 distinct corresponding 12mer oligonucleotides. Subsequently, the distinct 12mer oligonucleotides were further classified by GC content and immobilized in accordance with the GC content classification onto the corresponding sub arrays. In another preferred embodiment, before immobilization on the arrays, each of the above 3,721 distinct 9mer oligonucleotides was covalently coupled to an adaptor consisting of three consecutive universal bases such as 5'-nitroindoles at its 5'-end to form 3,721 distinct corresponding 12mer oligonucleotides. Subsequently, the distinct 12mer oligonucleotides were further classified by GC content and immobilized in accordance with the GC content onto the corresponding sub arrays. In another preferred embodiment, before immobilization on the arrays, each of the above 226,981 distinct 12mer oligonucleotides was covalently coupled to an adaptor consisting of three consecutive universal bases such as 5'-nitroindoles at its 5'-end to form 226,981 distinct corresponding 15mer oligonucleotides. Subsequently, the distinct 15mer oligonucleotides were further classified by GC content and immobilized in accordance with the GC content classification onto the corresponding sub arrays. In another preferred embodiment, before immobilization on the arrays, each of the above 13,845,841 distinct 15mer oligonucleotides was covalently coupled to an adaptor consisting of three consecutive universal bases such as 5'-nitroindoles at its 5'-end to form 13,845,841 distinct corresponding 18mer oligonucleotides. Subsequently, the distinct 18mer oligonucleotides were further classified by GC content and immobilized in accordance with the GC content classification onto the corresponding sub arrays. After hybridization with RNA/cDNA/DNA sample(s), the positive position(s) was identified. Therefore the candidate positive sequence(s) of oligonucleotide(s) with the length of 3

codons (9mers), 4 codons (12mers), 5 codons (15mers) and 6 codons (18mers) were identified and determined. Using those oligonucleotides as the upstream primers together with oligo-d(T)₁₂ or oligo-d(T)₁₅ or oligo-d(T)₁₈ respectively as the downstream oligonucleotide primer, the sequence(s) of the genes of interest could be identified, amplified and eventually cloned by means known in the art. The PCR and cloning procedures were carried out as described by Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., 1990; and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, 1989.

4. 3'-terminal Sense Sequence Deducing Oligonucleotide Arrays and PCR Aided Cloning

As the counterpart of 5'-terminal sense sequence deducing strategy, the present invention of oligonucleotide libraries such as, 9mer sense ORF oligonucleotide library with 3'-stop codon such as 3'-AGT (5'-TGA) or 3'-GAT (5'-TAG) or 3'-AAT (5'-TAA) orientation provide a unique probing system to identify, deduce and determine the sequence of the first three consecutive codons upstream of 5'-TGA or 5'-TAG or 5'-TAA at the 3'-end of the target sequence. As discussed above, the high density oligonucleotide arrays with 3'-stop codon orientation were also classified into several sub arrays based on GC content.

After hybridization with RNA/cDNA/DNA sample(s), the positive position(s) was identified. The candidate positive sequence(s) of oligonucleotide(s) such as, 9mer oligonucleotide(s) was/were identified and determined. Using those oligonucleotides as the common downstream primer(s) together with any pre-determined sequence such as promoter, consensus sequence or motif sequence respectively as the upstream oligonucleotide primer, the sequence(s) of a gene family of interest could be identified, amplified and eventually cloned in the subsequent PCR aided cloning procedures. Those methods are well known to one of ordinary skilled in the art.

5. 5'-Two-codon of Restriction Enzyme Recognition Sense Sequence

Oriented Sequence Deducing Oligonucleotide Arrays and PCR Aided Cloning

The present invention of oligonucleotide libraries such as, 12mer sense ORF oligonucleotide library with 5'-Two-codon of Restriction Enzyme Recognition Sense Sequence orientation provide a unique probing system to identify, deduce and determine the sequence of the first four consecutive codons downstream of a 5'-Two-codon Restriction Enzyme Recognition Sense Sequence of the target sequence through

hybridization. All procedures and strategies for 5'-terminal Sense Sequence Deducing Oligonucleotide Arrays and PCR Aided Cloning are applicable herein.

6. The Complete Sense Sequence Deducing and Determining Oligonucleotide Arrays

Regarding the DNA Sequencing by Hybridization, Chetverin et al. proposed that nucleic acids could be sequenced by hybridizing them to oligonucleotide arrays. (Chetverin et al., Bio/Technology 12: 1093-1099, 1994) (Barinaga Science 253:1489, 1991) In one preferred embodiment, the present invention presents a 9mer ORF sense oligonucleotide library between 5'- and 3'- having 5'-end orientation which consists of 226,981 distinct 9mer oligonucleotides. Those 9mer oligonucleotides were further classified into seven categories in accordance with GC content. Each category of the oligonucleotides was used to construct sub-arrays for DNA sequencing analysis. The sequence(s) could be determined by hybridization with a single suitable labeled target to the each of the seven distinctive 9mer oligonucleotide sub-arrays under individual optimized hybridization conditions per sub-array as described by Khrapko et al., DNA Sequencing and Mapping, 1:375-388, 1991. The oligonucleotide arrays formation, hybridization conditions, signal detection and sequence analysis were as described by Southern et al., Genomics 13: 1008-1017, 1992; and Macevicz, U.S. Pat. No. 5,002,867, 1991.

7. The Gene Expression Determining cDNA Arrays and PCR Aided Cloning

Using cDNA Arrays usually produces much less cross-hybridization than Oligonucleotide Arrays. Moreover, once a cDNA library is obtained, it can be used as a resource for duplication. Therefore, using cDNA arrays could be potentially much more economical than oligonucleotide arrays. However, the present invention has established the corresponding relationship between the two. A specific cDNA library could be derived and produced from the original corresponding oligonucleotide library as described in the section of Constructions of the Normalized Comprehensive cDNA Libraries.

In one preferred embodiment, the coupling of cDNA molecules to the solid phase was done by printing on the pre-treated glass plates as described by Schena et al., Science 270: 467-470, 1995; Schena et al., Proc. Natl. Acad. Sci .U.S.A. 93: 10614-10619, 1996; DeRisi et al., Nature Genetics: 14: 457-460, 1996. In one embodiment, RNA had been isolated independently from individual sources such as cell lines, tissues, organs etc.

or from the same individual source but in a different time series after a specific treatment was given. RNA isolation and reverse transcription was carried out as described herein. RNA isolated from the test sample(s) had the red-fluorescent dye Cy5 present during the reverse transcription reaction while RNA isolated from the reference sample(s) was incubated with the green-fluorescent dye Cy3. (BioDirectory, Amersham Pharmacia Biotech, 2001) Before hybridization, the cDNA sample with Cy5 labeling was mixed with its corresponding cDNA sample with Cy3 labeling at equal amounts. Using two-color fluorescent probes in hybridization with DNA Microarrays was processed as described by Schena et al., Proc. Natl. Acad. Sci .U.S.A. 93: 10614-10619, 1996; and Shalon et al., Genome Res. 6: 639-645, 1996.

After hybridization with RNA/ cDNA/ DNA sample(s), the positive position(s) was identified. Therefore the candidate positive cDNA(s) was identified. Since each cDNA was derived and produced from its corresponding oligonucleotide through PCR aided cDNA library construction, the sequence(s) of the corresponding oligonucleotide(s) was/were known. Using those oligonucleotides as the upstream primers together with oligo-d(T) as the downstream oligonucleotide primer, the sequence(s) of genes of interest could be identified, amplified and eventually cloned in the subsequent PCR aid cloning procedures as described above.

Overall, the methods of preparing, fabricating, operating and applying both oligonucleotide arrays and cDNA arrays regarding an oligonucleotide and its corresponding derived cDNA synthesis, immobilization, sample labeling, hybridization, hybridization signal scan and detection include, but are not limited to US Pat. Nos. 5,242,974; 5,384,261; 5,405,783; 5,412,087; 5,424,186; 5,429,807; 5,436,327; 5,445,934; 5,472,672; 5,510,270; 5,527,681; 5,529,756; 5,545,531; 5,554,501; 5,556,752; 5,561,071; 5,599,895; 5,624,711; 5,639,603; 5,658,734; 5,700,637; 5,800,992; 5,807,522; 6,057,100; 6,197,506; 6,309,823; 6,337,188; 6,344,316; 6,352,828; 6,403,957; 6,406,921; US Pat. Application Nos. 20010005588; 20020006622; 20020015949; 20020051981; 20020064482; 20020072060. The present invention provides a unique design of oligonucleotide arrays, such as preferred 9mer, preferred 12mer, preferred 15mer, preferred 18mer and preferred 21mer oligonucleotide microarrays. The present invention provides their derived corresponding cDNA microarrays as well.

VI. PARALLEL GENE SEQUENCE DEDUCING, CLONING AND EXPRESSION AIDED BY ELISA-PCR-ELISA

The advantages of ELISA screening compared to protein microarrays are as follows: the number and size of ELISA screens are manually manageable. ELISA screening is a well-established standard laboratory method which has simple, straightforward procedures with high reliability and relatively quick turn-around. To perform ELISA screening does not require expensive equipment such as a laser scanner. Precoated ELISA plates could be produced massively by robot technology. The cost of operation is much lower than protein microarray technology whereas being miniaturized Protein microarrays.

Enzyme Linked Immunoassay (ELISA) is one of the standard amplification methods used to detect and display the presence of Antigen-antibody complex. Particularly, it has been used widely to measure antibody titers of clinical samples collected from serum, urine and culture supernatant at ng/ml to pg/ml sensitivity. Regarding simplicity, safety and specificity, ELISA is superior to the standard serological protocols and most of the immunoassays in both clinical and research applications. Generally, the typical epitope could consist of a hexa-peptide or more and be recognized by the corresponding monoclonal antibody such as, Anti-His(C-term) Antibody. (Invitrogen, Catalogue 2001) Even the antibody which binds the specific penta-peptide with nanomolar affinity has been demonstrated and introduced in Microarrays technology (Meo et al., Proc. Natl. Acad. Sci. U.S.A. 80: 4084, 1983) (Fodor et al., Science 251: 767-773, 1991). The ELISA protocols employed in this invention were as described by Perlmann et al., Enzyme-linked Immunosorbent Assay, Academic Press, Inc., 1994; and Harlow et al., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1988.

The present invention of series universal peptide libraries, such as the series universal peptide libraries of restriction enzyme recognition sequence of two-amino acids orientation, is capable of providing a unique miniaturized systematically immobilized capture epitope platform which is pre-coated on the surface of ELISA plates such as high binding polystyrene plates to identify, deduce and determine the sequence of the epitopes from positively displayed antigen-antibody complex. The sequences of the identified epitopes could be further deduced into a set of corresponding sequences of oligonucleotides. Using the anti-sense of those deduced oligonucleotides as the downstream primers and the inventive sense ORF oligonucleotide 5'-ATG orientation such

as 9mer sense ORF oligonucleotide 5'-ATG orientation library comprising 3,721 distinct oligonucleotides as the upstream primers, the candidate sequences of genes of interest with 5'-orientation could be detected and identified in PCR and/or using those deduced oligonucleotides as the upstream primers and oligo-d(T)₁₈ as the downstream primer, the candidate sequences of gene of interest with 3'-orientation could be detected and identified by PCR. Those sequences could be further displayed on Agarose gel and isolated for subsequent cloning and processing respectively. Sequencing analysis would determine all the candidate sequences of the 5'-portion and 3'-portion of the gene of interest respectively. Therefore, a set of the entire candidate DNA sequences from 5'-ATG to 3'-polyA could be determined. Using the variety of expression vectors carrying those DNA sequences, such as AdEasy-1® vector, pESP-1 vector and pDual vector, (Stratagene, Catalog 2001/2002) the corresponding expressed antigenic proteins could be obtained. Those antigenic proteins could be used as the capture antigens and coated and immobilized on a new set of ELISA plates. The specific antigen(s) would be determined following another ELISA test.

1. ELISA Primary Screening for the Peptide Sequence(s) of Epitope(s)

In one preferred embodiment, the penta-peptide library of EcoR I recognition sequence with N-terminal orientation which consists of 8,000 distinctive penta-peptides were chosen as the capture epitopes. In one preferred embodiment, the hexa-peptide library of EcoR I recognition sequence with N-terminal orientation which consists of 160,000 distinctive hexa-peptides were chosen as the capture epitopes. Each distinctive hexa-peptide was dissolved in 1 x Phosphate Buffered Saline (PBS) pH 7.4. The concentration of hexa-peptides could be in the range of 20 ug/ml to 200 ug/ml. A pilot experiment was done to determine the precise concentration of the capture epitopes of interest. 40 ul of each solution was added on a corresponding well of the plastic plate such as, Falcon® 3915 Pro-Bind® microtitre plate or NUNC® Maxisorp 96-well ELISA plate. The incubation time is from 60 to 120 minutes at 37°C or 4 hours at room temperature. The plate was covered by Parafilm® to prevent evaporation and possible contamination. The plate was incubated at 4°C for another 12 hours. Before use, each epitope solution in the well was absorbed by a vacuum pump. The wells were washed three times with 1 x PBS 0.05%Tween®-20. The plate was drained away by inverting the plate 180° on a 3MM® paper pad after each washing for 10 to 30 second. The plate was not allowed to dry completely. 50µl of blocking solution (1% BSA / 1 x PBS / 0.005% thimerasol) was added in each well to block non-specific binding on the sensitized plate. The blocking step was

incubated at 37°C for 120 minutes followed by subsequent incubation at 4°C for 12 hours. The plate was covered by Parafilm®. The blocking solution in the well was absorbed by a vacuum pump. The wells were washed three times by 1 x PBS 0.05% Tween®-20. The plate was drained away by inverting the plate 180° on 3MM® paper pad for 10 to 30 seconds after each washing and the plate was not allowed to dry completely. The patient serum was serially diluted with 1 x PBS in accordance with the algorithm of $n.\text{sup}2$ from 1:2 to 1:16,384. Using a known positive antigen to work out the optimal dilution in accordance with standard ELISA protocols as described by Perlmann et al., *Enzyme-linked Immunosorbent Assay*, Academic Press, Inc., 1994; and Harlow et al., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1988. The ideal diluted serum of 40ul was added in each well of a Falcon® 3915 Pro-Bind® microtitre plate and 100ul in each well of a NUNC® Maxisorp 96-well ELISA plate. Each plate had wells that contained 1 x PBS as negative controls and a known antigen to that serum as the positive control.

Wrapped with Parafilm®, the plates were incubated at 37 °C for 120 minutes followed by incubation at 4 °C for 12 hours. The incubation solutions from the plates were discarded the wells were washed using 1x PBS 0.05% Tween®-20 three times. The conjugate secondary antibody solution of 40 ul was added to each well of the Falcon® 3915 Pro-Bind® microtitre plate and 100ul to each well of the NUNC® Maxisorp plate. If the anti-sera are raised in rabbits, goat anti-rabbit alkaline phosphatase antibody will be the corresponding conjugate secondary antibody. The incubation time was usually 60 to 90 minutes at 37°C. The conjugate antibody solution was discarded from all the wells and the plate was washed three times with 1x PBS 0.05% Tween®-20 at room temperature. 40 ul or 100 ul of substrate solution was added to each well of the Falcon® 3915 Pro-Bind microtitre plate and the NUNC® Maxisorp plate respectively. If the conjugate enzyme were alkaline phosphatase, p-nitrophenyl phosphate disodium would be chosen as the substrate. It was dissolved in the substrate buffer (20 mg NaN_3 , 291 mg NaHCO_3 , 159 mg Na_2CO_3 , 10 mg MgCl_2 in 100 ml dH_2O). Usually, the incubation was at room temperature for 10 to 30 minutes. Sometimes, incubation at 37°C was required. 10 ul to 50 ul of the stop solution (0.1M EDTA pH 7.4) was added to each well at the end of the incubation to stop the enzyme reaction. The absorbance values were read on a spectrophotometer (microplate reader) at OD 405/450/490. The positive wells were determined. In this manner, the corresponding sequence(s) of the antigenic epitope(s) were determined. In another embodiment, biotinylated serum was diluted to the appropriate titer and was

added to each well of the ELISA plate which was precoated with the distinct capture epitope. After the plate was washed three times and the unbound antibodies were removed, Streptavidin-peroxidase was then added to each well as described in the art. The plate was washed three times again to remove all the unbound Streptavidin-peroxidase from each well. The substrate solution containing peroxide and TMB was added to each well. The Streptavidin-peroxidase of antigen-antibody complex catalyzed the enzyme reaction with the substrate to produce colour which was proportional to the amount of antigen in the test sample and could be measured under an ELISA reader.

2. PCR Screening for the Corresponding cDNA Sequence(s) of Epitope(s)

In one embodiment, if EFCMHW was determined as the antigenic epitope by ELISA, the set of 16 candidate corresponding 6 codon (18mer) oligonucleotides were deduced as follows:

- | | |
|---------------------------------|---------------------------------|
| (1) 5'-GAA TTC TGC ATG CAC TGG | (2) 5'-GAA TTT TGC ATG CAC TGG |
| (3) 5'-GAA TTC TGT ATG CAC TGG | (4) 5'-GAA TTC TGT ATG CAT TGG |
| (5) 5'-GAG TTC TGC ATG CAC TGG | (6) 5'-GAG TTT TGC ATG CAC TGG |
| (7) 5'-GAG TTC TGT ATG CAC TGG | (8) 5'-GAG TTC TGT ATG CAT TGG |
| (9) 5'-GAA TTC TGC ATG CAT TGG | (10) 5'-GAA TTT TGC ATG CAT TGG |
| (11) 5'-GAA TTC TGT ATG CAT TGG | (12) 5'-GAA TTC TGT ATG CAC TGG |
| (13) 5'-GAG TTC TGC ATG CAT TGG | (14) 5'-GAG TTT TGC ATG CAT TGG |
| (15) 5'-GAG TTC TGT ATG CAT TGG | (16) 5'-GAG TTC TGT ATG CAC TGG |

The above 16 distinctive six codons (18mer) oligonucleotide sequences could be consolidated into 5'-GAS TTS TGS ATG CAS TGG (S represents substitutions of the universal base analogues such as 5-nitroindole, 3-nitopyrrole, inosine and hypoxanthine etc.). The substitutions of the universal base are not appropriate at any 3'-end of any oligonucleotide sequences. In one preferred embodiment, 5-nitroindole was chosen to be the universal base in the consolidated six codon (18mer) oligonucleotide sequence.

In one preferred embodiment, using the above 16 distinctive six codons (18mers) oligonucleotide sequences as upstream PCR primers and oligo-d(T)₁₈ as the downstream PCR primer, detected and amplified the specific corresponding sequence(s) of interest by PCR. In one preferred embodiment, using the above consolidated six codon (18mers) oligonucleotide sequence as the upstream PCR primer and oligo-d(T)₁₈ as the downstream PCR primer, detected and amplified the specific corresponding sequence(s) of the 3'-portion (from 5'-GAS TTS TGS ATG CAS TGG to polyA) of interest of the gene(s) by PCR. In one embodiment, using the above consolidated six codons (18mers) oligonucleotide sequences as the downstream PCR primer and the oligonucleotides of the

entire 5'-ATG oriented oligonucleotide library of a suitable length, such as three codons (9mers) or four codons (12mers) or five codons(15mers) as the upstream PCR primers detected and amplified the specific corresponding sequence(s) of 5'-portion (from 5'-ATG to 5'-GAS TTS TGS ATG CAS TGG) of the gene(s) of interest by PCR. In one preferred embodiment, the incorporated indirect detectable label was biotinylated-dNTP. The biotinylated sequence could be detected and visualized by Streptavidin which conjugates to a fluorescein isothiocyanate or an enzymatically activatable molecule. In another preferred embodiment, the incorporated indirect detectable label was digoxigenin-dNTP. The sequence incorporated with digoxigenin-dNTP could be detected and visualized by an anti-digoxigenin antibody which conjugated to a fluorescent molecule or an enzymatically activatable molecule (PCR ELISA, Roche Diagnostics GmbH, Cat. No. 1636 120, 2001). In accordance with the know arts, RNA was isolated from the corresponding biological samples. The isolated RNA was reverse transcribed into cDNA which in turn served as the targeting template in PCR. In one embodiment, the initial denaturation was at 94 °C for 1 minute for one cycle. Since the multiple oligonucleotide primers have variable melting temperatures, the optimal annealing temperature varied from 37°C to 65°C. The elongation was at 72°C for 2 minutes. The number of cycles was from 25 to 30. The last elongation was at 72 °C for 5 minutes followed by immediately keeping the temperature at 4°C. In one embodiment, in order to work out the optimal annealing temperature of each specific PCR, touchdown PCR protocols were employed (Don et al., Nucleic Acids Res. 19: 4008, 1991). In one embodiment, the PCR solution was prepared to have the final concentration of 100 uM of dATP, dGTP, dTTP and dCTP respectively. The PCR solution contains the final concentration of 1.5 mM MgCl₂, 25mM KCl, 50-100 nM of each primer, 1-10 nM of template DNA, 2-10 units of Taq DNA polymerase and 20 mM Tris-HCl. All the positive clones from 3'-portions and 5'-portions of the gene(s) of interest were identified respectively using the PCR ELISA (DIG-Labeling) Kit. (PCR ELISA, Roche Diagnostics GmbH, Cat. No. 1636 120, 2001) The DNA sequence(s) of a set of candidates of full length clones (from 5'-ATG to polyA) of the gene(s) of interest were deduced by all the possible combinations of the positive 5'-portions and 3'-portions.

3. ELISA Final Screening to Determine the Peptide Sequence(s) of Epitope(s)

Using a variety of expression vectors carrying those DNA sequences such as AdEasy®-1 vector, pESP-1 vector and pDual vector (Stratagene, Catalog 2001/2002), the corresponding expression antigenic proteins could be obtained. Those antigenic proteins could be used as the capture antigens which would be coated and immobilized on a new

set of ELISA plates. The peptide sequence(s) of specific antigen(s) could be determined by the final ELISA test.

VII. PARALLEL SEQUENCE DEDUCING AND CLONING BY PROTEIN MICROARRAYS

Protein arrays including peptide and polypeptide arrays are solid-phase ligand binding platforms with immobilized and precisely positioned peptide or polypeptide or protein on a solid surface like silicon, glass slide, membrane, microtiter wells, mass spectrometer plates, microbeads and other particles. Protein arrays are capable of revealing functionalities as well as the interactions of protein to ligand, protein to protein, antigen to antibody and protein to enzyme. The present invention provides series universal peptide libraries of N-terminal orientation, series universal peptide libraries of C-terminal orientation and series universal peptide libraries of restriction enzyme recognition sequences of two-amino acids orientations based on the inventive algorithms of 20.sup.(n-1), 20.sup.n. and 20.sup.(n-2) respectively for the construction. To identify the responsible antigens and deduce sequences from peptide to DNA, the working principles of the above ELISA-PCR-ELISA are applicable.

VIII. SIGNATURE SEQUENCE AND SIGNATURE SIGNAL

The present invention provides series universal sense ORF oligonucleotide libraries of 5'-start codon orientation, series universal sense ORF oligonucleotide libraries of 3'-stop codon orientation and series universal sense ORF oligonucleotide libraries of restriction enzyme recognition sequence of two-codon orientations based on the inventive algorithms of 61.sup.(n-1) and 61.sup.(n-2) respectively. The present invention also provides series universal peptide libraries of N-terminal orientation, series universal peptide libraries of C-terminal orientation and series universal peptide libraries of restriction enzyme recognition sequence of two-amino-acids orientations based on the inventive algorithms of 20.sup.(n-1), 20.sup.n. and 20.sup.(n-2) respectively. Those are the resources and systems for the identification, selection and determination of signature sequences and signature signals for both nucleic acids and proteins. For example, each of the known sequences of an expressed gene of a genome could be used to hybridized with or aligned with the above 9mer, 12mer, 15mer, and 18mer inventive oligonucleotides in order to identify and determine the specified signature sequence(s) of its own at either the

5'-terminal or 3-terminal or between the two terminus or in combination therewith. Once the signature sequence for each expressed gene of the entire genome of any biological species such as human was identified, determined and registered, the complete sequence analysis may not be required. Codon consists of triplet nucleotides namely A, T, G, C, and U. Each of A, T, G, C and U possesses a characteristic maximum energy absorption spectrum at 259 nm, 267 nm, 252 nm, 271 nm and 258 nm respectively. Therefore, each signature sequence possesses a detectable unique energy emission pattern. The existing extrinsic label substances, such as a light emitting compound, quenching compounds and radioactive compounds could be coupled with the signature sequence by means known in the art. The coupling of extrinsic labels to the signature sequences of either nucleic acids or peptide could provide a more sophisticated signature signal for distinction and characterization. Those signature signals could be the chemical and/or physical signals in different forms such as energy emitting, absorbing, transferring and quenching. The signature sequence and the corresponding signature signal could be detected and measured either chemically, magnetically, electromagnetically, electronically, optically or combinatorially. In one preferred embodiment, a fluorescent compound such as Cy-3 was employed as the extrinsic label substance to a DNA signature sequence. The protocol of extrinsic labelling such as fluorescent compound labelling and corresponding signal excitement, detection and analysis was carried out as described by Gilmanshin et al., U.S. Pat. No. 6, 263, 286, 2001; Chan, U.S. Pat. No. 6,210,896, 2001; U.S. Pat. No.6, 355,420, 2002; and U.S. Pat. No.6, 403,311, 2002.

IX. TERMINAL SEQUENCE DATABASES

In accordance with the present inventive series universal oligonucleotide and peptide libraries of any given length, the corresponding terminal sequence databases are set forth below.

1. 5'-terminal Sense Sequence Series Databases

For example, (1) All possible ORF Sense Sequential Sequences of 5'-terminal Start Codon Orientation, such as 9mers, 12mers, 15mers, 18mers and so on. (2) All possible 5'-UTR Sense Sequential Sequences of 3'-terminal Start Codon Orientation, such as 9mers, 12mers, 15mers, 18mers and so on.

2. 3'-terminal Sense Sequence Series Databases

For example, (1) All possible ORF Sense Sequential Sequences of 3'-terminal Stop Codon Orientation, such as 9mers, 12mers, 15mers, 18mers and so on. (2) All possible 3'-UTR Sense Sequential Sequences of 5'-terminal Stop Codon Orientation, such as 9mers, 12mers, 15mers, 18mers and so on.

3. Other Databases

The Sense Sequence of Restriction Enzyme Recognition Sequence of Two-codon Orientations Series Databases, 5'-terminal Anti-sense Sequence Series Databases, 3'-terminal Anti-sense Sequence Series Databases, The Anti-sense Sequence of Restriction Enzyme Recognition Sequence of Two-codons Orientations Series Databases, N-terminal Amino Acid Sequence Series Databases, C-terminal Amino Acid Sequence Series Databases, The Amino Acid Sequence of Restriction Enzyme Recognition Sequence of Two Amino Acids Orientations Series Databases.

In order to form computer readable image media, the above series terminal sequence databases and their analogues with any given length could be written, rewritten, modified, formatted and recorded in current available computer software, such as Microsoft® Word, Excel and Access etc.. Those databases could also be represented in other computer readable formats, such as ASCII file and other data processor formats, such as Sybase, DB2 and Oracle etc.. In order to ensure any computer including ones not connected to the Internet has access to those databases, those databases could be stored in various forms and resources as long as the sequence image is computer readable. For example, those databases could be stored either magnetically or electrically or optically or combinations thereof. Except for those databases stored in a central computer, the storage media could be 3 1/2 floppy discs, compact discs, tapes, computer hard drives, ROM, RAM and CD-ROM etc.. The utilities of those databases are various. If a sequence of either nucleic acid or amino acid were determined be it known or unknown, a query could be made to those databases from a computer with Graphical User Interfaces (GUIs) such as, the Macintosh and Microsoft Windows such as Window® 95, 98, 2000 and Window® XP to search and identify the specific 5'- terminal, 3'-terminal, RERS-terminal, N-terminal and C-terminal sequences of a given length by the comparison and sequence matching. It is known in the art that a specific pair of 5'- terminal and 3'-terminal sequence that usually possesses 3 codons (9mers) and/or beyond it in the length of a particular gene could be used as upstream and downstream primers in PCR. With a specific optimized PCR protocol for a particular gene sequence, the full length of that particular ORF sequence from or before the start codon to the stop codon could be

identified and cloned experimentally. The full length of that particular ORF sequence from or before the stop codon to the start codon could also be identified and cloned experimentally. Those databases could also be used to identify and determine the signature sequence of a gene of interest. Therefore, those databases could be used as routine tools for PCR and cloning operations of research and clinical laboratories.

Recent progress in genetic research, particularly in the era of the Human Genome Program, there are enormous sequencing data mines of different sources, formats and file types dispersed worldwide. One skilled in the art could mine, interact, exchange and develop that information to advance science. One of the current major issues is the standardization of those databases. The present invention provides the series universal genetic algorithms, such as 61.sup.(n-1), 61.sip.n, 64.sup(n-1), 20.sup.(n-1) and 20.sup.n for ORF, 5'-UTR, 3'-UTR, N-terminal and C-terminal peptide sequences. Those algorithms are all codon based methodology. Therefore it could be useful in unifying, integrating, standardizing and rectifying the current existing databases into a common codon-based database(s).

X. KITS

1. RNA Quality Control Kit

The present invention provided a novel working kit for RNA quality control. It is known that the degradation of mRNA is often due to the residual RNases in the RNA extraction. 2-5 ug of test RNA sample was dissolved in DEPC-treated H₂O. The same amount of the control RNA sample was dissolved in 0.1% SDS solution. Both are incubated at 37°C for 30 minutes. 2-5 ug test RNA samples and the control are loaded on 1% Agarose gel containing Ethidium Bromide and electrophoresis. Photography is taken under UV light. In comparison with the ratio and intensity of the bands of 28s and 18s of rRNA in the gel, the question of whether the residual RNases exist in the testing sample may be analyzed and detected. The present invention provides a working kit of RNA Quality Control System for the immediate usage for both research and industry.

2. cDNA Quality Control Kits

mRNA reverse transcription is one of the crucial steps in many procedures, such as hybridization, PCR and cloning. The quality of cDNA syntheses often has the decisive impact on the subsequent processes. Mouse beta-actin gene (GenBank Accession: X03672) were chosen as an indicator for the cDNA Quality Control by PCR detection. The

targeting regions and PCR primers are as following:

1. 5' end region of 405 bp(135 codons) / 36-440

Forward primer (F1): 5'-TTT GCA GCT CCT TCG TTG CCG-3' 21nt

Backward primer(B1): 5'-GGT CAT CTT TTC ACG GTT GGC-3' 21nt

2. 3' end region of 477bp(159 codons) / 693-1,169

Forward primer (F1): 5'-GAG AGG GAA ATC GTG CGT GAC-3' 21nt

Backward primer(B1): 5'-ATC GTA CTC CTG CTT GCT GAT -3' 21nt

3. The gene of 1,134bp(378 codons) / 36-1,169

Forward primer (F1): 5'-TTT GCA GCT CCT TCG TTG CCG-3' 21nt

Backward primer(B1): 5'-ATC GTA CTC CTG CTT GCT GAT -3' 21nt

PCR Profile: Initial Denaturation: 94°C, 45 Sec. Denaturation: 94°C for 30 Sec.

Primer Annealing: 52° for 30 Sec. Primer Extension: 72°C for 90 Sec. for 25 Cycles.

PCR Protocol: the final volume of each PCR reaction is 20 ul. It contains 1 to 25 ng of cDNA, 1.5 mM MgCl₂, 50 mM KCl, 20 mM Tris-HCl (pH 7.4), 0.1 mM EDTA, 0.1 mM DTT, 150 uM each of dNTPs (dATP, dCTP, dGTP and dTTP), 0.05% Tween 20, 10 to 25 pM of each primer and 1 to 2 units of Taq DNA polymerase in a PCR reaction tube.

The sequences of β -actin or and glyceraldehyde-3-phosphate dehydrogenase (GAPD) was / were chosen as the indicators for the cDNA quality control. cDNA pool or library was diluted in accordance with 1: 1000 to 1: 10,000 to 100,000. The diluted cDNA library or pool was used as the template in the PCR.

The resulting PCR-amplicons are loaded evenly in parallel on the agarose gel and electrophoresized. Under the UV light, the evaluation of the integrity of the generated 5'-end and 3'-end cDNA fragments are performed by the comparison of the intensity of the bands. If the intensity of 5'-end fragment is equal to 3'-end fragment, it may imply that the targeting molecule (indicator) is intact. If the intensity of 5'-end fragment is higher than 3'-end fragment, it may imply that degradation happened on the 3'-end of the sequence. If the intensity of 5'-end fragment is lower than 3'-end fragment, it may imply that degradation happened on the 5'-end of the sequence.

3. cDNA Synthesis Kits

In reverse transcription procedures, Random Priming has been widely used to generate the first-strand of cDNA. However, the traditional random primers consist of random hexamers designed based on four single nucleotides. Therefore, nonsense codons are inevitably included in the hexamers. With the present invention, the cDNA

synthesis primers were designed based on all the possible combinations of two codons at each time among the 61 codons. The library of hexamer deoxyoligonucleotides possessed 3,721 distinct hexamer deoxyoligonucleotides in total. In accordance with the algorithm of $4^{3n} - (4^3 - 3)^n$ or $4^{3n} - (61)^n$, $4^6 - (4^3 - 3)^2 = 4,096 - 3,721 = 375$ when $n = 2$. The present invention has eliminated 375 redundant hexamers that consisted of 125 non-sense codons from hexamer deoxyoligonucleotide and oligonucleotide library designed by all possible combinations of the four-nucleotides (algorithm of $4^{\text{sup}.n}$). The anti-sense hexamers could be used to synthesize the first strand cDNA in the presence of mRNA as the template in reverse transcription. All the sequences of the 3,721 hexamer deoxyoligonucleotide and oligonucleotide primers were determined. All the sequences of the 3,721 anti-sense hexamer deoxyoligonucleotide and oligonucleotide primers were determined. The present invention provides a working kit for cDNA synthesis for both research and industry.

4. DNA Sequencing Primer Kits

The Kits include the inventive hexamer sense ORF oligonucleotide library which consists of 3,721 distinctive hexamer oligonucleotides, 9mer sense ORF oligonucleotide library which consists of 226,981 distinctive 9mer oligonucleotides and their anti-sense counterpart libraries respectively. The methods and protocols for processing DNA sequencing by using oligonucleotide library were performed as described by Blocker, U.S. Pat. No. 5, 114, 839, 1992.

5. PCR Primer Kits

The Kits include the inventive 9mer sense 5'-ATG orientation ORF oligonucleotide library which consists of 3,721 distinctive 9mer oligonucleotides and its derivative libraries respectively as the upstream primers. The kits include oligo-d(T)₁₂₋₁₈ as the downstream primers as well. The methods and protocols for processing PCR were performed as described by Innis et al., PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., 1990; Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, 1989.

XI. BIOMOLECULAR COMPUTATION

Adleman proposed a biomolecular computation model of using ligation of two DNA

fragments as a single operation at micromolar concentration (Adleman, Science 266: 1021-1024, 1994). In accordance with Adleman's model, the speed of a DNA computer may possibly reach at 10^{20} operations per second which is 10^8 times faster than the current supercomputers. This is mainly achieved by the massively parallel operations of DNA. The genetic information was stored in the codon of DNA molecules. If assuming that one average codon occupies one nanometer along the linear DNA sequence, then it would allow for a data density of about one codon per cubic nanometer. This implies the capacity of 16×10^4 G bits per square centimeter in DNA whereas the hard drive of an electronic computer has 1.12 G bits at the same size. The DNA computer could have at least more than 10^5 times more data storage density than the current supercomputer. Furthermore, the DNA computer could be potentially built at least 10^5 times smaller than existing silicon-based computers but with the same data storage capacity. Another important issue is the energy efficiency. In accordance with Adleman's DNA ligation model, one Joule of energy may be sufficient enough for 2×10^{19} DNA ligation operations whereas 10^9 operations at most could be executed by the current supercomputer (Adleman, Science 266: 1021-1024, 1994). It is likely that DNA computer may consume energy 10^{10} times less than the current supercomputer. In 1994, Leonard M. Adleman reported the first small-scale biomolecular computation. In 2002, Adleman et al. reported the largest-scale of DNA computing to date (Reif, Science 296: 478-479, 2002) (Braich et al. Science 296: 499-502, 2002). Adleman successfully resolved an instance of a computing problem by using the tools of PCR and DNA ligation. However the solution relied on the genetic algorithm of 4.sup.n when 20mer oligonucleotides were chosen for encoding the graph. Accordingly, Adleman's model has included huge non-sense codons. It seems unlikely that Adleman's solutions to those computational problems were based on the coding regions of the genes. The present invention provided the means for rectification of Adleman's solution to the directed Hamiltonian Path problem. The present invention provided alternative algorithms and models for exploring the possibilities of making use of nucleic acid molecules as a future computer programming language. All the algorithms and models provided by the present invention are applicable to the fields of biomolecular computation, computational Biology and Bioinformatics.

XII. Library Construction

Each of the distinct natural polydeoxyoligonucleotides or polyoligonucleotides

thereafter of a given length, which was measured and quantified by the number of codons, is linear polymers of molecules that covalently joined by deoxynucleotides or nucleotides respectively. Each of the distinct natural deoxyoligonucleotides is covalently joined together with each other by phosphodiester bonds between 3'-hydroxyl group of the preceding nucleotide and 5'-phosphate group of the immediately adjacent nucleotide in 5' towards 3' orientation. The same is true for the oligonucleotides.

Each of the distinct natural polydeoxyoligonucleotides or polyoligonucleotides thereafter of a given length, which was measured and quantified by the number of codons, is translated into peptide sequences that consist of L-amino acids respectively. Each of the distinct natural L-amino acid of the translated peptides is covalently joined together with each other by peptide bonds between carboxylic acid groups of the preceding amino acid and amino groups of the immediately adjacent amino acid in N-terminal towards C-terminal orientation.

Each of the distinct translated peptides thereafter is used as a distinct antigen in the production of the primary specific monoclonal or multiclinal antibodies respectively.

Each of the distinct monoclonal or multiclinal antibodies produced by using each distinct translated peptide is used as a distinct antigen in the production of the secondary specific monoclonal and multiclinal antibodies respectively.

1. The Series Universal ORF Sense Oligonucleotide Libraries
with 5'-end Start Codon Orientation

At the 5' end of the most of the genes, 5'-ATG is the first codon which orients the entire ORF from 5' end to 3' end. The second codon will be occupied by any one of the 61 codons. Although each of the consecutive subsequent codon of its corresponding preceding codon in 5' towards 3' orientation will be occupied by any one of the 61 codons, the number of combinations of 61 codons and the variety of sequences will increase according to the genetic algorithm of $61^{sup. (n-1)}$ as the linear codon sequences elongate. For example, there are $61^{3-1} / 61^2 / 3,721$ various sequences consisting of codons having 5' towards 3' orientated manner within three-codon or 9 nucleotides (3.times.3 codons). There are $61^{334-1} / 61^{333}$ various sequences consisting of codons having 5' towards 3' orientated manner within 334-codon or 1,002-nucleotide sequences (3.times.334 codons). Logically, there are $61^{n-1} / 61^E$ various sequences consisting of codons having 5' towards 3' orientated manner within n-codon-length long sequences. Therefore, the nth codon / position of each of the 61^{n-1} oligonucleotides will be occupied by each one of the 61 codons in 5' toward 3' orientation within 3 n-codon-length long sequences. n is the number

of the length of deoxyoligonucleotide(s) or oligonucleotide(s) measured by codon. The n^{th} codon occupies 5'-end oriented nucleotide positions $(3n-2)$ to $(3n)$. The positions of each of the nucleotides of the n^{th} codon in the triplet formation is $(3n-2)$, $(3n-1)$ and $(3n)$ respectively. n is the serial number of the codon/position for both 5'-end and 3'-end orientated ORF single strand of nucleic acids. The n is the serial number of codon/position. It begins at both 5'-end and 3'-end orientated ORF of single strand of nucleic acids. The positions of each of the nucleotides of the n^{th} codon in 5'-oriented triplet formation is $(3n-2)$, $(3n-1)$ and $(3n)$ respectively. n is an integer. n is not equal zero. $n=1$, $n>1$ and n is from 1 to infinite. $E = n-1$. E : Exponent. The algorithms and chemical compositional formulas for the construction of the Series Universal ORF Sense Oligonucleotide Libraries of 5'-end Start Codon Orientation are as follows: $61.\text{sup.}(n-1)$ under the conditions: $n=1$, $n>1$, $n<\infty$. n is an integer. n is not equal zero.

2. The Series Universal ORF Sense Oligonucleotide Libraries with 3'-end Stop Codon Orientation

As discussed above, there are three major stop codons (5'-TAA, 5'-TGA, 5'-TAG). Only one stop codon at the 3'-end of ORF per a single gene is a unique molecular characteristic of the gene. In a given gene, the stop codon at the 3' end orients the entire ORF sequences of the gene and its transcript from 3' towards 5'. The second codon will be occupied by any one of the 61 codons. Although each of the consecutive subsequent codons of its corresponding precedent codon in the 3' towards 5' orientation will be occupied by any one of the 61 codons, the combinations of 61 codons and the variety of sequences increase in accordance with the genetic algorithm of $61.\text{sup.}(n-1)$ accompanied with the linear elongation of the codon sequence. Thus, there are $61^{3-1} / 61^2 / 3,721$ various sequences consisting of codons having 3' towards 5' orientated manner within 9 nucleotides sequence (3.times.3 codons). Therefore, the n^{th} codon / position of each of the 61^{n-1} oligonucleotides will be occupied by each one of the 61 codons in 3' toward 5' orientated manner within 3.times. n nucleotides sequences. n is the number of the length of deoxyoligonucleotide or oligonucleotide. The n^{th} codon occupies 3'-end oriented nucleotide positions $(3n)$ to $(3n-2)$. The positions of each of the nucleotide of the n^{th} codon in 5'-oriented triplet formation is $(3n)$, $(3n-1)$ and $(3n-2)$ respectively. The present invention has eliminated 16,550,235 redundant 12mer oligonucleotides, which consisted of 5,516,745 non-sense codons from the four-nucleotide based (algorithm of 4.^{sup.} n based) 12mer oligonucleotide library that has no orientation. The algorithms and chemical compositional formulas for the construction of the Series Universal Libraries of ORF

Sense Oligonucleotide of 3'end Orientation are as follows: $61.\sup{(n-1)}$ under the conditions: $n=1$, $n>1$, $n<\infty$, n is an integer, n is not equal zero. One stop codon is at 3'-end of each oligonucleotide.

3. The Series Universal ORF Sense Oligonucleotide Libraries of Restriction Enzyme Recognition Sequence of Two-codon Orientations

A method of preparing the series universal ORF oligonucleotide libraries oriented by the restriction enzyme recognition sequence of two-codon wherein the restriction enzyme is selected from the group of restriction enzymes, which exclude any and all stop codons within the recognition sequence comprising: Aat II, Acc65 I, Ad I, Afe I, Afl II, Age I, Apa I, Apal I, Ase I, Avr II, BamHI, BfrBI, Bgl II, Bme1580 I, BmgB I, BseY I, Btr I, BstWI, BspDI, BspEI, BsrBI, BsrGI, BssH II, BssS I, BstBI, BstZ17 I, Cla I, Dra I, Eag I, EcoRI, EcoRV, Fsp I, Hind III, Hpa I, Kas I, Kpn I, Mfe I, Mlu I, Msc I, Nae I, Nar I, Nco I, Nde I, NgoM IV, Nhe I, Nru I, Nsi I, PaeR7 I, Pci I, Pml I, PspOM I, Pst I, Pvu I, Pvu II, Sac I, Sac II, Sal I, Sca I, Sfo I, Sma I, SnaBI, Spe I, Sph I, Ssp I, Stu I, Tli I, Xba I, Xho I, Xma I, Acc I, BsaWI, BsiHKA I, Bsp1286 I, MspA1 I, Sty I.

The excluded restriction enzymes with two-codon recognition sequence are Bcl I, BspHI and Psi I. The composition of the restriction enzyme recognition sequence of two-codon is selected from the group of restriction enzymes, without limiting the generality of the foregoing, comprising:

5'-GACGTC
5'-GGTACC
5'-AACGTT
5'-AGCGCT
5'-CTTAAG
5'-ACCGGT
5'-GGGCCC
5'-GTGCAC
5'-ATTAAT
5'-CCTAGG
5'-GGATCC
5'-AGATCT
5'-CACGTC
5'-CGTACG
5'-ATCGAT
5'-TCCGGA

5'-GAGCGG
5'-TGTACA
5'-GCGCGC
5'-CTCGTG
5'-TTCGAA
5'-GTATAC
5'-TTTAAA
5'-CGGCCG
5'-GAATTC
5'-GATATC
5'-TGGCGA
5'-AAGCTT
5'-GGCGCC
5'-GGTACC
5'-CAATTG
5'-ACGCGT
5'-TGGCCA
5'-GCCGGC
5'-CCATGG
5'-CATATG
5'-GCTAGC
5'-TCGCGA
5'-ATGCAT
5'-CTCGAG
5'-ACATGT
5'-CACGTG
5'-CTGCAG
5'-CGATCG
5'-CAGCTG
5'-GAGCTC
5'-CCGCGG
5'-AGTACT
5'-TACGTA
5'-ACTAGT
5'-GCATGC
5'-AATATT
5'-AGGCCT

5'-TCTAGA
 5'-CCCGGG
 5'-GGGCAC
 5'-GTGCCC
 5'-GTCTAC
 5'-GTAGAC
 5'-GTCGAC
 5'-ACCGGA
 5'-TCCGGT
 5'-CCCAGC
 5'-GAGCAC
 5'-GTGGTC
 5'-GAGCCC
 5'-CAGCGG
 5'-CCGCTG
 5'-CCAAGG
 5'-CCTTGG

(a) The Series Universal ORF Sense Oligonucleotide Libraries of Restriction Enzyme Recognition Sequence of Two-codon of 5'-end Orientation

For example, 5'-GACGTC is the two-codon recognition sequence of Aat II. Placing 5'-GACGTC/ 5'-GACGUC in 5' towards 3' orientated manner within $6 / [3x (2-2) + 3x2]$ nucleotides. The third codon adjacent to the 3' end of 5'-GACGTC will be occupied by any one of the $61^{3-2} / 61^1$ codons in 5' towards 3' orientated manner within $9 / [3x (3-2) + 3x2]$ nucleotides. Although each of the consecutive subsequent codons of the preceding terminal sequence of two-codon in 5' towards 3' orientation will be occupied by any one of the 61 codons, the combinations of 61 codons and the variety of sequences increase according to genetic algorithm of $61 \cdot \sup. (n-2)$ accompanied with the linear elongation of the codon sequence. For example, there are $61^{4-2} / 61^2 / 3,721$ various sequences consisting of codons having 5' towards 3' orientated manner within 12 nucleotides (3.times.4 codons). The 334th codon / the 334th position will be occupied by any one of the $61^{334-2} / 61^{332}$ codons in 5' towards 3' orientated manner within $1,002 / [3x (334-2) + 3x2]$ nucleotides sequence. Therefore, the nth codon / position of each of the $61^{n-2} / 61^{n-2}$ oligonucleotides will be occupied by each one of the 61 codons in 5' towards 3' orientated manner within $[3x (n-2) + 3x2]$ nucleotides sequences. n is the number of the length of deoxyoligonucleotide(s) or oligonucleotide(s) measured by codon. The n is the serial

number of codon /position for both 5' and 3' orientated ORF single strand of nucleic acids. The positions of each of the nucleotide of the n^{th} codon in the 5'-oriented triplet formation is $(3n-2)$, $(3n-1)$ and $(3n)$ respectively. The n is from 2 to infinity. The n is an integer. n is not equal zero. $n=2$. $n>2$. $\text{Erers}=n-2$. Erers: Exponent of restriction enzyme recognition sequence. The algorithms and chemical compositional formulas for the construction of the Series Universal Libraries of ORF Sense Oligonucleotide of Restriction Enzyme Recognition Sequence of Two-codon Orientation of 5'-end are as follows: $61^{\text{sup.}(n-2)} \cdot \text{times.}(\text{REN})$ under the conditions: $n=2$, $n>2$, $n<\infty$. n is an integer. $61^{\text{sup.}(n-2)} \cdot \text{times.}(\text{REN})$. $61^{\text{sup.}(n-2)} \cdot \text{times.}1$, when $\text{REN} = 1$. REN: Restriction Enzyme Number. If only one of the Restriction Enzyme Recognition Sequence of Two-codon exists in each oligonucleotide, REN is one. If two of the Restriction Enzyme Recognition Sequences of Two-codon exist in each oligonucleotide, REN are two.

(b) The Series Universal ORF Sense Oligonucleotide Libraries of Restriction Enzyme Recognition Sequence of Two-codon of 3'-end Orientation

For example, placing 5'-GACGTC/ 5'-GACGUC as the first two consecutive codons orients the entire ORF sequences of a gene and its transcript from 3' end towards 5' end within $6 / [3 \times (2-2) + 3 \times 2]$ nucleotides sequence. The third codon adjacent to the 5' end of 5'-GACGTC will be occupied by any one of the $61^{3-2} / 61^1$ codons in 3' towards 5' orientated manner within $9 / [3 \times (3-2) + 3 \times 2]$ nucleotides. Although each of the consecutive subsequent codon of the precedent terminal sequence of two-codon in 3' towards 5' orientation will be occupied by any one of the 61 codons, the combinations of 61 codons and the variety of sequences increase according to genetic algorithm of $61^{\text{sup.}(n-2)}$ accompanied with the linear elongation of the codon sequence. For example, there are $61^{4-2} / 61^2 / 3,721$ various sequences consisting of codons having 3' towards 5' orientated manner within 12 nucleotides ($3 \cdot \text{times.}4$ codons). The 334th codon / the 334th position will be occupied by any one of the $61^{334-2} / 61^{332}$ codons in 3' towards 5' orientated manner within $1,002 / [3 \times (334-2) + 3 \times 2]$ nucleotides sequence. Therefore, the n^{th} codon / position of each of the $61^{n-2} / 61^{\text{Erers}}$ oligonucleotides will be occupied by each one of the 61 codons in 3' towards 5' orientated manner within $[3 \times (n-2) + 3 \times 2]$ nucleotide sequences. n is the number of the length of deoxyoligonucleotide(s) or oligonucleotide(s) measured by codon. The n is the serial number of codon /position for both 3' and 5' orientated ORF single strand of nucleic acids. The positions of each of the nucleotide of the n^{th} codon in the 5'-oriented triplet formation is $(3n)$, $(3n-1)$ and $(3n-2)$ respectively. The n is from 2 to infinity. The n is an integer. n is not equal zero. $n=2$. $n>2$. $\text{Erers}=n-2$. Erers: Exponent of restriction

enzyme recognition sequence. The algorithms and chemical compositional formulas for construction of the Series Universal Libraries of ORF Sense Oligonucleotide of the Restriction Enzyme Recognition Sequence of Two-codon Orientation of 3'-end are as follows: $61 \cdot \sup{(n-2)} \cdot \text{times} \cdot (\text{REN})$ under the conditions: $n=2$, $n>2$, $n<\infty$. n is an integer. $61 \cdot \sup{(n-2)} \cdot \text{times} \cdot (\text{REN})$. $61 \cdot \sup{(n-2)} \cdot \text{times} \cdot 1$, when $\text{REN} = 1$. REN: Restriction Enzyme Number. If only one of the Restriction Enzyme Recognition Sequence of Two-codon exists in each oligonucleotide, REN is one. If two of the Restriction Enzyme Recognition Sequences of Two-codon exist in each oligonucleotide, REN are two.

4. The Series Universal ORF Sense Oligonucleotide Libraries
Between 5'- and 3'- Orientations

(a) The Series Universal ORF Sense Oligonucleotide Libraries
Between 5' and 3' of 5'-end Orientation

The first codon / the first position of ORF is occupied by each one of the 61 codons in 5' towards 3' orientated manner within 3 nucleotides sequence (3×1). There are $61(61^1)$ distinct 3 nucleotide long oligonucleotides. The second codon / position adjacent to the 3' of the first codon / the first position of each of the 61 oligonucleotides will be occupied by each one of the 61 codons in 5' towards 3' orientated manner within 6 nucleotides sequence (3×2). There are $3,721(61^2)$ distinct 6 nucleotide long oligonucleotides. Subsequently, the third codon / the third position of each one of the 3,721 oligonucleotides will be occupied by each one of the 61 codons in 5' towards 3' orientated manner within 9 nucleotides sequence (3×3). There are $226,981(61^3)$ distinct 9 nucleotide long oligonucleotides. The 334th codon / the 334th position of each of the 61^{334} oligonucleotides will be occupied by each one of the 61 codons in 5' towards 3' orientated manner within 1,002 nucleotides (3×334). There are 61^{334} distinct 1,002 nucleotide long oligonucleotides. Therefore, the n th codon / position of each of the 61^n oligonucleotides will be occupied by each one of the 61 codons in 5' toward 3' orientation within $3 \cdot \text{times} \cdot n$ nucleotides sequences. The n^{th} codon occupies 5' oriented nucleotide positions $(3n-2)$ to $(3n)$. The positions of each of the nucleotides of the n^{th} codon in the 5'-oriented triplet formation is $(3n-2)$, $(3n-1)$ and $(3n)$ respectively. n is the number of the length of deoxyoligonucleotide(s) or oligonucleotide(s) measured by codon. The n is the serial number of codon / position for both 5' and 3' orientated ORF single strand of nucleic acids. n is an integer. n is not equal zero. $n=1$. $n>1$. n is from 1 to infinite. $E = n-1$. E: Exponent. The algorithms and chemical compositional formulas for the construction of the Series Universal Libraries of ORF Sense Oligonucleotide Between 5' and 3' of 5'-end Orientation

are as following: $61.\sup.n$ under the conditions: $n=1$, $n>1$, $n<\infty$. n is an integer. n is not equal zero. One of the 61 codon is at 5'-end of each oligonucleotide.

(b) The Universal ORF Oligonucleotide Libraries

Between 5' and 3' of 3'-end Orientation

The first codon / the first position of ORF is occupied by each one of the 61 codons in 3' towards 5' orientated manner within 3 nucleotides sequence (3×1). There are $61(61^1)$ distinct 3 nucleotide long oligonucleotides. The second codon / position adjacent to the 5' of the first codon / the first position of each of the 61 oligonucleotides will be occupied by each one of the 61 codons in 3' towards 5' orientated manner within 6 nucleotides (3×2). There are $3,721(61^2)$ distinct 6 nucleotide long oligonucleotides. Subsequently, the third codon / the third position of each one of the 3,721 oligonucleotides will be occupied by each one of the 61 codons in 3' towards 5' orientated manner within 9 nucleotides (3×3). There are $226,981(61^3)$ distinct 9 nucleotide long oligonucleotides. The 334th codon / the 334th position of each of the 61^{334} oligonucleotides will be occupied by each one of the 61 codons in 3' towards 5' orientated manner within 1,002 nucleotides sequence (3×334). There are totally 61^{334} distinct 1,002 nucleotide long oligonucleotides. Therefore, the n th codon / position of each of the 61^n oligonucleotides will be occupied by each one of the 61 codons in 3' towards 5' orientation within $3 \times n$ nucleotides. n is the number of the length of deoxyoligonucleotide or oligonucleotide. n is the serial number of codon / position for both 3' and 5' orientated single strand of nucleic acids. The positions of each of the nucleotide of the n^{th} codon in the 5'-oriented triplet formation is $(3n)$, $(3n-1)$ and $(3n-2)$ respectively. n is from 1 to infinity. n is an integer. n is not equal zero. $n=1$. $n>1$. The algorithms and chemical composition formulas for the construction of the Series Universal Libraries of ORF Sense Oligonucleotide between 5' and 3' of 3'-end Orientation are as following: $61.\sup.n$ under the conditions: $n=1$, $n>1$, $n<\infty$. n is an integer. One of the 61 codon is at 3'-end of each oligonucleotide.

5. The Universal ORF Sense Hexamer Oligonucleotide Library

Two codons were selected from the group consisting of the 61 codons at each time. By adding all the possible combinations of the two codons from the 61 codons without any overlap and repetition, The Universal ORF Sense Hexamer Oligonucleotide Library is synthesized. It comprises 3,721 distinct deoxyoligonucleotides or 3,721 distinct oligonucleotides. Each of the deoxyoligonucleotides or oligonucleotides is two-codons (3×2 nucleotides) long with 5' orientation. Any and all of the stop codons is excluded. The algorithms and chemical compositional formulas for the construction of the Universal

ORF Sense Hexamer Oligonucleotide Library are as follows: $61.\text{sup}.n$ under the conditions: $n = 2$. $61.\text{sup}.2 = 3,721$ One of the 61 codons is at 5'-end of each of the distinct 3,721 hexamer oligonucleotides.

6. The Series Universal 5'-UTR Sense Oligonucleotide Libraries of 3'-end Start Codon Orientation

A start codon, such as 5'-ATG / 5'-AUG was chosen as the 3'-end terminal codon for each deoxyoligonucleotide or oligonucleotide. By adding sequentially to the 5' of each start codon a codon selected from a group consisting of the 64 codons, 64 distinct deoxyoligonucleotides or oligonucleotides with two-codon length long or six-nucleotide length long ($3.\text{times}.2$) in 3' toward 5' orientation were synthesized. By continually adding sequentially to the 5' of each subsequent n codon position a codon selected from the 64 codons until reaching the appropriate position in accordance with the variety of different experimental purposes, the designed deoxyoligonucleotide or oligonucleotide could eventually elongate to the 5'-end of each 3'-end orientated deoxyoligonucleotide or oligonucleotide of n -codon-length long which is equal to $3.\text{times}.n$ -nucleotide-length.

By adding all possible combinations of the 64 codons sequentially and consecutively without any overlap, the Series Universal 5'-UTR Sense Oligonucleotide Libraries of 3'-end Start Codon Orientation comprising $64.\text{sup}.(n-1)$ distinct deoxyoligonucleotides or oligonucleotides could be constructed. Each of the deoxyoligonucleotides or oligonucleotides has n codon long sequence wherein $n=1$, $n>1$, and n is from 1 to infinite. n is an integer. n is not equal zero. When $n=1$, the position of codon is $(1-n)+1$. When $n>1$ and $n<\text{infinity}$, the position of codon is $(1-n)$. n represents codon. n is the number of the length of deoxyoligonucleotide or oligonucleotide. The negative sign in front of n indicates that codon position is in the 5'-UTR. The n^{th} codon occupies 3'-end oriented nucleotide positions $3(1-n)+1$ to $3(1-n)+3$ when $n=1$. The n^{th} codon occupies 3'-end oriented nucleotide positions $3(1-n)$ to $3(1-n)+2$ when $n>1$. The positions of each of the nucleotide of the n^{th} codon in the 5' oriented triplet formation is $3(1-n)+1$, $3(1-n)+2$ and $3(1-n)+3$ respectively when $n=1$. The positions of each of the nucleotide of the n^{th} codon in the 5' oriented triplet formation is $3(1-n)$, $3(1-n)+1$ and $3(1-n)+2$ respectively when $n=1$. The algorithms and chemical composition formulas for the construction of the Series Universal 5'-UTR Oligonucleotide Libraries of 3'-end Start Codon Orientation are as follows: $64.\text{sup}.(n-1)$ under the conditions: $n=1$, $n>1$, $n<\text{infinity}$. n is an integer. n is not equal zero. One start codon is at 3'-end of each oligonucleotide.

7. The Series Universal 3'-UTR Oligonucleotide Libraries of 5'-end Stop Codon Orientation

A stop codon, such as 5'-TGA / 5'-UGA was chosen as the 5'-end terminal codon for each deoxyoligonucleotide or oligonucleotide. By adding sequentially to the 3' of each stop codon a codon selected from a group consisting of the 64 codons, 64 distinct deoxyoligonucleotides or oligonucleotides of two-codon long or six-nucleotide long (3×2) in 3' toward 5' orientation were synthesized. By continually adding sequentially to the 3' of each subsequent n codon position a codon selected from the 64 codons until reaching the appropriate position in accordance with the variety of different experimental purposes, the designed deoxyoligonucleotide or oligonucleotide could eventually elongate to the 3'-end of each 5'-end orientated deoxyoligonucleotide or oligonucleotide of n -codon-length long which is equal to $3 \times n$ -nucleotide-length.

By adding all possible combinations of the 64 codons sequentially and consecutively without any overlap, the Series Universal 3'-UTR Sense Oligonucleotide Libraries of 5'-end Stop Codon Orientation comprising $64 \times (n-1)$ distinct deoxyoligonucleotides or oligonucleotides could be constructed. Each of the deoxyoligonucleotides or oligonucleotides has n codon long sequence wherein $n=1$, $n>1$, and n is from 1 to infinite. n is an integer. n is not equal zero. n is the number of the length of deoxyoligonucleotide(s) or oligonucleotide(s) measured by codon. The n^{th} codon occupies 5'-end oriented nucleotide positions $(3n-2)$ to $(3n)$. The positions of each of the nucleotide of the n^{th} codon in the 5' oriented triplet formation is $(3n-2)$, $(3n-1)$ and $(3n)$ respectively. The algorithms and chemical compositional formulas for the construction of the Series Universal 3'-UTR Oligonucleotide Libraries of 5'-end Stop Codon Orientation are as following: $64 \times (n-1)$ under the conditions: $n=1$, $n>1$, $n< \text{infinity}$. n is an integer. n is not equal zero. One stop codon is at 5'-end of each oligonucleotide.

8. The Series Universal Peptide Libraries of N-terminal Orientation

Except for some less frequent exceptions, peptides, polypeptides and proteins comprise various combinations of the 20 essential amino acids. Methionine or Formylmethionine is frequently the first amino acid at the first position of the N-terminal of the majority of the peptide chains and protein molecules. Hence, the majority of the peptide chains and protein molecules have a unique and common start position / the first position at the beginning of their N-terminals. The second amino acid / the second position adjacent to Methionine or Formylmethionine is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within two-peptide

sequence. There are 20 or $(20)^{2-1}$ or (20^1) distinct peptides of the two-peptide. Subsequently, the third amino acid / the third position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within three-peptide sequence. There are 400 or $(20)^{3-1}$ or (20^2) distinct peptides of the three-peptide. The 334th amino acid/ the 334th position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within 334-peptide sequence. There are 20^{333} or $(20)^{334-1}$ distinct peptides of the 334-peptides. Therefore, the nth amino acid / the nth position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within n amino acid long peptide sequence. There are 20^{n-1} or 20^E distinct peptides of the n-peptide. n is the serial number for N-terminal orientated amino acids sequences of peptides, polypeptides and proteins and begins at N-terminal orientated amino acids sequences of peptides, polypeptides and proteins. The n is from 1 to infinity. n is the number of the length of peptide, polypeptide and protein measured by amino acid encoded by codon. The n is an integer. $n \neq 1$. $n > 1$. n is not equal zero. $E = n - 1$. E: Exponent. Therefore there are $20^{n-1} / 20^E$ distinct n-peptide sequences. Since the number of the essential amino acids (EAA) could be variable during evolution, EAA is a variable. There are $(EAA)^{n-1} / (EAA)^E$ distinct n-peptide sequences at any evolutionary stage. However, the inventive methods are applicable at any evolutionary stage. The algorithms and chemical compositional formulas for the construction of the Series Universal peptide Libraries of the N-terminal Orientation are as following: $20.\sup.(n-1)$ under the conditions: $n = 1$, $n > 1$, $n < \text{infinity}$. n is an integer. n is not equal zero. One Methionine or Formylmethionine is at the first position of N-terminal of each peptide or polypeptide or protein.

9. The Series Universal Peptide Libraries of C-terminal Orientation

At 3' end of the most of genes, with some exceptions, 5'-TAA, 5'-TGA, 5'-TAG is /are the stop codons where peptide chains are released during protein synthesis. Statistically, any one of the 20 essential amino acids could be the first amino acid / the first position at C-terminal of the majority of the peptide chains and protein molecules (the gene products) of the majority of biological species. Those peptide chains and protein molecules are involved in all life processes. Therefore, majority of the peptide chains and protein molecules have such unique and common start positions at the beginning of their C-terminals. The first amino acid at the first position of the C-terminal (the last position from the N-terminal) of the peptide chain is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation within one-peptide sequence. There are

20 (20^1) distinct peptides of one-amino-acid long. Subsequently, the second amino acid / the second position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within a two-peptide sequence. There are 400 (20^2) distinct peptides of two-amino-acids long. Subsequently, the third amino acid / the third position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within a three-peptide sequence. There are 8,000 (20^3) distinct peptides of three-amino-acids long. The 334th amino acid / the 334th position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within 334-peptide sequence. There are totally 20^{334} distinct peptides having a length of 334 amino acids. Therefore, the n^{th} amino acid / the n^{th} position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation within an n-amino-acid long sequence. There are totally $20^n / 20^E$ distinct peptides of n-amino-acids long. The n is the serial number for N-terminal orientated amino acids sequences of peptides, polypeptides and proteins. The n is the serial number begins at C-terminal orientated amino acids sequences of peptides, polypeptides and proteins. n is the number of the length of peptide, polypeptide and protein measured by amino acid encoded by codon. n is from 1 to infinity. n is an integer. $n = 1$. $n > 1$. $E = n$. E: Exponent. There are 20^n distinct n-peptide sequences at current evolutionary stage. Since the number of the essential amino acids (EAA) could be variable during the evolution, EAA is a variable. There are $(EAA)^n / (EAA)^E$ distinct n-peptide sequences of C-terminals orientated peptide chains at any evolutionary stage while the inventive methods are always applicable. The methods are also flexible. For example, although the methods was established by the model of the peptides and proteins sequences with Methionine or Formylmethionine as the first amino acid at the first position orientated from the N-terminal towards C-terminal of peptide chains and protein molecules, Methionine can be substituted by Valine or any other amino acids encoded by any other start codons in the methods. Therefore any other relevant peptide sequences could still be predicted, deduced, synthesized and determined by the same methods. By the same principles, the methods can even be applied to the exceptions i.e. mammalian mitochondria and some bacteria. Only the substitutions or replacement of the amino acids (i.e. Valine substitutes Methionine, Tryptophan replaces one of the termination positions at the C-terminal of peptide chains etc.) are needed in order to apply those methods. In the case of ORF shifting, the methods will work too. Another unique feature of the methods is its universal application to the unusual situations that may have more or less than 20 essential amino acids in the peptides. The algorithms

and chemical compositional formulas for the construction of the Series Universal Peptide Libraries of the C-terminal Orientation are as following: $20.\text{sup}.n$ under the conditions: $n = 1$, $n > 1$, $n < \text{infinity}$. n is an integer. n is not equal zero. One of the 20 essential amino acids is at the first position of C-terminal of each peptide or polypeptide or protein.

10. The Series Universal Peptide Libraries

Between N-terminal and C-terminal Orientations

(a) The Series Universal Peptide Libraries

Between N-terminal and C-terminal of N-terminal Orientation

The first position at N-terminal is occupied by any one of the 20 essential amino acids. There are 20 (20^1) distinct peptides of one-amino-acid long. The second amino acid / the second position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within a two-amino-acids long peptide sequence. There are 400 (20^2) distinct peptides that are two-amino-acids long. Subsequently, the third amino acid / the third position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within three-amino-acids long peptide sequence. There are 8,000 (20^3) distinct peptides that are two-amino-acids long. The 334th amino acid / the 334th position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within 334-peptide sequence. There are totally 20^{334} distinct peptides of 334-amino-acids long. Therefore, the n^{th} amino acid / the n^{th} position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientation within a peptide sequence of n amino acids long. There are $20^n / 20^E$ distinct peptides of n -amino-acids long. The n is the serial number for N-terminal orientated amino acids sequences of peptides, polypeptides and proteins. n is the number of the length of peptide or polypeptide measured by amino acid encoded by codon. The n is from 1 to infinity. The n is an integer. $n = 1$. $n > 1$. $E = n$. E : Exponent. The algorithms and chemical compositional formulas for the construction of the Series Universal peptide Libraries Between N-terminal and C-terminal of the N-terminal Orientation are as following: $20.\text{sup}.n$ under the conditions: $n = 1$, $n > 1$, $n < \text{infinity}$. n is an integer. n is not equal zero. One of the 20 essential amino acids is at the first position of amino acids of N-terminal of each peptide or polypeptide or protein.

(b) The Series Universal Peptide Libraries

Between N-terminal and C-terminal of C-terminal Orientation

The first amino acid at the first position of C-terminal (the last position from N-terminal) of the peptide is occupied by any one of the 20 essential amino acids in C-

terminal towards N-terminal orientation within a one-amino-acid long peptide sequence. There are 20 (20^1) distinct peptides having a length of one amino acid. Subsequently, the second amino acid / the second position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation in two-amino-acids long peptide sequence. There are 400 (20^2) distinct two-amino-acids long peptides. Subsequently, the third amino acid / the third position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation in a three-amino-acids peptide sequence. There are 8,000 (20^3) distinct peptides of three-amino-acids long. The 334th amino acid / the 334th position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within a 334-amino-acids long peptide sequence. There are 20^{334} distinct peptides having a length of 334 amino acids. Therefore, the n^{th} amino acid / the n^{th} position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within an n-amino-acids long peptide sequence. There are $20^n / 20^E$ distinct peptides of n-peptide. The n is the serial number for N-terminal orientated amino acids sequences of peptides, polypeptides and proteins. n is the number of the length of peptide or polypeptide measured by amino acid encoded by codon. n is from 1 to infinity. The n is an integer. $n = 1$. $n > 1$. $E = n$. E: Exponent. The algorithms and chemical compositional formulas for the construction of the Series Universal peptide Libraries Between N-terminal and C-terminal of the C-terminal Orientation are as following: $20.\text{sup}.n$ under the conditions: $n = 1$, $n > 1$, $n < \text{infinity}$. n is an integer. n is not equal zero. One of the 20 essential amino acids is at the first position of C-terminal of each peptide or polypeptide or protein.

11. The Series Universal Peptide Libraries of Restriction

Enzyme Recognition Sequence Two-Amino-Acids Orientations

The restriction enzyme is selected from a group of restriction enzymes, which have two-codon recognition sequences that excluded any and all stop codons within the two codons. Examples of suitable restriction enzymes include but are by no means limited to Aat II, Acc65 I, Acl I, Afe I, Afl II, Age I, Apa I, ApaL I, Ase I, Avr II, BamHI, BfrBI, Bgl II, Bme1580 I, BmgB I, BseY I, Btr I, BsiW I, BspD I, BspE I, BsrB I, BsrG I, BssH II, BssS I, Bst B I, BstZ17 I, Cla I, Dra I, Eag I, EcoR I, EcoR V, Fsp I, Hind III, Hpa I, Kas I, Kpn I, Mfe I, Mlu I, Msc I, Nae I, Nar I, Nco I, Nde I, NgoM IV, Nhe I, Nru I, Nsi I, PaeR7 I, Pci I, Pml I, PspOM I, Pst I, Pvu I Pvu II, Sac I, Sac II, Sal I, Sca I, Sfo I, Sma I, SnaB I, Spe I, Sph I, Ssp I, Stu I, Tli I, Xba I, Xho I, Xma I, Acc I, BsaW I, BsiHKA I, Bsp1286 I, MspA1 I, and Sty I.

The corresponding restriction enzyme recognition sequences of two-codon are shown in TABLE 3. The excluded restriction enzymes with two-codon recognition sequence are Bcl I, BspH I and Pst I.

The panel of peptides comprising two amino acids deduced from the above restriction enzyme recognition sequences include but are by no means limited to:

NH₂-DV
NH₂-GT
NH₂-NV
NH₂-SA
NH₂-LK
NH₂-TG
NH₂-GP
NH₂-VH
NH₂-IN
NH₂-PR
NH₂-GS
NH₂-RS
NH₂-HV
NH₂-RT
NH₂-ID
NH₂-SG
NH₂-ER
NH₂-CT
NH₂-AR
NH₂-LV
NH₂-FE
NH₂-VY
NH₂-FK
NH₂-RP
NH₂-EF
NH₂-DI
NH₂-CA
NH₂-KL
NH₂-VN
NH₂-QL
NH₂-TR

NH₂-WP
 NH₂-AG
 NH₂-PW
 NH₂-HM
 NH₂-AS
 NH₂-SR
 NH₂-MH
 NH₂-LE
 NH₂-TC
 NH₂-LQ
 NH₂-EL
 NH₂-VD
 NH₂-ST
 NH₂-PG
 NH₂-YV
 NH₂-TS
 NH₂-AC
 NH₂-NI
 NH₂-GH
 NH₂-VP
 NH₂-EH
 NH₂-VL
 NH₂-EP
 NH₂-QR
 NH₂-PL
 NH₂-GA

(a) The Series Universal Peptide Libraries of Restriction Enzyme Recognition Sequence of Two-Amino- Acids of N-terminal Orientation

For example, 5'-GACGTC is the recognition sequence of two-codon of restriction enzyme Aat II. NH₂-DV is encoded by 5'-GACGTC. In some embodiments, a Two-Amino-Acids peptide from a restriction enzyme recognition sequence is placed at N-terminal of a designed peptide. For example, NH₂-DV is placed as the first two consecutive amino acids / the consecutive first and second positions of the designed peptide which orients the entire peptide sequence from N-terminal towards C-terminal. The consecutive first and second amino acid /positions of peptide is solely occupied by the designed two-amino-acids of the restriction enzyme recognition sequence of two-codon, e.g. NH₂-DV in N-

terminal towards C-terminal orientation within a two-peptide length sequence. The third amino acid / position adjacent to C-terminal of $\text{NH}_2\text{-DV}$ / the first and second amino acid / the first and second positions is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientation within a three-amino-acids long peptide. There are 20 three-amino-acids long peptides. Subsequently, the fourth amino acid / the fourth position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within a four-amino-acids long peptide. There are 400 (20^{4-2}) 4-peptides. The 334th amino acids / the 334th position is occupied by any one of the 20 essential amino acids in N-terminal towards C-terminal orientated manner within 334-peptide-length sequence. There are totally 20^{334-2} or 20^{332} 334-peptides. Therefore, the nth amino acid / position of peptide is occupied by any one of the $20^{n-2} / 20^{\text{Erers}}$ amino acids in N-terminal towards C-terminal orientated manner within n-peptide-length sequences. The n is the serial number of amino acid /position for both N-terminal and C-terminal orientated amino acid sequence of peptides. n is the number of the length of peptide or polypeptide or protein measured by amino acid encode by codon. n is from 2 to infinity. n is an integer. $n=2$. $n>2$. $E = n-2$. Erers: Exponent of restriction enzyme recognition sequence. There are $20^{n-2} / 20^{\text{Erers}}$ distinct amino acid sequences for all of n-peptide-length amino acid sequences of peptides. Since the number of the 20 Essential Amino Acids (EAA) could be variable during the process of evolution on the Earth. It could be variable among different species or different organs, tissues, cells and sub cellular ultra structures. EAA is a variable. At current evolutionary stage, for most of species on the Earth, $\text{EAA}=20$. There are $(\text{EAA})^{n-2} / (\text{EAA})^{\text{Erers}}$ distinct amino acid sequences for all of n-peptide-length amino acid sequences regardless the changes of the EAA, the methods are always applicable during the process of evolution. The algorithms and chemical compositional formulas for the construction of the Series Universal Peptide Libraries of the Two-Amino-Acid encoded by the Restriction Enzyme Recognition Sequence of Two-codon of N-terminal Orientation are as follows: $20.\text{sup.}(n-2).\text{times.}(\text{REN})$ under the conditions: $n=2$, $n > 2$, $n < \text{infinity}$. n is an integer. $20.\text{sup.}(n-2).\text{times.}(\text{REN}) = 20.\text{sup.}(n-2).\text{times.}1$ when $\text{REN} = 1$. Two amino acids encoded by Restriction Enzyme Recognition Sequence of Two-codon is at N-terminal of each peptide. REN: Restriction Enzyme Number.

(b) The Series Universal Peptide Libraries of Restriction Enzyme Recognition Sequence of Two-Amino-Acids of C-terminal Orientation

Similarly, 5'-GACGTC is the recognition sequence of two-codon restriction enzyme Aat II. $\text{NH}_2\text{-DV}$ is encoded by 5'-GACGTC. A specific two-amino-acids from a two-codon

restriction enzyme recognition sequence is placed at C-terminal of the designed peptide. For example, NH₂-DV is placed as the first two consecutive amino acids / the consecutive first and second positions of the designed peptide which orients the entire peptide sequences from C-terminal towards N-terminal. The consecutive first and second amino acid / positions of peptide is occupied by the designed two-amino-acids derived from the restriction enzyme recognition sequence of two-codon, e.g. NH₂-DV in C-terminal towards N-terminal orientated manner within two-peptide length sequence. There is 1 two-peptide. The third amino acid / position adjacent to N-terminal of NH₂-DV / the first and second amino acid / the first and second positions is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation within a three-peptide length sequence. There are totally 20 or 20^{3-2} three-peptides. Subsequently, the fourth amino acid / the fourth position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientation within a four-peptide-length sequence. There are the 400 or 20^{4-2} four-peptides. The 334th amino acid / the 334th position is occupied by any one of the 20 essential amino acids in C-terminal towards N-terminal orientated manner within 334-peptide-length sequence. There are 20^{334-2} / 20^{332} 334-peptides. Therefore, the nth amino acid / position of peptide is occupied by any one of the 20^{n-2} / 20^{Ers} amino acids in C-terminal towards N-terminal orientated manner within n-peptide-length sequences. The n is the serial number of amino acids / position for both C-terminal and N-terminal orientated amino acid sequence of peptides. N is the number of the length of peptide or polypeptide measured by amino acid encoded by codon. n is from 2 to infinity. n is an integer. $n \geq 2$. $n > 2$. Ers = n-2. E: Exponent of restriction enzyme recognition sequence. The algorithms and chemical compositional formulas for the construction of the Series Universal Peptide Libraries of the Two-Amino-Acids encoded by the Restriction Enzyme Recognition Sequence of Two-codon of C-terminal Orientation are as following: $20^{\text{sup.}(n-2)} \cdot \text{times.}(\text{REN})$ under the conditions: $n=2$, $n > 2$, $n < \text{infinity}$. n is an integer. $20^{\text{sup.}(n-2)} \cdot \text{times.}(\text{REN}) = 20^{\text{sup.}(n-2)} \cdot \text{times.}1$ when $\text{REN} = 1$. Two amino acids encoded by Restriction Enzyme Recognition Sequence of Two-codon is at C-terminal of each peptide. REN: Restriction Enzyme Number.

12. The Series Universal Antibody Libraries Produced by the Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of N-terminal Methionine Orientation
 - (a) The Corresponding Series Universal Monoclonal Antibody Libraries
 - (b) The Corresponding Series Universal Multiclonal Antibody Libraries

- (c) The Secondary Corresponding Series Universal Monoclonal Antibody Libraries

The algorithms and chemical compositional formulas for construction of the Series Universal Antibody Libraries produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of N-terminal Methionine orientation are as following: $20^{\sup.(n-1)}$ under the conditions: $n = 5$, $n > 5$, $n < \text{infinity}$. n is an integer. One-Amino-Acid encoded by a Start Codon, such as Methionine encoded by 5'-ATG is at the first position of amino acids sequence of N-terminal of each Antigen or Epitopes.

13. The Series Universal Antibody Libraries Produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of C-terminal Orientation
 - (a) The Corresponding Series Universal Monoclonal Antibody Libraries
 - (b) The Corresponding Series Universal Multiclonal Antibody Libraries
 - (c) The Secondary Corresponding Series Universal Monoclonal Antibody Libraries

The algorithms and chemical compositional formulas for construction of the Series Universal Antibody Libraries produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of C-terminal orientation are as following: $20^{\sup.n}$ under the conditions: $n = 5$, $n > 5$, $n < \text{infinity}$. n is an integer.

Any Amino Acid of the 20 Essential Amino Acids encoded by a codon is at the first position of amino acids sequence of C-terminal of each Antigen or Epitopes.

14. The Series Universal Antibody Libraries Produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of N-terminal Restriction Enzyme Recognition Sequence of Two-Amino-Acids Orientation
 - (a) The Corresponding Series Universal Monoclonal Antibody Libraries
 - (b) The Corresponding Series Universal Multiclonal Antibody Libraries
 - (c) The Secondary Corresponding Series Universal Monoclonal Antibody Libraries

The algorithms and chemical compositional formulas for construction of the Series Universal Antibody Libraries produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of N-terminal Restriction Enzyme Recognition Sequence of Two-Amino-Acids Orientation are as following: $20^{\sup.(n-2).times.(REN)}$ under the conditions: $n = 5$, $n > 5$, $n < \text{infinity}$. n is an integer. $20^{\sup.(n-2).times.(REN)} = 20^{\sup.(n-2).times.1}$ when $REN = 1$. REN: Restriction Enzyme Number. Two amino acids

encoded by Restriction Enzyme Recognition Sequence of Two-codon are at N-terminal of each Antigens or Epitopes.

15. The Series Universal Antibody Libraries Produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of C-terminal Restriction Enzyme Recognition Sequence of Two-Amino-Acids Orientation
 - (a) The Corresponding Series Universal Monoclonal Antibody Libraries
 - (b) The Corresponding Series Universal Multiclonal Antibody Libraries
 - (c) The Secondary Corresponding Series Universal Monoclonal Antibody Libraries

The algorithms and chemical compositional formulas for construction of the Series Universal Antibody Libraries produced by Antigens or Epitopes of the Corresponding Series Universal Peptide Libraries of C-terminal Restriction Enzyme Recognition Sequence of Two-Amino-Acids Orientation are as following: $20^{\sup.(n-2)} \times (\text{REN})$ under the conditions: $n = 5$, $n > 5$, $n < \infty$. n is an integer. $20^{\sup.(n-2)} \times (\text{REN}) = 20^{\sup.(n-2)} \times 1$ when $\text{REN} = 1$. REN: Restriction Enzyme Number. Two amino acids encoded by Restriction Enzyme Recognition Sequence of Two-codon are at C-terminal of each Antigens or Epitopes.

16. Analogues of Oligonucleotide and/or Peptide

Regardless of length, the oligonucleotide(s) mentioned above as a single oligonucleotide or a group of oligonucleotides or partial oligonucleotide library or entire oligonucleotide library or combination of oligonucleotides within a unique library or between different libraries or in the combinations of various libraries could be used for various purposes, such as signature signal / tag for the sequencing analysis, PCR primers and hybridization probes. The inventive oligonucleotide(s) could be either dissolved in the aqueous phase as primer(s) (PCR) or immobilized on the surface of solid phase as probes (Dot blotting and DNA microarrays). As the PCR primer, the length of oligonucleotide usually ranges from two codons (6 nucleotides) to ten codons (30 nucleotides). Any of said sequences of the oligonucleotide(s) may contain restriction endonuclease recognition sequence(s). Any of said sequences of the oligonucleotide(s) may contain the restriction endonuclease recognition sequence(s) and/or promoter sequence(s) such as bacteriophage SP6 and/or T3 and/or T7 sequence(s) as the result(s) of an adaptor (linker) covalently linked to the 5'-end or and 3'-end of the said oligonucleotide(s). Any of said sequences of the oligonucleotide(s) may have one or two or three or four or five or six universal base analogue(s), such as 5'-Nitroindole covalently linked to the 5'-end of said

oligonucleotide(s). Any of said sequences of the oligonucleotide(s) may contain chemical modifications and substitutions of any sugars, phosphates, phosphodiester bonds, bases, base analogues, universal bases and polyamide. For example, 2'-O-alkylribose, 2'-O-Methylribonucleotide, Methylphosphonates, Phosphorothioate, Phosphordithioate, Sulfamate, H-phosphonate, [(alpha)]-anomeric nucleosides and any analogous structures e.g. 5'-Nitoindole, 3-nitropyrrole, inosine and pypoxamthine and other universal bases which deviates from, within, between and among the said genetic codes, base, nucleoside, nucleotide, nucleic acids sequences of the said oligonucleotide(s). The said oligonucleotide(s) analogue(s) could be Peptide Nucleic Acid (PNA). The said oligonucleotide(s) analogue(s) include the modified nucleotide units, which posses energy emission patterns of a light emitting chemical compound or a quenching compound, such as hypoxanthine, mercaptopurine, selenopurine, 2-aminopurine, 2,4-diselenouracil and 2,4-dithiouracil. Additionally, the said modifications and substitutions include modifications and substitutions known and/or under development and/or to be developed to the extent that such alterations facilitate or have no negative affect when the said oligonucleotide(s) hybridizes to the corresponding complementary targeting sequence(s). Any of said sequences of the oligonucleotide(s) may contain minor deletions, insertions and additions of codons or bases to the extent that such alterations facilitate or do not negatively affect when the said oligonucleotide(s) hybridize the corresponding complementary targeting sequence(s). Any of said sequences of the oligonucleotide(s) may be DNA, cDNA, mRNA, Anti-sense DNA, Anti-sense mRNA and Peptide Nucleic Acids (PNA) in the format of either single strand or double strands. Any of said sequences of the oligonucleotide(s) may be labelled by a chemical composition(s) which produces specific detectable signal by radioactive ray, electromagnetic radiation, immunochemistry, biochemistry and photochemistry. Those labelling chemical composition include radioisotopes such as 3.sup.H, 14.sup.C, 32.sup.P, 33.sup.P, and 35.sup.S; biotin; fluorescent molecules such as fluorescein isothiocyanate (FITC), Texas Red®, green fluorescent protein, rhodamines, tetramethylrhodamine isothiocyanate (TRITC), 4, 4-difluoro-4-bora-3a, 4a-diaza-s-indacene, lissamine, 5'-carboxy-fluorescein, 2',7'-dimethoxy-4',5'-dichloro-6 carboxy-fluorescein, phycoerythrin, Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7; enzymes such as alkaline phosphates, horse radish peroxidase; substrates; nucleotide chromophores; chemiluminescent moieties; bioluminescent moieties; phosphorescent compounds, magnetic particles etc. The analogue(s) also include natural peptide, polypeptide and protein which contained the chemical substitutions and chemical modifications of any

amino acid and any of its analogous structures which deviates from and within the said peptide, polypeptide and protein sequences. The said chemical modifications of any amino acid and any of its analogous include the hydroxylation, methylation, acetylation, carboxylation and phosphorylation. It is also included the addition of lipids and carbohydrate polymers to the side chains of amino acid residues of the peptides, polypeptides and proteins.

17. Synthesis of Oligonucleotide or and Peptide

In one preferred embodiment, synthesis of oligonucleotides was carried out by phosphoramidite methods such as Caruthers et al., *Nucleic Acids Res. Symp. Ser.* 7: 215-223, 1980; Beaucage et al., *Tetrahedron Lett.* 22: 1859-1862, 1981; McBride et al., *Tetrahedron Lett.* 24: 245-248, 1983; and Beaucage et al., *Tetrahedron Lett.* 48: 2223-2311, 1992.

In one preferred embodiment, the synthesis of oligonucleotides was processed by the H-phosphonate methods, such as Garegg et al., *Chem. Scripta* 25: 280-282, 1985; Garegg et al., *Chem. Scripta* 26: 59-62, 1986; Garegg et al., *Tetrahedron Lett.* 27: 4051-4054, 1986; Froehler et al., *Nucleic Acids Res.*, 14: 5399-5407, 1986; Froehler et al., *Tetrahedron Lett.* 27: 469-4472, 1986; Froehler et al., *Tetrahedron Lett.* 27: 5575-5578, 1986.

In another preferred embodiment, the synthesis of oligonucleotides was carried out by an automated nucleic acid synthesizer, such as ABI 381-A, ABI 391, ABI 392, ABI 3900 and Expedite® 8909 Nucleic Acid Synthesizer of PE Applied Biosystems at a 0.2 μ M scale using standard protocols in accordance with the manual of the manufacturer. Prior to the coupling step on the solid phase, the synthesized oligonucleotides then were purified, desalted and lyophilized at different grades of purity in accordance with different purposes, such as PCR®-grade (ethanol precipitation to remove the salt), Probe-grade (purified by HPLC) and/or Gene-synthesis-grade (purified by polyacrylamide gel electrophoresis). Such purification is known to those of skilled in the art.

In one preferred embodiment, at specific defined discrete positions on a solid phase for example the surface on silicon, in-situ synthesis of oligonucleotides was carried out by photolithographic methods such as Fodor et al., *Science* 251: 767-773, 1991; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* 91: 5022-5026, 1994; Lockhart et al., *Nature Biotech.* 14: 1675, 1996; Pirrung et al., U.S. Pat. No. 5,143, 854, 1992; Fodor et al., U.S. Pat. No. 5,445, 934, 1995; Fodor et al., U.S. Pat. No. 5,510, 270, 1996; Fodor et al., U.S. Pat. No. 5,800, 992, 1998. In another preferred embodiment, at specific defined discrete

position on the surface of glass plate, in-situ synthesis of oligonucleotides was processed in accordance with methods as described, such as Southern et al., Genomic13: 1008-1017, 1992; Maskos et al., Nucleic Acids Res. 20: 1679-1684, 1992; Southern et al., Nucleic Acids Res. 22: 1368-1373, 1994. In another embodiment, in-situ synthesis of oligonucleotides and deposition on the perfluorinated hydrophobic surface of silicon dioxide was processed by Ink-jet printer heads as described by Blanchard et al., Biosensors & Bioelectronics 11: 687-690.

At the present time, the synthesis of oligonucleotides and peptides and the production of monoclonal antibodies have become mature technology and standard laboratory operation procedures. Moreover, many companies such as Sigma-Genosys, Life Technologies and Washington Biotechnology Inc. provide routine service to produce the custom designed oligonucleotide, peptide and monoclonal antibodies tailored to different requirements and purposes. Additionally, the advances of computer and robotics technology have laid the framework for large scale production and production precision across all high-throughput processes in the Biotechnology domain. Those profound achievements mean one of skilled in the art could prepare oligonucleotides, peptides and monoclonal antibodies with undue experimentation.

XIII. EQUIVALENTS

While the preferred embodiments of the invention have been described above, it will be recognized and understood that various modifications may be made therein, and the appended claims are intended to cover all such modifications, which may fall within the spirit and scope of the invention. Taken together, the inventive methods, without limiting the generality of the foregoing, comprising a series of complex and combinatorial methods, working platforms and systems are comprehensive. Obviously, a unique, novel and useful method of gene signature, identification, cloning and determination has been described through the foregoing detailed illustrations and descriptions of various aspects, different examples and specific embodiments of the present invention. Although the specific embodiments and examples have been introduced and disclosed herein, it has been accomplished by way of example for the objectives of explanation and illustration only,

without limiting the generality of the foregoing, regarding the spirit and scope of the claims made for the invention. Specifically, it is contemplated by the inventors that various substitutions, alterations, modification, revisions and developments may be made in part or as the whole regarding both the structures or and the functions of the invention without departing from the spirit and the scope of the invention as defined by the claims. For example, the choices of nucleotides and amino acids from natural, synthetic or chemically modified resources respectively, the form of nucleic acids strands, such as sense or anti-sense, the particular labelling substances and the corresponding signal detection measurements, or the particular single or combinatorial oligonucleotide or peptide libraries are conceived as a matter of routine for one skilled in the art with knowledge of the embodiments described herein.

TABLE 1: 61 Codons

No.	Codons in DNA	Codons in mRNA	GC Content
1	5'-GCA	5'-GCA	66.67%
2	5'-TGC	5'-UGC	66.67%
3	5'-GAC	5'-GAC	66.67%
4	5'-GAA	5'-GAA	33.33%
5	5'-TTC	5'-UUC	33.33%
6	5'-GGA	5'-GGA	66.67%
7	5'-CAC	5'-CAC	66.67%
8	5'-ATA	5'-AUA	0.00%
9	5'-AAA	5'-AAA	0.00%
10	5'-TTA	5'-UUA	0.00%
11	5'-ATG	5'-AUG	33.33%
12	5'-AAC	5'-AAC	33.33%
13	5'-CCA	5'-CCA	66.67%
14	5'-CAA	5'-CAA	33.33%
15	5'-AGA	5'-AGA	33.33%
16	5'-AGC	5'-AGC	66.67%
17	5'-ACA	5'-ACA	33.33%
18	5'-GTA	5'-GUA	33.33%
19	5'-TGG	5'-UGG	66.67%
20	5'-TAC	5'-UAC	33.33%
21	5'-GCC	5'-GCC	100.00%
22	5'-TGT	5'-UGU	33.33%
23	5'-GAT	5'-GAU	33.33%
24	5'-GAG	5'-GAG	66.67%
25	5'-TTT	5'-UUU	0.00%
26	5'-GGC	5'-GGC	100.00%
27	5'-CAT	5'-CAU	33.33%
28	5'-ATC	5'-AUC	33.33%
29	5'-AAG	5'-AAG	33.33%
30	5'-TTG	5'-UUG	33.33%
31	5'-AAT	5'-AAU	0.00%
32	5'-CCC	5'-CCC	100.00%
33	5'-CAG	5'-CAG	66.67%
34	5'-AGG	5'-AGG	66.67%
35	5'-AGT	5'-AGU	33.33%
36	5'-ACC	5'-ACC	66.67%
37	5'-GTC	5'-GUC	66.67%
38	5'-TAT	5'-UAU	0.00%
39	5'-GCG	5'-GCG	100.00%
40	5'-GGG	5'-GGG	100.00%
41	5'-ATT	5'-AUU	0.00%
42	5'-CTA	5'-CUA	33.33%
43	5'-CCG	5'-CCG	100.00%
44	5'-CGA	5'-CGA	66.67%
45	5'-TCA	5'-UCA	33.33%

46	5'-ACG	5'-ACG	66.67%
47	5'-GTG	5'-GUG	66.67%
48	5'-GCT	5'-GCU	66.67%
49	5'-GGT	5'-GGU	66.67%
50	5'-CTC	5'-CUC	66.67%
51	5'-CCT	5'-CCU	66.67%
52	5'-CGC	5'-CGC	100.00%
53	5'-TCC	5'-UCC	66.67%
54	5'-ACT	5'-ACU	33.33%
55	5'-GTT	5'-GUU	33.33%
56	5'-CTG	5'-CUG	66.67%
57	5'-CGG	5'-CGG	100.00%
58	5'-TCG	5'-UCG	66.67%
59	5'-CTT	5'-CUU	33.33%
60	5'-CGT	5'-CGU	66.67%
61	5'-TCT	5'-UCU	33.33%

GC Content of 61 Codons: 100%: 8 66.7%: 24 33.3%: 22 0%: 7

TABLE 2: 61 Antisense Codons

No.	Antisense Codons in DNA	Antisense Codons in mRNA	GC Content
1	5'-TGC	5'-UGC	66.67%
2	5'-GCA	5'-GCA	66.67%
3	5'-GTC	5'-GUC	66.67%
4	5'-TTC	5'-UUC	33.33%
5	5'-GAA	5'-GAA	33.33%
6	5'-TCC	5'-UCC	66.67%
7	5'-GTG	5'-GUG	66.67%
8	5'-TAT	5'-UAU	0.00%
9	5'-TTT	5'-UUU	0.00%
10	5'-TAA	5'-UAA	0.00%
11	5'-CAT	5'-CAU	33.33%
12	5'-GTT	5'-GUU	33.33%
13	5'-TGG	5'-UGG	66.67%
14	5'-TTG	5'-UUG	33.33%
15	5'-TCT	5'-UCU	33.33%
16	5'-GCT	5'-GCU	66.67%
17	5'-TGT	5'-UGU	33.33%
18	5'-TAC	5'-UAC	33.33%
19	5'-CCA	5'-CCA	66.67%
20	5'-GTA	5'-GUA	33.33%
21	5'-GGC	5'-GGC	100.00%
22	5'-ACA	5'-ACA	33.33%
23	5'-ATC	5'-AUC	33.33%
24	5'-CTC	5'-CUC	66.67%
25	5'-AAA	5'-AAA	0.00%

95

26	5'-GCC	5'-GCC	100.00%
27	5'-ATG	5'-AUG	33.33%
28	5'-GAT	5'-GAU	33.33%
29	5'-CTT	5'-CUU	33.33%
30	5'-CAA	5'-CAA	33.33%
31	5'-ATT	5'-AUU	0.00%
32	5'-GGG	5'-GGG	100.00%
33	5'-CTG	5'-CUG	66.67%
34	5'-CCT	5'-CCU	66.67%
35	5'-ACT	5'-ACU	33.33%
36	5'-GGT	5'-GGU	66.67%
37	5'-GAC	5'-GAC	66.67%
38	5'-ATA	5'-AUA	0.00%
39	5'-CGC	5'-CGC	100.00%
40	5'-CCC	5'-CCC	100.00%
41	5'-AAT	5'-AAU	0.00%
42	5'-TAG	5'-UAG	33.33%
43	5'-CGG	5'-CGG	100.00%
44	5'-TCG	5'-UCG	66.67%
45	5'-TGA	5'-UGA	33.33%
46	5'-CGT	5'-CGU	66.67%
47	5'-CAC	5'-CAC	66.67%
48	5'-AGC	5'-AGC	66.67%
49	5'-ACC	5'-ACC	66.67%
50	5'-GAG	5'-GAG	66.67%
51	5'-AGG	5'-AGG	66.67%
52	5'-GCG	5'-GCG	100.00%
53	5'-GGA	5'-GGA	66.67%
54	5'-AGT	5'-AGU	33.33%
55	5'-AAC	5'-AAC	33.33%
56	5'-CAG	5'-CAG	66.67%
57	5'-CCG	5'-CCG	100.00%
58	5'-CGA	5'-CGA	66.67%
59	5'-AAG	5'-AAG	33.33%
60	5'-ACG	5'-ACG	66.67%
61	5'-AGA	5'-AGA	33.33%

GC Content of 61 Antisense Codons: 100%: 8 66.7%: 24 33.3%: 22 0%: 7

TABLE 3: The Restriction Enzyme Recognition Sequences (Two-codon)

No.	Restriction Enzyme	DNA Recognition Sense Sequence	Corresponding Recognition Sequence of Amino Acids	DNA Anti-sense Recognition Sequence
1	<i>Aat</i> II	5'-GACGTC	NH2-DV	5'-GACGTC
2	<i>Acc</i> 65 I	5'-GGTACC	NH2-GT	5'-GGTACC
3	<i>Acc</i> I	5'-AACGTT	NH2-NV	5'-AACGTT
4	<i>Afe</i> I	5'-AGCGCT	NH2-SA	5'-AGCGCT
5	<i>Afl</i> II	5'-CTTAAG	NH2-LK	5'-CTTAAG
6	<i>Age</i> I	5'-ACCGGT	NH2-TG	5'-ACCGGT
7	<i>Apa</i> I	5'-GGGCCC	NH2-GP	5'-GGGCCC
8	<i>Apa</i> L I	5'-GTGCAC	NH2-VH	5'-GTGCAC
9	<i>Ase</i> I	5'-ATTAAT	NH2-IN	5'-ATTAAT
10	<i>Avr</i> II	5'-CCTAGG	NH2-PR	5'-CCTAGG
11	<i>Bam</i> H I	5'-GGATCC	NH2-GS	5'-GGATCC
12	<i>Bfr</i> B I	5'-ATGCAT	NH2-MH	5'-ATGCAT
13	<i>Bmg</i> B I	5'-CACGTC	NH2-HV	5'-GACGTG
14	<i>Bgl</i> II	5'-AGATCT	NH2-RS	5'-AGATCT
15	<i>Bse</i> Y I	5'-CCCAGC	NH2-PS	5'-GCTGGG
16	<i>Btr</i> I	5'-CACGTC	NH2-HV	5'-GACGTG
17	<i>Bsi</i> W I	5'-CGTACG	NH2-RT	5'-CGTACG
18	<i>Bsp</i> D I	5'-ATCGAT	NH2-ID	5'-ATCGAT
19	<i>Bsp</i> E I	5'-TCCGGA	NH2-SG	5'-TCCGGA
20	<i>Bsr</i> B I	5'-GAGCGG	NH2-ER	5'-CCGCTC
21	<i>Bsr</i> G I	5'-TGTACA	NH2-CT	5'-TGTACA
22	<i>Bss</i> H II	5'-GCGCGC	NH2-AR	5'-GCGCGC
23	<i>Bss</i> S I	5'-CTCGTG	NH2-LV	5'-CACGAG
24	<i>Bst</i> B I	5'-TTCGAA	NH2-FE	5'-TTCGAA
25	<i>Bst</i> Z17 I	5'-GTATAC	NH2-VY	5'-GTATAC
26	<i>Cla</i> I	5'-ATCGAT	NH2-ID	5'-ATCGAT
27	<i>Dra</i> I	5'-TTTAAA	NH2-FK	5'-TTTAAA
28	<i>Eag</i> I	5'-CGGCCG	NH2-RP	5'-CGGCCG
29	<i>Eco</i> R I	5'-GAATTC	NH2-EF	5'-GAATTC
30	<i>Eco</i> R V	5'-GATATC	NH2-DI	5'-GATATC
31	<i>Fsp</i> I	5'-TGCGCA	NH2-CA	5'-TGCGCA
32	<i>Hind</i> III	5'-AAGCTT	NH2-KL	5'-AAGCTT
33	<i>Hpa</i> I	5'-GTTAAC	NH2-VN	5'-GTTAAC
34	<i>Kas</i> I	5'-GGCGCC	NH2-GA	5'-GGCGCC
35	<i>Kpn</i> I	5'-GGTACC	NH2-GT	5'-GGTACC
36	<i>Mfe</i> I	5'-CAATTG	NH2-QL	5'-CAATTG
37	<i>Mlu</i> I	5'-ACGCGT	NH2-TR	5'-ACGCGT
38	<i>Msc</i> I	5'-TGGCCA	NH2-WP	5'-TGGCCA
39	<i>Nae</i> I	5'-GCCGGC	NH2-AG	5'-GCCGGC
40	<i>Nar</i> I	5'-GGCGCC	NH2-GA	5'-GGCGCC
41	<i>Nco</i> I	5'-CCATGG	NH2-PW	5'-CCATGG
42	<i>Nde</i> I	5'-CATATG	NH2-HM	5'-CATATG
43	<i>Ngo</i> M IV	5'-GCCGGC	NH2-AG	5'-GCCGGC
44	<i>Nhe</i> I	5'-GCTAGC	NH2-AS	5'-GCTAGC
45	<i>Nru</i> I	5'-TCGCGA	NH2-SR	5'-TCGCGA

46	<i>Nsi</i> I	5'-ATGCAT	NH2-MH	5'-ATGCAT
47	<i>Pae</i> R7 I	5'-CTCGAG	NH2-LE	5'-CTCGAG
48	<i>Pci</i> I	5'-ACATGT	NH2-TC	5'-ACATGT
49	<i>Pml</i> I	5'-CACGTG	NH2-HV	5'-CACGTG
50	<i>Psp</i> OM I	5'-GGGCCC	NH2-GP	5'-GGGCCC
51	<i>Pst</i> I	5'-CTGCAG	NH2-LQ	5'-CTGCAG
52	<i>Pvu</i> I	5'-CGATCG	NH2-RS	5'-CGATCG
53	<i>Pvu</i> II	5'-CAGCTG	NH2-QL	5'-CAGCTG
54	<i>Sac</i> I	5'-GAGCTC	NH2-EL	5'-GAGCTC
55	<i>Sac</i> II	5'-CCGCGG	NH2-PR	5'-CCGCGG
56	<i>Sal</i> I	5'-GTCGAC	NH2-VD	5'-GTCGAC
57	<i>Sca</i> I	5'-AGTACT	NH2-ST	5'-AGTACT
58	<i>Sfo</i> I	5'-GGCGCC	NH2-GA	5'-GGCGCC
59	<i>Sma</i> I	5'-CCCGGG	NH2-PG	5'-CCCGGG
60	<i>Sna</i> B I	5'-TACGTA	NH2-YV	5'-TACGTA
61	<i>Spe</i> I	5'-ACTAGT	NH2-TS	5'-ACTAGT
62	<i>Sph</i> I	5'-GCATGC	NH2-AC	5'-GCATGC
63	<i>Ssp</i> I	5'-AATATT	NH2-NI	5'-AATATT
64	<i>Stu</i> I	5'-AGGCCT	NH2-RP	5'-AGGCCT
65	<i>Tli</i> I	5'-CTCGAG	NH2-LE	5'-CTCGAG
66	<i>Xba</i> I	5'-TCTAGA	NH2-SR	5'-TCTAGA
67	<i>Xho</i> I	5'-CTCGAG	NH2-LE	5'-CTCGAG
68	<i>Xma</i> I	5'-CCCGGG	NH2-PG	5'-CCCGGG
69	<i>Acc</i> I	5'-GTATAC	NH2-VY	5'-GTATAC
	<i>Acc</i> I	5'-GTCTAC	NH2-VY	5'-GTAGAC
	<i>Acc</i> I	5'-GTAGAC	NH2-VD	5'-GTCTAC
	<i>Acc</i> I	5'-GTCGAC	NH2-VD	5'-GTCGAC
70	<i>Bme</i> 1580 I	5'-GGGCAC	NH2-GH	5'-GTGCCC
	<i>Bme</i> 1580 I	5'-GGGCCC	NH2-GP	5'-GGGCCC
	<i>Bme</i> 1580 I	5'-GTGCAC	NH2-VH	5'-GTGCAC
	<i>Bme</i> 1580 I	5'-GTGCCC	NH2-VP	5'-GGGCAC
71	<i>Bsa</i> W I	5'-ACCGGA	NH2-TG	5'-TCCGGT
	<i>Bsa</i> W I	5'-ACCGGT	NH2-TG	5'-ACCGGT
	<i>Bsa</i> W I	5'-TCCGGA	NH2-SG	5'-TCCGGA
	<i>Bsa</i> W I	5'-TCCGGT	NH2-SG	5'-ACCGGA
72	<i>Bsi</i> HKA I	5'-GAGCAC	NH2-EH	5'-GTGCTC
	<i>Bsi</i> HKA I	5'-GAGCTC	NH2-EL	5'-GAGCTC
	<i>Bsi</i> HKA I	5'-GTGCAC	NH2-VH	5'-GTGCAC
	<i>Bsi</i> HKA I	5'-GTGCTC	NH2-VL	5'-GAGCAC
	<i>Bsp</i> 1286 I	5'-GGGCCC	NH2-GP	5'-GGGCCC
73	<i>Bsp</i> 1286 I	5'-GAGCCC	NH2-EP	5'-GGGCTC
	<i>Bsp</i> 1286 I	5'-GTGCCC	NH2-VP	5'-GGGCAC
	<i>Bsp</i> 1286 I	5'-GAGCAC	NH2-EH	5'-GTGCTC
	<i>Bsp</i> 1286 I	5'-GTGCAC	NH2-VH	5'-GTGCAC
	<i>Bsp</i> 1286 I	5'-GAGCTC	NH2-EL	5'-GAGCTC
74	<i>Msp</i> A1 I	5'-CAGCGG	NH2-QR	5'-CCGCTG
	<i>Msp</i> A1 I	5'-CAGCTG	NH2-QL	5'-CAGCTG
	<i>Msp</i> A1 I	5'-CCGCGG	NH2-PR	5'-CCGCGG
	<i>Msp</i> A1 I	5'-CCGCTG	NH2-PL	5'-CAGCGG
75	<i>Sty</i> I	5'-CCAAGG	NH2-PR	5'-CCTTGG

Styl

5'-CCTAGG

NH2-PR

5'-CCTAGG

Styl

5'-CCATGG

NH2-PW

5'-CCATGG

Styl

5'-CCTTGG

NH2-PW

5'-CCAAGG

TABLE 4: 64 Codons

No.	Codons in DNA	Codons in mRNA	GC Content
1	5'-GCA	5'-GCA	66.67%
2	5'-TGC	5'-UGC	66.67%
3	5'-GAC	5'-GAC	66.67%
4	5'-GAA	5'-GAA	33.33%
5	5'-TTC	5'-UUC	33.33%
6	5'-GGA	5'-GGA	66.67%
7	5'-CAC	5'-CAC	66.67%
8	5'-ATA	5'-AUA	0.00%
9	5'-AAA	5'-AAA	0.00%
10	5'-TTA	5'-UUA	0.00%
11	5'-ATG	5'-AUG	33.33%
12	5'-AAC	5'-AAC	33.33%
13	5'-CCA	5'-CCA	66.67%
14	5'-CAA	5'-CAA	33.33%
15	5'-AGA	5'-AGA	33.33%
16	5'-AGC	5'-AGC	66.67%
17	5'-ACA	5'-ACA	33.33%
18	5'-GTA	5'-GUA	33.33%
19	5'-TGG	5'-UGG	66.67%
20	5'-TAC	5'-UAC	33.33%
21	5'-GCC	5'-GCC	100.00%
22	5'-TGT	5'-UGU	33.33%
23	5'-GAT	5'-GAU	33.33%
24	5'-GAG	5'-GAG	66.67%
25	5'-TTT	5'-UUU	0.00%
26	5'-GGC	5'-GGC	100.00%
27	5'-CAT	5'-CAU	33.33%
28	5'-ATC	5'-AUC	33.33%
29	5'-AAG	5'-AAG	33.33%
30	5'-TTG	5'-UUG	33.33%
31	5'-AAT	5'-AAU	0.00%
32	5'-CCC	5'-CCC	100.00%
33	5'-CAG	5'-CAG	66.67%
34	5'-AGG	5'-AGG	66.67%
35	5'-AGT	5'-AGU	33.33%
36	5'-ACC	5'-ACC	66.67%
37	5'-GTC	5'-GUC	66.67%
38	5'-TAT	5'-UAU	0.00%
39	5'-GCG	5'-GCG	100.00%
40	5'-GGG	5'-GGG	100.00%

41	5'-ATT	5'-AUU	0.00%
42	5'-CTA	5'-CUA	33.33%
43	5'-CCG	5'-CCG	100.00%
44	5'-CGA	5'-CGA	66.67%
45	5'-TCA	5'-UCA	33.33%
46	5'-ACG	5'-ACG	66.67%
47	5'-GTG	5'-GUG	66.67%
48	5'-GCT	5'-GCU	66.67%
49	5'-GGT	5'-GGU	66.67%
50	5'-CTC	5'-CUC	66.67%
51	5'-CCT	5'-CCU	66.67%
52	5'-CGC	5'-CGC	100.00%
53	5'-TCC	5'-UCC	66.67%
54	5'-ACT	5'-ACU	33.33%
55	5'-GTT	5'-GUU	33.33%
56	5'-CTG	5'-CUG	66.67%
57	5'-CGG	5'-CGG	100.00%
58	5'-TCG	5'-UCG	66.67%
59	5'-CTT	5'-CUU	33.33%
60	5'-CGT	5'-CGU	66.67%
61	5'-TCT	5'-UCU	33.33%
62	5'-TAA	5'-UAA	0.00%
63	5'-TGA	5'-UGA	33.33%
64	5'-TAG	5'-UAG	33.33%

GC Content of 64 Codons: 100%: 8 66.7%: 24 33.3%: 24 0%: 8

TABLE 5: 64 Antisense Codons

No.	Antisense Codons in DNA	Antisense Codons in mRNA	GC Content
1	5'-TGC	5'-UGC	66.67%
2	5'-GCA	5'-GCA	66.67%
3	5'-GTC	5'-GUC	66.67%
4	5'-TTC	5'-UUC	33.33%
5	5'-GAA	5'-GAA	33.33%
6	5'-TCC	5'-UCC	66.67%
7	5'-GTG	5'-GUG	66.67%
8	5'-TAT	5'-UAU	0.00%
9	5'-TTT	5'-UUU	0.00%
10	5'-TAA	5'-UAA	0.00%
11	5'-CAT	5'-CAU	33.33%
12	5'-GTT	5'-GUU	33.33%
13	5'-TGG	5'-UGG	66.67%
14	5'-TTG	5'-UUG	33.33%
15	5'-TCT	5'-UCU	33.33%
16	5'-GCT	5'-GCU	66.67%
17	5'-TGT	5'-UGU	33.33%

100

18	5'-TAC	5'-UAC	33.33%
19	5'-CCA	5'-CCA	66.67%
20	5'-GTA	5'-GUA	33.33%
21	5'-GGC	5'-GGC	100.00%
22	5'-ACA	5'-ACA	33.33%
23	5'-ATC	5'-AUC	33.33%
24	5'-CTC	5'-CUC	66.67%
25	5'-AAA	5'-AAA	0.00%
26	5'-GCC	5'-GCC	100.00%
27	5'-ATG	5'-AUG	33.33%
28	5'-GAT	5'-GAU	33.33%
29	5'-CTT	5'-CUU	33.33%
30	5'-CAA	5'-CAA	33.33%
31	5'-ATT	5'-AUU	0.00%
32	5'-GGG	5'-GGG	100.00%
33	5'-CTG	5'-CUG	66.67%
34	5'-CCT	5'-CCU	66.67%
35	5'-ACT	5'-ACU	33.33%
36	5'-GGT	5'-GGU	66.67%
37	5'-GAC	5'-GAC	66.67%
38	5'-ATA	5'-AUA	0.00%
39	5'-CGC	5'-CGC	100.00%
40	5'-CCC	5'-CCC	100.00%
41	5'-AAT	5'-AAU	0.00%
42	5'-TAG	5'-UAG	33.33%
43	5'-CGG	5'-CGG	100.00%
44	5'-TCG	5'-UCG	66.67%
45	5'-TGA	5'-UGA	33.33%
46	5'-CGT	5'-CGU	66.67%
47	5'-CAC	5'-CAC	66.67%
48	5'-AGC	5'-AGC	66.67%
49	5'-ACC	5'-ACC	66.67%
50	5'-GAG	5'-GAG	66.67%
51	5'-AGG	5'-AGG	66.67%
52	5'-GCG	5'-GCG	100.00%
53	5'-GGA	5'-GGA	66.67%
54	5'-AGT	5'-AGU	33.33%
55	5'-AAC	5'-AAC	33.33%
56	5'-CAG	5'-CAG	66.67%
57	5'-CCG	5'-CCG	100.00%
58	5'-CGA	5'-CGA	66.67%
59	5'-AAG	5'-AAG	33.33%
60	5'-ACG	5'-ACG	66.67%
61	5'-AGA	5'-AGA	33.33%
62	5'-TTA	5'-UUA	0.00%
63	5'-TCA	5'-UCA	33.33%
64	5'-CTA	5'-CUA	33.33%

GC Content of 64 Antisense Codons: 100%: 8 66.7%: 24 33.3%: 24 0%: 8

**TABLE 6: A Systematic Sequence Deducing Method for
ORF Sense Sequential Sequence with 5'-terminal Start Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
1	+1,	+2,	+3,	3mers	$61^{1-1} / 61^0$
2	+4,	+5,	+6,	6mers	$61^{2-1} / 61^1$
3	+7,	+8,	+9,	9mers	$61^{3-1} / 61^2$
4	+10,	+11,	+12,	12mers	$61^{4-1} / 61^3$
5	+13,	+14,	+15,	15mers	$61^{5-1} / 61^4$
6	+16,	+17,	+18,	18mers	$61^{6-1} / 61^5$
7	+19,	+20,	+21,	21mers	$61^{7-1} / 61^6$
8	+22,	+23,	+24,	24mers	$61^{8-1} / 61^7$
9	+25,	+26,	+27,	27mers	$61^{9-1} / 61^8$
10	+28,	+29,	+30,	30mers	$61^{10-1} / 61^9$
11	+31,	+32,	+33,	33mers	$61^{11-1} / 61^{10}$
12	+34,	+35,	+36,	36mers	$61^{12-1} / 61^{11}$
13	+37,	+38,	+39,	39mers	$61^{13-1} / 61^{12}$
14	+40,	+41,	+42,	42mers	$61^{14-1} / 61^{13}$
15	+43,	+44,	+45,	45mers	$61^{15-1} / 61^{14}$
16	+46,	+47,	+48,	48mers	$61^{16-1} / 61^{15}$
17	+49,	+50,	+51,	51mers	$61^{17-1} / 61^{16}$
18	+52,	+53,	+54,	54mers	$61^{18-1} / 61^{17}$
19	+55,	+56,	+57,	57mers	$61^{19-1} / 61^{18}$
20	+58,	+59,	+60,	60mers	$61^{20-1} / 61^{19}$
21	+61,	+62,	+63,	63mers	$61^{21-1} / 61^{20}$
22	+64,	+65,	+66,	66mers	$61^{22-1} / 61^{21}$
23	+67,	+68,	+69,	69mers	$61^{23-1} / 61^{22}$
24	+70,	+71,	+72,	72mers	$61^{24-1} / 61^{23}$
25	+73,	+74,	+75,	75mers	$61^{25-1} / 61^{24}$
26	+76,	+77,	+78,	78mers	$61^{26-1} / 61^{25}$
27	+79,	+80,	+81,	81mers	$61^{27-1} / 61^{26}$
28	+82,	+83,	+84,	84mers	$61^{28-1} / 61^{27}$
29	+85,	+86,	+87,	87mers	$61^{29-1} / 61^{28}$
30	+88,	+89,	+90,	90mers	$61^{30-1} / 61^{29}$
31	+91,	+92,	+93,	93mers	$61^{31-1} / 61^{30}$
32	+94,	+95,	+96,	96mers	$61^{32-1} / 61^{31}$
33	+97,	+98,	+99,	99mers	$61^{33-1} / 61^{32}$
34	+100,	+101,	+102,	102mers	$61^{34-1} / 61^{33}$
n	+(3n-2),+(3n-1), +3n			3nmers	$(\text{GCOEAA})^{n-1} / (\text{GCOEAA})^E$

*The first position of codon at 5'-terminal of ORF sequences is occupied by any one of the Start Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. E=n-1 E: Exponent

The n is the serial number of Codon for both 5'-end and 3'-end orientated ORF sense sequences.

The n is the serial number of Codon begins at both 5'-end and 3'-end orientated ORF sense sequences.

GCOEAA: Genetic Codes Of the 20 Essential Amino Acids / Codons Of the 20 Essential Amino Acids

 $(\text{GCOEAA})^{n-1} = (\text{GCOEAA})^E$ When GCOEAA=61, $61^{n-1} = 61^E$. [61.sup.(n-1) = 61.sup.E]. E=n-1

The methods are applicable when GCOEAA=61, GCOEAA<61 and GCOEAA>61

**TABLE 7: A Systematic Sequence Deducing Method for
ORF Sense Sequential Sequence with 3'-terminal Stop Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+3,	+2,	+1,	3mers	$61^{1-1} / 61^0$
2	+6,	+5,	+4,	6mers	$61^{2-1} / 61^1$
3	+9,	+8,	+7,	9mers	$61^{3-1} / 61^2$
4	+12,	+11,	+10,	12mers	$61^{4-1} / 61^3$
5	+15,	+14,	+13,	15mers	$61^{5-1} / 61^4$
6	+18,	+17,	+16,	18mers	$61^{6-1} / 61^5$
7	+21,	+20,	+19,	21mers	$61^{7-1} / 61^6$
8	+24,	+23,	+22,	24mers	$61^{8-1} / 61^7$
9	+27,	+26,	+25,	27mers	$61^{9-1} / 61^8$
10	+30,	+29,	+28,	30mers	$61^{10-1} / 61^9$
11	+33,	+32,	+31,	33mers	$61^{11-1} / 61^{10}$
12	+36,	+35,	+34,	36mers	$61^{12-1} / 61^{11}$
13	+39,	+38,	+37,	39mers	$61^{13-1} / 61^{12}$
14	+42,	+41,	+40,	42mers	$61^{14-1} / 61^{13}$
15	+45,	+44,	+43,	45mers	$61^{15-1} / 61^{14}$
16	+48,	+47,	+46,	48mers	$61^{16-1} / 61^{15}$
17	+51,	+50,	+49,	51mers	$61^{17-1} / 61^{16}$
18	+54,	+53,	+52,	54mers	$61^{18-1} / 61^{17}$
19	+57,	+56,	+55,	57mers	$61^{19-1} / 61^{18}$
20	+60,	+59,	+58,	60mers	$61^{20-1} / 61^{19}$
21	+63,	+62,	+61,	63mers	$61^{21-1} / 61^{20}$
22	+66,	+65,	+64,	66mers	$61^{22-1} / 61^{21}$
23	+69,	+68,	+67,	69mers	$61^{23-1} / 61^{22}$
24	+72,	+71,	+70,	72mers	$61^{24-1} / 61^{23}$
25	+75,	+74,	+73,	75mers	$61^{25-1} / 61^{24}$
26	+78,	+77,	+76,	78mers	$61^{26-1} / 61^{25}$
27	+81,	+80,	+79,	81mers	$61^{27-1} / 61^{26}$
28	+84,	+83,	+82,	84mers	$61^{28-1} / 61^{27}$
29	+87,	+86,	+85,	87mers	$61^{29-1} / 61^{28}$
30	+90,	+89,	+88,	90mers	$61^{30-1} / 61^{29}$
31	+93,	+92,	+91,	93mers	$61^{31-1} / 61^{30}$
32	+96,	+95,	+94,	96mers	$61^{32-1} / 61^{31}$
33	+99,	+98,	+97,	99mers	$61^{33-1} / 61^{32}$
34	+102,	+101,	+100,	102mers	$61^{34-1} / 61^{33}$
n	+3n,	+(3n-1),	+(3n-2)	3nmers	$(GCOEAA)^{n-1} / (GCOEAA)^E$

* The first position of codon at 3'-terminal of ORF sequences is occupied by any one of the Stop Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E: Exponent

The n is the serial number of Codon for both 5'-end and 3'-end orientated ORF sense sequences.

The n is the serial number of Codon begins at both 5'-end and 3'-end orientated ORF sense sequences.

GCOEAA: Genetic Codes Of the 20 Essential Amino Acids / Codons Of the 20 Essential Amino Acids

$(GCOEAA)^{n-1} = (GCOEAA)^E$ When $GCOEAA=61$, $61^{n-1} = 61^E$. $[61.sup.(n-1) = 61.sup.E]$. $E=n-1$

The methods are applicable when $GCOEAA=61$, $GCOEAA<61$ and $GCOEAA>61$

**TABLE 8: A Systematic Sequence Deducing Method for
5'-UTR Sense Sequential Sequence with 3'-terminal Start Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+1,	+2,	+3,	3mers	$64^{1-1} / 64^0$
2	-3,	-2,	-1,	6mers	$64^{2-1} / 64^1$
3	-6,	-5,	-4,	9mers	$64^{3-1} / 64^2$
4	-9,	-8,	-7,	12mers	$64^{4-1} / 64^3$
5	-12,	-11,	-11,	15mers	$64^{5-1} / 64^4$
6	-15,	-14,	-13,	18mers	$64^{6-1} / 64^5$
7	-18,	-17,	-16,	21mers	$64^{7-1} / 64^6$
8	-21,	-20,	-19,	24mers	$64^{8-1} / 64^7$
9	-24,	-23,	-22,	27mers	$64^{9-1} / 64^8$
10	-27,	-26,	-25,	30mers	$64^{10-1} / 64^9$
11	-30,	-29,	-28,	33mers	$64^{11-1} / 64^{10}$
12	-33,	-32,	-31,	36mers	$64^{12-1} / 64^{11}$
13	-36,	-35,	-34,	39mers	$64^{13-1} / 64^{12}$
14	-39,	-38,	-37,	42mers	$64^{14-1} / 64^{13}$
15	-42,	-41,	-40,	45mers	$64^{15-1} / 64^{14}$
16	-45,	-44,	-43,	48mers	$64^{16-1} / 64^{15}$
17	-48,	-47,	-46,	51mers	$64^{17-1} / 64^{16}$
18	-51,	-50,	-49,	54mers	$64^{18-1} / 64^{17}$
19	-54,	-53,	-52,	57mers	$64^{19-1} / 64^{18}$
20	-57,	-56,	-55,	60mers	$64^{20-1} / 64^{19}$
21	-60,	-59,	-58,	63mers	$64^{21-1} / 64^{20}$
22	-63,	-62,	-61,	66mers	$64^{22-1} / 64^{21}$
23	-66,	-65,	-64,	69mers	$64^{23-1} / 64^{22}$
24	-69,	-68,	-67,	72mers	$64^{24-1} / 64^{23}$
25	-72,	-71,	-70,	75mers	$64^{25-1} / 64^{24}$
26	-75,	-74,	-73,	78mers	$64^{26-1} / 64^{25}$
27	-78,	-77,	-76,	81mers	$64^{27-1} / 64^{26}$
28	-81,	-80,	-79,	84mers	$64^{28-1} / 64^{27}$
29	-84,	-83,	-82,	87mers	$64^{29-1} / 64^{28}$
30	-87,	-86,	-85,	90mers	$64^{30-1} / 64^{29}$
31	-90,	-89,	-88,	93mers	$64^{31-1} / 64^{30}$
32	-93,	-92,	-91,	96mers	$64^{32-1} / 64^{31}$
33	-96,	-95,	-94,	99mers	$64^{33-1} / 64^{32}$
34	-99,	-98,	-97,	102mers	$64^{34-1} / 64^{33}$
When $n = 1, (1-n)+1$	$3(1-n)+1, 3(1-n)+2, 3(1-n)+3$			3nmers	$(\text{GCOEAA})^{n-1} / (\text{GCOEAA})^E$
When $n > 1, (1-n)$	$3(1-n), 3(1-n)+1, 3(1-n)+2$			3nmers	$(\text{GCOEAA})^{n-1} / (\text{GCOEAA})^E$

*The first position of codon at 3'-terminal of 5'-UTR sense sequences is occupied by any one of the Stop Codons.
The n is from 1 to infinity. The n is an integer. $E=n-1$ E: Exponent The negative sign means position in 5'-UTR.
The n is the serial number of Codon for 3'-end orientated 5'-UTR sense sequences. n is not equal zero.
The n is the serial number of Codon begins 3'-end orientated 5'-UTR sense sequences.
GCOEAA: Genetic Codes Of the 20 Essential Amino Acids / Codons Of the 20 Essential Amino Acids
 $(\text{GCOEAA})^{n-1} = (\text{GCOEAA})^E$ When $\text{GCOEAA}=61, 61^{n-1} = 61^E$. [61.sup.(n-1) = 61.sup.E]. $E=n-1$
The methods are applicable when $\text{GCOEAA}=61, \text{GCOEAA}<61$ and $\text{GCOEAA}>61$

**TABLE 9: A Systematic Sequence Deducing Method for
3'-UTR Sense Sequential Sequence with 5'- terminal Stop Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+1,	+2,	+3,	3mers	$64^{1-1} / 64^0$
2	+4,	+5,	+6,	6mers	$64^{2-1} / 64^1$
3	+7,	+8,	+9,	9mers	$64^{3-1} / 64^2$
4	+10,	+11,	+12,	12mers	$64^{4-1} / 64^3$
5	+13,	+14,	+15,	15mers	$64^{5-1} / 64^4$
6	+16,	+17,	+18,	18mers	$64^{6-1} / 64^5$
7	+19,	+20,	+21,	21mers	$64^{7-1} / 64^6$
8	+22,	+23,	+24,	24mers	$64^{8-1} / 64^7$
9	+25,	+26,	+27,	27mers	$64^{9-1} / 64^8$
10	+28,	+29,	+30,	30mers	$64^{10-1} / 64^9$
11	+31,	+32,	+33,	33mers	$64^{11-1} / 64^{10}$
12	+34,	+35,	+36,	36mers	$64^{12-1} / 64^{11}$
13	+37,	+38,	+39,	39mers	$64^{13-1} / 64^{12}$
14	+40,	+41,	+42,	42mers	$64^{14-1} / 64^{13}$
15	+43,	+44,	+45,	45mers	$64^{15-1} / 64^{14}$
16	+46,	+47,	+48,	48mers	$64^{16-1} / 64^{15}$
17	+49,	+50,	+51,	51mers	$64^{17-1} / 64^{16}$
18	+52,	+53,	+54,	54mers	$64^{18-1} / 64^{17}$
19	+55,	+56,	+57,	57mers	$64^{19-1} / 64^{18}$
20	+58,	+59,	+60,	60mers	$64^{20-1} / 64^{19}$
21	+61,	+62,	+63,	63mers	$64^{21-1} / 64^{20}$
22	+64,	+65,	+66,	66mers	$64^{22-1} / 64^{21}$
23	+67,	+68,	+69,	69mers	$64^{23-1} / 64^{22}$
24	+70,	+71,	+72,	72mers	$64^{24-1} / 64^{23}$
25	+73,	+74,	+75,	75mers	$64^{25-1} / 64^{24}$
26	+76,	+77,	+78,	78mers	$64^{26-1} / 64^{25}$
27	+79,	+80,	+81,	81mers	$64^{27-1} / 64^{26}$
28	+82,	+83,	+84,	84mers	$64^{28-1} / 64^{27}$
29	+85,	+86,	+87,	87mers	$64^{29-1} / 64^{28}$
30	+88,	+89,	+90,	90mers	$64^{30-1} / 64^{29}$
31	+91,	+92,	+93,	93mers	$64^{31-1} / 64^{30}$
32	+94,	+95,	+96,	96mers	$64^{32-1} / 64^{31}$
33	+97,	+98,	+99,	99mers	$64^{33-1} / 64^{32}$
34	+100,	+101,	+102,	102mers	$64^{34-1} / 64^{33}$
n	+(3n-2), +(3n-1), +3n			3nmers	$(\text{GCOEAA})^{n-1} / (\text{GCOEAA})^E$

*The first position of codon at 5'-terminal of 3'-UTR sense sequences is occupied by any one of the Stop Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E: Exponent

The n is the serial number of Codon for 5'-end orientated 3'-UTR sense sequences.

The n is the serial number of Codon begins at 5'-end orientated 3'-UTR sense sequences.

GCOEAA: Genetic Codes Of the 20 Essential Amino Acids / Codons Of the 20 Essential Amino Acids

$(\text{GCOEAA})^{n-1} = (\text{GCOEAA})^E$ When $\text{GCOEAA}=61$, $61^{n-1} = 61^E$. $[61.\text{sup.}(n-1) = 61.\text{sup.}E]$. $E=n-1$

The methods are applicable when $\text{GCOEAA}=61$, $\text{GCOEAA}<61$ and $\text{GCOEAA}>61$

**TABLE 10: A Systematic Sequence Deducing Method for
Anti-sense ORF Sequential Sequences with 3'-terminal Anti-sense Start Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+3,	+2,	+1,	3mers	$61^{1-1} / 61^0$
2	+6,	+5,	+4,	6mers	$61^{2-1} / 61^1$
3	+9,	+8,	+7,	9mers	$61^{3-1} / 61^2$
4	+12,	+11,	+10,	12mers	$61^{4-1} / 61^3$
5	+15,	+14,	+13,	15mers	$61^{5-1} / 61^4$
6	+18,	+17,	+16,	18mers	$61^{6-1} / 61^5$
7	+21,	+20,	+19,	21mers	$61^{7-1} / 61^6$
8	+24,	+23,	+22,	24mers	$61^{8-1} / 61^7$
9	+27,	+26,	+25,	27mers	$61^{9-1} / 61^8$
10	+30,	+29,	+28,	30mers	$61^{10-1} / 61^9$
11	+33,	+32,	+31,	33mers	$61^{11-1} / 61^{10}$
12	+36,	+35,	+34,	36mers	$61^{12-1} / 61^{11}$
13	+39,	+38,	+37,	39mers	$61^{13-1} / 61^{12}$
14	+42,	+41,	+40,	42mers	$61^{14-1} / 61^{13}$
15	+45,	+44,	+43,	45mers	$61^{15-1} / 61^{14}$
16	+48,	+47,	+46,	48mers	$61^{16-1} / 61^{15}$
17	+51,	+50,	+49,	51mers	$61^{17-1} / 61^{16}$
18	+54,	+53,	+52,	54mers	$61^{18-1} / 61^{17}$
19	+57,	+56,	+55,	57mers	$61^{19-1} / 61^{18}$
20	+60,	+59,	+58,	60mers	$61^{20-1} / 61^{19}$
21	+63,	+62,	+61,	63mers	$61^{21-1} / 61^{20}$
22	+66,	+65,	+64,	66mers	$61^{22-1} / 61^{21}$
23	+69,	+68,	+67,	69mers	$61^{23-1} / 61^{22}$
24	+72,	+71,	+70,	72mers	$61^{24-1} / 61^{23}$
25	+75,	+74,	+73,	75mers	$61^{25-1} / 61^{24}$
26	+78,	+77,	+76,	78mers	$61^{26-1} / 61^{25}$
27	+81,	+80,	+79,	81mers	$61^{27-1} / 61^{26}$
28	+84,	+83,	+82,	84mers	$61^{28-1} / 61^{27}$
29	+87,	+86,	+85,	87mers	$61^{29-1} / 61^{28}$
30	+90,	+89,	+88,	90mers	$61^{30-1} / 61^{29}$
31	+93,	+92,	+91,	93mers	$61^{31-1} / 61^{30}$
32	+96,	+95,	+94,	96mers	$61^{32-1} / 61^{31}$
33	+99,	+98,	+97,	99mers	$61^{33-1} / 61^{32}$
34	+102,	+101,	+100,	102mers	$61^{34-1} / 61^{33}$
n	+3n,	+(3n-1),	+(3n-2)	3nmers	$(ASGCOEAA)^{n-1} / (ASGCOEAA)^E$

*The first position of Anti-sense codon at 3'-terminal of Anti-sense ORF sequences is occupied by any one of the Anti-sense Start Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E: Exponent

The n is the serial number of Anti-sense Codon for both 3'-end and 5'-end orientated Anti-sense ORF sequences.

The n is the serial number of Anti-sense Codon begins at both 3'-end and 5'-end orientated Anti-sense ORF sequences.

ASGCOEAA: Anti-sense Genetic Codes Of the 20 Essential Amino Acids / Anti-sense Codons Of the 20 Essential Amino Acids

$(ASGCOEAA)^{n-1} = (ASGCOEAA)^E$ When $ASGCOEAA=61$, $61^{n-1} = 61^E$. $[61.sup.(n-1) = 61.sup.E]$. $E=n-1$

The methods are applicable when $ASGCOEAA=61$, $ASGCOEAA<61$ and $ASGCOEAA>61$

**TABLE 11: A Systematic Sequence Deducing Method for
Anti-sense ORF Sequential Sequences with 5'-terminal Anti-sense Stop Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+1,	+2,	+3,	3mers	$61^{1-1} / 61^0$
2	+4,	+5,	+6,	6mers	$61^{2-1} / 61^1$
3	+7,	+8,	+9,	9mers	$61^{3-1} / 61^2$
4	+10,	+11,	+12,	12mers	$61^{4-1} / 61^3$
5	+13,	+14,	+15,	15mers	$61^{5-1} / 61^4$
6	+16,	+17,	+18,	18mers	$61^{6-1} / 61^5$
7	+19,	+20,	+21,	21mers	$61^{7-1} / 61^6$
8	+22,	+23,	+24,	24mers	$61^{8-1} / 61^7$
9	+25,	+26,	+27,	27mers	$61^{9-1} / 61^8$
10	+28,	+29,	+30,	30mers	$61^{10-1} / 61^9$
11	+31,	+32,	+33,	33mers	$61^{11-1} / 61^{10}$
12	+34,	+35,	+36,	36mers	$61^{12-1} / 61^{11}$
13	+37,	+38,	+39,	39mers	$61^{13-1} / 61^{12}$
14	+40,	+41,	+42,	42mers	$61^{14-1} / 61^{13}$
15	+43,	+44,	+45,	45mers	$61^{15-1} / 61^{14}$
16	+46,	+47,	+48,	48mers	$61^{16-1} / 61^{15}$
17	+49,	+50,	+51,	51mers	$61^{17-1} / 61^{16}$
18	+52,	+53,	+54,	54mers	$61^{18-1} / 61^{17}$
19	+55,	+56,	+57,	57mers	$61^{19-1} / 61^{18}$
20	+58,	+59,	+60,	60mers	$61^{20-1} / 61^{19}$
21	+61,	+62,	+63,	63mers	$61^{21-1} / 61^{20}$
22	+64,	+65,	+66,	66mers	$61^{22-1} / 61^{21}$
23	+67,	+68,	+69,	69mers	$61^{23-1} / 61^{22}$
24	+70,	+71,	+72,	72mers	$61^{24-1} / 61^{23}$
25	+73,	+74,	+75,	75mers	$61^{25-1} / 61^{24}$
26	+76,	+77,	+78,	78mers	$61^{26-1} / 61^{25}$
27	+79,	+80,	+81,	81mers	$61^{27-1} / 61^{26}$
28	+82,	+83,	+84,	84mers	$61^{28-1} / 61^{27}$
29	+85,	+86,	+87,	87mers	$61^{29-1} / 61^{28}$
30	+88,	+89,	+90,	90mers	$61^{30-1} / 61^{29}$
31	+91,	+92,	+93,	93mers	$61^{31-1} / 61^{30}$
32	+94,	+95,	+96,	96mers	$61^{32-1} / 61^{31}$
33	+97,	+98,	+99,	99mers	$61^{33-1} / 61^{32}$
34	+100,	+101,	+102,	102mers	$61^{34-1} / 61^{33}$
n	+(3n-2), +(3n-1), +3n			3nmers	$(ASGCOEAA)^{n-1} / (ASGCOEAA)^E$

* The first position of Anti-sense codon at 5'-terminal of Anti-sense ORF sequences is occupied by any one of the Anti-sense Stop Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E: Exponent

The n is the serial number of Anti-sense Codon for both 3'-end and 5'-end orientated Anti-sense ORF sequences.

The n is the serial number of Anti-sense Codon begins at both 3'-end and 5'-end orientated Anti-sense ORF sequences.

ASGCOEAA: Anti-sense Genetic Codes Of the 20 Essential Amino Acids / Anti-sense Codons Of the 20 Essential Amino Acids

$(ASGCOEAA)^{n-1} = (ASGCOEAA)^E$ When $ASGCOEAA=61$, $61^{n-1} = 61^E$. $[61.\text{sup.}(n-1) = 61.\text{sup.}E]$. $E=n-1$

The methods are applicable when $ASGCOEAA=61$, $ASGCOEAA<61$ and $ASGCOEAA>61$

**TABLE 12: A Systematic Sequence Deducing Method for
5'-UTR Anti-sense Sequential Sequence with 5'-terminal Anti-sense Start Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+3,	+2,	+1,	3mers	$64^{1-1} / 64^0$
2	-1,	-2,	-3,	6mers	$64^{2-1} / 64^1$
3	-4,	-5,	-6,	9mers	$64^{3-1} / 64^2$
4	-7,	-8,	-9,	12mers	$64^{4-1} / 64^3$
5	-10,	-11,	-12,	15mers	$64^{5-1} / 64^4$
6	-13,	-14,	-15,	18mers	$64^{6-1} / 64^5$
7	-16,	-17,	-18,	21mers	$64^{7-1} / 64^6$
8	-19,	-20,	-21,	24mers	$64^{8-1} / 64^7$
9	-22,	-23,	-24,	27mers	$64^{9-1} / 64^8$
10	-25,	-26,	-27,	30mers	$64^{10-1} / 64^9$
11	-28,	-29,	-30,	33mers	$64^{11-1} / 64^{10}$
12	-31,	-32,	-33,	36mers	$64^{12-1} / 64^{11}$
13	-34,	-35,	-36,	39mers	$64^{13-1} / 64^{12}$
14	-37,	-38,	-39,	42mers	$64^{14-1} / 64^{13}$
15	-40,	-41,	-42,	45mers	$64^{15-1} / 64^{14}$
16	-44,	-44,	-45,	48mers	$64^{16-1} / 64^{15}$
17	-46,	-47,	-48,	51mers	$64^{17-1} / 64^{16}$
18	-49,	-50,	-51,	54mers	$64^{18-1} / 64^{17}$
19	-52,	-53,	-54,	57mers	$64^{19-1} / 64^{18}$
20	-55,	-56,	-57,	60mers	$64^{20-1} / 64^{19}$
21	-58,	-59,	-60,	63mers	$64^{21-1} / 64^{20}$
22	-61,	-62,	-63,	66mers	$64^{22-1} / 64^{21}$
23	-64,	-65,	-66,	69mers	$64^{23-1} / 64^{22}$
24	-67,	-68,	-69,	72mers	$64^{24-1} / 64^{23}$
25	-70,	-71,	-72,	75mers	$64^{25-1} / 64^{24}$
26	-73,	-74,	-75,	78mers	$64^{26-1} / 64^{25}$
27	-76,	-77,	-78,	81mers	$64^{27-1} / 64^{26}$
28	-79,	-80,	-81,	84mers	$64^{28-1} / 64^{27}$
29	-82,	-83,	-84,	87mers	$64^{29-1} / 64^{28}$
30	-85,	-86,	-87,	90mers	$64^{30-1} / 64^{29}$
31	-88,	-89,	-90,	93mers	$64^{31-1} / 64^{30}$
32	-91,	-92,	-93,	96mers	$64^{32-1} / 64^{31}$
33	-94,	-95,	-96,	99mers	$64^{33-1} / 64^{32}$
34	-97,	-98,	-99,	102mers	$64^{34-1} / 64^{33}$
When $n = 1$, $(1-n)+1$	$3(1-n)+3, 3(1-n)+2, 3(1-n)+1$			3nmers	$(\text{GCOEAA})_{n-1} / (\text{GCOEAA})^E$
When $n > 1$, $(1-n)$	$3(1-n)+2, 3(1-n)+1, 3(1-n)$			3nmers	$(\text{GCOEAA})_{n-1} / (\text{GCOEAA})^E$

*The first position of Anti-sense codon at 5'-terminal of 5'-UTR Anti-sense sequences is occupied by any one of the Anti-sense Start Codons. The negative sign means position in 5'-UTR.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E : Exponent

The n is the serial number of Anti-sense Codon for 5'-end orientated 5'-UTR Anti-sense sequences.

The n is the serial number of Anti-sense Codon begins 5'-end orientated 5'-UTR Anti-sense sequences.

ASGCOEAA: Anti-sense Genetic Codes Of the 20 Essential Amino Acids / Codons Of the 20 Essential Amino Acids

$(\text{ASGCOEAA})^{n-1} = (\text{ASGCOEAA})^E$ When $\text{ASGCOEAA}=61$, $61^{n-1} = 61^E$. $[61.\text{sup.}(n-1) = 61.\text{sup.}E]$. $E=n-1$

The methods are applicable when $\text{ASGCOEAA}=61$, $\text{ASGCOEAA}<61$ and $\text{ASGCOEAA}>61$

**TABLE 13: A Systematic Sequence Deducing Method for
3'-UTR Anti-sense Sequential Sequence with 3'- terminal Anti-sense Stop Codon Orientation**

Codon Position	Nucleotide Position			Sequence Length (3n)	Total Number
*1	+3,	+2,	+1,	3mers	$64^{1-1} / 64^0$
2	+6,	+5,	+4,	6mers	$64^{2-1} / 64^1$
3	+9,	+8,	+7,	9mers	$64^{3-1} / 64^2$
4	+12,	+11,	+10,	12mers	$64^{4-1} / 64^3$
5	+15,	+14,	+13,	15mers	$64^{5-1} / 64^4$
6	+18,	+17,	+16,	18mers	$64^{6-1} / 64^5$
7	+21,	+20,	+19,	21mers	$64^{7-1} / 64^6$
8	+24,	+23,	+22,	24mers	$64^{8-1} / 64^7$
9	+27,	+26,	+25,	27mers	$64^{9-1} / 64^8$
10	+30,	+29,	+28,	30mers	$64^{10-1} / 64^9$
11	+33,	+32,	+31,	33mers	$64^{11-1} / 64^{10}$
12	+36,	+35,	+34,	36mers	$64^{12-1} / 64^{11}$
13	+39,	+38,	+37,	39mers	$64^{13-1} / 64^{12}$
14	+42,	+41,	+40,	42mers	$64^{14-1} / 64^{13}$
15	+45,	+44,	+43,	45mers	$64^{15-1} / 64^{14}$
16	+48,	+47,	+46,	48mers	$64^{16-1} / 64^{15}$
17	+51,	+50,	+49,	51mers	$64^{17-1} / 64^{16}$
18	+54,	+53,	+52,	54mers	$64^{18-1} / 64^{17}$
19	+57,	+56,	+55,	57mers	$64^{19-1} / 64^{18}$
20	+60,	+59,	+58,	60mers	$64^{20-1} / 64^{19}$
21	+63,	+62,	+61,	63mers	$64^{21-1} / 64^{20}$
22	+66,	+65,	+64,	66mers	$64^{22-1} / 64^{21}$
23	+69,	+68,	+67,	69mers	$64^{23-1} / 64^{22}$
24	+72,	+71,	+70,	72mers	$64^{24-1} / 64^{23}$
25	+75,	+74,	+73,	75mers	$64^{25-1} / 64^{24}$
26	+78,	+77,	+76,	78mers	$64^{26-1} / 64^{25}$
27	+81,	+80,	+79,	81mers	$64^{27-1} / 64^{26}$
28	+84,	+83,	+82,	84mers	$64^{28-1} / 64^{27}$
29	+87,	+86,	+85,	87mers	$64^{29-1} / 64^{28}$
30	+90,	+89,	+88,	90mers	$64^{30-1} / 64^{29}$
31	+93,	+92,	+91,	93mers	$64^{31-1} / 64^{30}$
32	+96,	+95,	+94,	96mers	$64^{32-1} / 64^{31}$
33	+99,	+98,	+97,	99mers	$64^{33-1} / 64^{32}$
34	+102,	+101,	+100,	102mers	$64^{34-1} / 64^{33}$
n	+3n,	+(3n-1),	+(3n-2)	3nmers	$(GCOEAA)^{n-1} / (GCOEAA)^E$

*The first position of Anti-sense Codon at 3'-terminal of 3'-UTR Anti-sense sequences is occupied by any one of the Anti-sense Stop Codons.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$ E: Exponent

The n is the serial number of Anti-sense Codon for 3'-end orientated 3'-UTR Anti-sense sequences.

The n is the serial number of Anti-sense Codon begins at 3'-end orientated 3'-UTR Anti-sense sequences.

ASGCOEAA: Anti-sense Genetic Codes Of the 20 Essential Amino Acids / Anti-sense Codons Of the 20 Essential Amino Acids

$(ASGCOEAA)^{n-1} = (ASGCOEAA)^E$ When $ASGCOEAA=61$, $61^{n-1} = 61^E$. $[61.\text{sup.}(n-1) = 61.\text{sup.}E]$. $E=n-1$

The methods are applicable when $ASGCOEAA=61$, $ASGCOEAA<61$ and $ASGCOEAA>61$

**TABLE 14: A Systematic Sequence Deducing Method for
Amino Acids Sequential Sequences with N-terminal Orientation Peptides**

Amino Acid Position	Peptide Length	Total Number
*1	1 Peptide	$20^{1-1} / 20^0$
2	2 Peptides	$20^{2-1} / 20^1$
3	3 Peptides	$20^{3-1} / 20^2$
4	4 Peptides	$20^{4-1} / 20^3$
5	5 Peptides	$20^{5-1} / 20^4$
6	6 Peptides	$20^{6-1} / 20^5$
7	7 Peptides	$20^{7-1} / 20^6$
8	8 Peptides	$20^{8-1} / 20^7$
9	9 Peptides	$20^{9-1} / 20^8$
10	10 Peptides	$20^{10-1} / 20^9$
11	11 Peptides	$20^{11-1} / 20^{10}$
12	12 Peptides	$20^{12-1} / 20^{11}$
13	13 Peptides	$20^{13-1} / 20^{12}$
14	14 Peptides	$20^{14-1} / 20^{13}$
15	15 Peptides	$20^{15-1} / 20^{14}$
16	16 Peptides	$20^{16-1} / 20^{15}$
17	17 Peptides	$20^{17-1} / 20^{16}$
18	18 Peptides	$20^{18-1} / 20^{17}$
19	19 Peptides	$20^{19-1} / 20^{18}$
20	20 Peptides	$20^{20-1} / 20^{19}$
21	21 Peptides	$20^{21-1} / 20^{20}$
22	22 Peptides	$20^{22-1} / 20^{21}$
23	23 Peptides	$20^{23-1} / 20^{22}$
24	24 Peptides	$20^{24-1} / 20^{23}$
25	25 Peptides	$20^{25-1} / 20^{24}$
26	26 Peptides	$20^{26-1} / 20^{25}$
27	27 Peptides	$20^{27-1} / 20^{26}$
28	28 Peptides	$20^{28-1} / 20^{27}$
29	29 Peptides	$20^{29-1} / 20^{28}$
30	30 Peptides	$20^{30-1} / 20^{29}$
31	31 Peptides	$20^{31-1} / 20^{30}$
32	32 Peptides	$20^{32-1} / 20^{31}$
33	33 Peptides	$20^{33-1} / 20^{32}$
34	34 Peptides	$20^{34-1} / 20^{33}$
n	n Peptides	$(EAA)^{n-1} / (EAA)^E$

* One Peptide: An Amino Acid encoded by the Start Codon only. For Example, Methionine is often the first amino acid of N-terminal in most of the peptides.

The n is from 1 to infinity. The n is an integer. n is not equal zero. $E=n-1$. E: Exponent.

The n is the serial number of amino acid for N-terminal orientated amino acids sequences of peptides, polypeptides and proteins.

The n is the serial number of amino acid begins at N-terminal orientated amino acids sequences of peptides, polypeptides and proteins.

EAA: The 20 Essential Amino Acids

$(EAA)^{n-1} = (EAA)^E$. When $EAA=20$, $20^{n-1} = 20^E$. [$20.\text{sup.}(n-1) = 20.\text{sup.}E$]. $E=n-1$.

The methods are applicable when $EAA=20$, $EAA<20$ and $EAA>20$

TABLE 15: A Systematic Sequence Deducing Method for Amino Acids Sequential Sequence with C-terminal Orientation Peptides

Amino Acid Position	Peptide Length	Total Number
*1	1 Peptide	20^1
2	2 Peptides	20^2
3	3 Peptides	20^3
4	4 Peptides	20^4
5	5 Peptides	20^5
6	6 Peptides	20^6
7	7 Peptides	20^7
8	8 Peptides	20^8
9	9 Peptides	20^9
10	10 Peptides	20^{10}
11	11 Peptides	20^{11}
12	12 Peptides	20^{12}
13	13 Peptides	20^{13}
14	14 Peptides	20^{14}
15	15 Peptides	20^{15}
16	16 Peptides	20^{16}
17	17 Peptides	20^{17}
18	18 Peptides	20^{18}
19	19 Peptides	20^{19}
20	20 Peptides	20^{20}
21	21 Peptides	20^{21}
22	22 Peptides	20^{22}
23	23 Peptides	20^{23}
24	24 Peptides	20^{24}
25	25 Peptides	20^{25}
26	26 Peptides	20^{26}
27	27 Peptides	20^{27}
28	28 Peptides	20^{28}
29	29 Peptides	20^{29}
30	30 Peptides	20^{30}
31	31 Peptides	20^{31}
32	32 Peptides	20^{32}
33	33 Peptides	20^{33}
34	34 Peptides	20^{34}
n	n Peptides	$(EAA)^n / (EAA)^E$

*The first position of amino acids of C-terminal is occupied by any one of the EAA.
The n is from 1 to infinity. The n is an integer. n is not equal zero. E=n. E: Exponent.
The n is the serial number of amino acid for C-terminal orientated amino acids sequences of peptides, polypeptides and proteins.
The n is the serial number of amino acid begins at C-terminal orientated amino acids sequences of peptides, polypeptides and proteins.
EAA: The 20 Essential Amino Acids
 $(EAA)^n = (EAA)^E$. When EAA=20, $20^n = 20^E$. [20.sup.n = 20.sup.E]. E=n.
The methods are applicable when EAA=20, EAA<20 and EAA>20

TABLE 16: A Systematic Comparison of Codon-based and Nucleotide-based Oligonucleotide Libraries

Codon-based Oligonucleotide Library*		Nucleotide-based Oligonucleotide Library**		Ratio
Length	Total Number	Length	Total Number	Nucleotide/Codon
1 Codon	$61^{(1-1)} = 1$	3mers	$4^{3 \times 1} = 64$	64.00
2 Codons	$61^{(2-1)} = 61$	6mers	$4^{3 \times 2} = 4,096$	67.15
3 Codons	$61^{(3-1)} = 3,721$	9mers	$4^{3 \times 3} = 262,144$	70.45
4 Codons	$61^{(4-1)} = 226,981$	12mers	$4^{3 \times 4} = 16,777,216$	73.91
5 Codons	$61^{(5-1)} = 13,845,841$	15mers	$4^{3 \times 5} = 1,073,741,824$	77.55
6 Codons	$61^{(6-1)} = 844,596,301$	18mers	$4^{3 \times 6} = 68,719,476,736$	81.36
7 Codons	$61^{(7-1)} = 51,520,374,361$	21mers	$4^{3 \times 7} = 4,398,046,511,104$	85.37
8 Codons	$61^{(8-1)} = 3,142,742,836,021$	24mers	$4^{3 \times 8} = 281,474,976,710,656$	89.56
9 Codons	$61^{(9-1)} = 191,707,312,997,281$	27mers	$4^{3 \times 9} = 18,014,398,509,481,984$	93.97
10 Codons	$61^{(10-1)} = 11,694,146,092,834,141$	30mers	$4^{3 \times 10} = 1,152,921,504,606,846,976$	98.59
11 Codons	$61^{(11-1)} = 713,342,911,662,882,601$	33mers	$4^{3 \times 11} = 73,786,976,294,838,206,464$	103.44
12 Codons	$61^{(12-1)} = 43,513,917,611,435,838,661$	36mers	$4^{3 \times 12} = 4,722,366,482,869,645,213,696$	108.53
13 Codons	$61^{(13-1)} = 2,654,348,974,297,586,158,321$	39mers	$4^{3 \times 13} = 302,231,454,903,657,293,676,544$	113.86
14 Codons	$61^{(14-1)} = 161,915,287,432,152,755,657,581$	42mers	$4^{3 \times 14} = 19,342,813,113,834,066,795,298,816$	119.46
15 Codons	$61^{(15-1)} = 9,876,832,533,361,318,095,112,441$	45mers	$4^{3 \times 15} = 1,237,940,039,285,380,274,899,124,224$	125.34
16 Codons	$61^{(16-1)} = 602,486,784,535,040,403,801,858,901$	48mers	$4^{3 \times 16} = 79,228,162,514,264,337,593,543,950,336$	131.50
n Codons	$61^{(n-1)} = (4^3 - 3)^{(n-1)}$	3n mers	4^{3n}	$4^{3n} / 61^{(n-1)}$ OR $4^{3n} / (4^3 - 3)^{(n-1)}$
Formulas:	$61^{(n-1)} = (4^3 - 3)^{(n-1)}$	3n mers	4^{3n}	$4^{3n} / 61^{(n-1)}$

*All Possible Combinations of 61 codons, one Start Codon at 5'-end or one Stop Codon at 3'-end
**All Possible Combinations of Four Nucleotides (A.T.G.C) or Four Bases

TABLE 17: A Systematic Classification of Oligonucleotide by GC Content

Content (GC)	2 codons (6mer)	3 codons (9mer)	4 codons (12mer)	5 codons (15mer)	6 codons (18mer)	7 codons (21mer)
0	0%	0%	0%	0%	0%	0%
1	16.67%	11.11%	8.33%	6.67%	5.56%	4.76%
2	33.33%	22.22%	16.67%	13.33%	11.11%	9.52%
3	50.00%	33.33%	25.00%	20.00%	16.67%	14.29%
4	66.67%	44.44%	33.33%	26.67%	22.22%	19.05%
5	83.33%	55.56%	41.67%	33.33%	27.78%	23.81%
6	100%	66.67%	50.00%	40.00%	33.33%	28.57%
7		77.78%	58.33%	46.67%	38.89%	33.33%
8		88.89%	66.67%	53.33%	44.44%	38.10%
9		100%	75.00%	60.00%	50.00%	42.86%
10			83.33%	66.67%	55.56%	47.62%
11			91.67%	73.33%	61.11%	52.38%
12			100%	80.00%	66.67%	57.14%
13				86.67%	72.22%	61.90%
14				93.33%	77.78%	66.67%
15				100%	83.33%	71.43%
16					88.89%	76.19%
17					94.44%	80.95%
18					100%	85.71%
19						90.48%
20						95.24%
21						100%

SEQUENCE LISTING

<110> Chen, Tao
 Li, Jinghan
 Chen, Te-Ming

<120> Systems Biology Approach: High Throughput Screening (HTS)
 Platforms With Multiple Dimensions

<130> 85201-232

<140> 2,470,965
 <141> 2002-12-17

<150> PCT/CA02/01941
 <151> 2002-12-17

<150> US 60/340009
 <151> 2001-12-17

<160> 62

<170> PatentIn version 3.2

<210> 1
 <211> 18
 <212> DNA
 <213> Artificial

<220>
 <223> conservative motif of six amino acids of a zinc finger gene
 family

<400> 1
 cacacaggag aaaagcca 18

<210> 2
 <211> 6
 <212> PRT
 <213> Artificial

<220>
 <223> conservative motif of six amino acids of zinc finger gene family

<400> 2
 His Thr Gly Glu Phe Pro
 1 5

<210> 3
 <211> 12

<212> DNA
 <213> Artificial

<220>
 <223> forward pcr primer

<400> 3
 atggcagcat cg

12

<210> 4
 <211> 12
 <212> DNA
 <213> Artificial

<220>
 <223> sample primer

<220>
 <221> misc_feature
 <222> (10)..(12)
 <223> 5'-nitroindole

<400> 4
 atggcagcan nn

12

<210> 5
 <211> 15
 <212> DNA
 <213> Artificial

<220>
 <223> sample primer

<220>
 <221> misc_feature
 <222> (10)..(12)
 <223> 5'-nitroindole

<400> 5
 atggcagcan nngca

15

<210> 6
 <211> 6
 <212> PRT
 <213> Artificial

<220>
 <223> exemplary antigenic epitope

<400> 6

Glu Phe Cys Met His Trp
1 5

<210> 7

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 7

gaattctgca tgcactgg

18

<210> 8

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 8

gaattttgca tgcactgg

18

<210> 9

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 9

gaattctgta tgcactgg

18

<210> 10

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 10

gaattctgta tgcattgg

18

<210> 11
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 11
gagttctgca tgcactgg

18

<210> 12
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 12
gagttttgca tgcactgg

18

<210> 13
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 13
gagttctgta tgcactgg

18

<210> 14
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 14
gagttctgta tgcattgg

18

<210> 15
<211> 18
<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 15

gaattctgca tgcattgg

18

<210> 16

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 16

gaattttgca tgcattgg

18

<210> 17

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 17

gaattctgta tgcattgg

18

<210> 18

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 18

gaattctgta tgcactgg

18

<210> 19

<211> 18

<212> DNA

<213> Artificial

<220>

<223> exemplary primer

<400> 19
gagttctgca tgcattgg 18

<210> 20
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 20
gagttttgca tgcattgg 18

<210> 21
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 21
gagttctgta tgcattgg 18

<210> 22
<211> 18
<212> DNA
<213> Artificial

<220>
<223> exemplary primer

<400> 22
gagttctgta tgcactgg 18

<210> 23
<211> 18
<212> DNA
<213> Artificial

<220>
<223> consolidated primer

<220>
<221> misc_feature
<222> (3)..(3)
<223> universal base analogue

<220>
 <221> misc_feature
 <222> (6)..(6)
 <223> universal base analogue

<220>
 <221> misc_feature
 <222> (9)..(9)
 <223> universal base analogue

<220>
 <221> misc_feature
 <222> (15)..(15)
 <223> universal base analogue such as 5'-nitroindole-2'-deoxyriboside,
 3'-nitropyrrole, inosine, pypoxanthine, etc.

<400> 23
 gantttntgna tgcantgg 18

<210> 24
 <211> 24
 <212> DNA
 <213> Artificial

<220>
 <223> mouse beta-actin forward primer

<400> 24
 atgggtgggaa tgggtcagaa ggac 24

<210> 25
 <211> 21
 <212> DNA
 <213> Artificial

<220>
 <223> mouse beta-actin reverse primer

<400> 25
 ggtcatcttt tcacggttgg c 21

<210> 26
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> mouse beta-actin forward primer

<400> 26
tgagagggaa atcgtgcgtg 20

<210> 27
<211> 24
<212> DNA
<213> Artificial

<220>
<223> mouse beta-actin backward primer

<400> 27
atctgctgga aggtggacag tgag 24

<210> 28
<211> 24
<212> DNA
<213> Artificial

<220>
<223> mouse beta-actin forward primer

<400> 28
atggtgggaa tgggtcagaa ggac 24

<210> 29
<211> 24
<212> DNA
<213> Artificial

<220>
<223> mouse beta-actin backward primer

<400> 29
atctgctgga aggtggacag tgag 24

<210> 30
<211> 22
<212> DNA
<213> Artificial

<220>
<223> gapd forward primer

<400> 30
aaggtgaagg tcggagtcaa cg 22

<210> 31

<211> 22
<212> DNA
<213> Artificial

<220>
<223> gapd backward primer

<400> 31
tggaagatgg tgatgggatt tc 22

<210> 32
<211> 24
<212> DNA
<213> Artificial

<220>
<223> gapd forward primer

<400> 32
tgccatcact gccacccaga agac 24

<210> 33
<211> 24
<212> DNA
<213> Artificial

<220>
<223> gapd backward primer

<400> 33
atgaggtcca ccaccctggt gctg 24

<210> 34
<211> 22
<212> DNA
<213> Artificial

<220>
<223> gapd forward primer

<400> 34
aaggtgaagg tcggagtcaa cg 22

<210> 35
<211> 24
<212> DNA
<213> Artificial

<220>

<223> gapd backward primer

<400> 35

atgagggtcca ccaccctggt gctg

24

<210> 36

<211> 9

<212> DNA

<213> artificial

<220>

<223> primer

<400> 36

atgggcagca

9

<210> 37

<211> 12

<212> DNA

<213> artificial

<220>

<223> primer

<400> 37

atgggcagcat cg

12

<210> 38

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 38

gaattctgca tgcactgg

18

<210> 39

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 39

gaattttgca tgcactgg

18

<210> 40
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 40
gaattctgta tgcactgg

18

<210> 41
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 41
gaattctgta tgcattgg

18

<210> 42
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 42
gagttctgca tgcactgg

18

<210> 43
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 43
gagttttgca tgcactgg

18

<210> 44
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 44
gagttctgta tgcactgg

18

<210> 45
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 45
gagttctgta tgcattgg

18

<210> 46
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 46
gaattctgca tgcattgg

18

<210> 47
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 47
gaattttgca tgcattgg

18

<210> 48
<211> 18
<212> DNA
<213> artificial

<220>
<223> primer

<400> 48

gaattctgta tgcattgg

18

<210> 49

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 49

gaattctgta tgcactgg

18

<210> 50

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 50

gagttctgca tgcattgg

18

<210> 51

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 51

gagttttgca tgcattgg

18

<210> 52

<211> 18

<212> DNA

<213> artificial

<220>

<223> primer

<400> 52

gagttctgta tgcattgg

18

<210> 53

<211> 18

<212> DNA
 <213> artificial

<220>
 <223> primer

<400> 53
 gagttctgta tgcactgg

18

<210> 54
 <211> 18
 <212> DNA
 <213> artificial

<220>
 <223> consensus of sequences 40-55 (SEQ ID No. 38-53)

<220>
 <221> misc_feature
 <222> (3)..(3)
 <223> universal base analogue such as 5-nitroindoline, 3-nitopyrrole,
 inosine or hypoxanthine

<220>
 <221> misc_feature
 <222> (6)..(6)
 <223> universal base analogue such as 5-nitroindoline, 3-nitopyrrole,
 inosine or hypoxanthine

<220>
 <221> misc_feature
 <222> (9)..(9)
 <223> universal base analogue such as 5-nitroindoline, 3-nitopyrrole,
 inosine or hypoxanthine

<220>
 <221> misc_feature
 <222> (12)..(12)
 <223> universal base analogue such as 5-nitroindoline, 3-nitopyrrole,
 inosine or hypoxanthine

<220>
 <221> misc_feature
 <222> (15)..(15)
 <223> n is a, c, g, or t

<400> 54
 gantttntgna tgcantgg

18

<210> 55

<211> 24
<212> DNA
<213> artificial

<220>
<223> forward primer

<400> 55
atggtgggaa tgggtcagaa ggac

24

<210> 56
<211> 21
<212> DNA
<213> artificial

<220>
<223> backward primer

<400> 56
ggtcacatcttt tcacggttgg c

21

<210> 57
<211> 20
<212> DNA
<213> artificial

<220>
<223> forward primer f2

<400> 57
tgagagggaa atcgtgcgtg

20

<210> 58
<211> 24
<212> DNA
<213> artificial

<220>
<223> backward primer b2

<400> 58
atctgctgga aggtggacag tgag

24

<210> 59
<211> 22
<212> DNA
<213> artificial

<220>

<223> f1 primer for gapd

<400> 59

aaggtgaagg tcggagtcaa cg

22

<210> 60

<211> 22

<212> DNA

<213> artificial

<220>

<223> b1 primer for gapd

<400> 60

tggaagatgg tgatgggatt tc

22

<210> 61

<211> 24

<212> DNA

<213> artificial

<220>

<223> f2 primer for GAPD

<400> 61

tgccatcact gccacccaga agac

24

<210> 62

<211> 24

<212> DNA

<213> artificial

<220>

<223> b2 primer for GAPD

<400> 62

atgaggtcca ccaccctggt gctg

24

CLAIMS

1. A method of preparing a cDNA library comprising the following method steps: a. preparing an oligonucleotide library comprising all possible combinations of 61 codons that encode the 20 essential amino acids and each oligonucleotide has an orientation sequence wherein the orientation sequence is selected from the group consisting of a 5' end start codon, a 3' end stop codon, a 3' end antisense start codon, a 5' end antisense stop codon, a 5' end two codon restriction endonuclease recognition sequence, a 3' end two codon restriction endonuclease recognition sequence, a 5' end antisense two codon restriction endonuclease recognition sequence, and a 3' end antisense two codon restriction endonuclease recognition sequence and wherein the length of each oligonucleotide is a multiple of three based on the codon length wherein the oligonucleotides have a maximum length of 24 nucleotides, b. mixing an mRNA sample containing a plurality of different mRNA molecules with the oligonucleotide library of method step a wherein the orientation sequence orients each oligonucleotide, c. incubating the mixture of method step b with oligo-d(T)s and reagents needed for reverse transcription under conditions suitable for mRNA reverse transcription thereby producing first strands of a cDNA library, d. incubating the first strands of a cDNA library of method step c with reagents needed for cDNA synthesis under conditions suitable for cDNA synthesis thereby producing a double stranded cDNA library, e. amplifying the double stranded cDNA library of method step d via polymerase chain reaction, and f. isolating each cDNA of method step e.

2. The method of claim 1, wherein said oligonucleotides are grouped according to GC content.

3. The method of claim 1, wherein each of said cDNA is immobilized to or on a suitable solid support in a set of each of said cDNA at a specific discrete position to form an array, the said set comprising at least two copies of said cDNA, the said array comprising at least two said sets.