A Computer-Assisted Therapy: New Methodology of Predict Restricting Enzymes to Replace Mutations Gene Caused Cancers

Ayad Ghany Ismaeel

Abstract— there is high requirement for integrated therapy methodology starting with finding DNA template which involved target gene sequence causes diseases, diagnosis and classify mutations at gene and protein yield him, then expected replacement at a mutated gene via predicting a restriction enzymes before going to the laboratory. There are multiple researches around this idea but either not integrated method or focus on biological actions within lab which associated with problems. This paper suggested Computer-Assisted Therapy CAT as new methodology of integrated therapy to replacement mutated gene caused disease by predicting restriction enzymes using bioinformatics tools, software, packages etc. Implement and apply this proposed CAT methodology on TP53 gene (Tumor protein P53) which caused more than fifty% of human cancers, shows an effective cost and friendly therapy methodology of replacement mutated gene comparing with other methods and this therapy methodology is comprehensive in finding target TP53 (tumor protein P53) gene from DNA template, detecting primers, predict a restriction enzymes, etc. Which reduce the problems within biologic lab when implementing the replacement of mutations TP53 gene.

Index Terms— Computer-Assisted Therapy CAT, DNA Template, Target Sequence, Polymerase Chain Reaction PCR, Restriction Enzymes, Tumor Protein P53.

1 INTRODUCTION

Gene therapy may include a treatment of diseases by replacing the damaged gene with another non-damaged and that will need restriction enzymes and that means may able to scan along a DNA looking for a particular sequence of bases (recognition site or sequence) in general from 4-6 base pairs in length within laboratory that will take approximately 3 days. The restriction digestion in the lab takes place overnight and can be kept in the freezer until the next class period when it will be used for gel electrophoresis, the gels may be stained overnight prior to photographing or recording results.

Here need to explain briefly some important terminology related like DNA template - the sample DNA that contains the target sequence (means the gene sequence which needed therapy). DNA polymerase - a type of enzyme, which is synthesizes new strands of DNA complementary to the target sequence. PCR is based on using the ability of DNA polymerase to synthesize new strand of DNA complementary to the offered template strand. Primers - short pieces of single-stranded DNA that are complementary to the target sequence [1].

- *Ayad Ghany Ismaeel is currently pursuing Professor, computer science PhD in department of Information System Engineering, Technical Engineering College- Erbil Polytechnic University (previous FTE- Erbil), Kurdistan-IRAQ. PH-009647703580299. E-mail: dr_a_gh_i@yahoo.com Alternative dr.ayad.ghany.ismaeel@gmail.com*

The replacement mutated gene caused disease based on restriction enzymes in summary can explain the foreign DNA inserted into a plasmid within biological lab following steps exposure this procedure:

1) Use a restriction enzyme to open up the DNA.
2) Restriction site means cut the DNA at a specific place.
3) This act gives a set of double-stranded DNA pieces with single-stranded ends.
4) The ends that jut out are not only "sticky", but they have gaps that can be now be filled with a piece of foreign DNA.
5) The DNA from an outside source to bond with an original fragment, one more enzyme is needed.
6) DNA ligase seals any breaks in the DNA molecule.

Fig. 1 shows a restriction enzyme action as example using EcoRI (GAATTG) overhang enzyme.

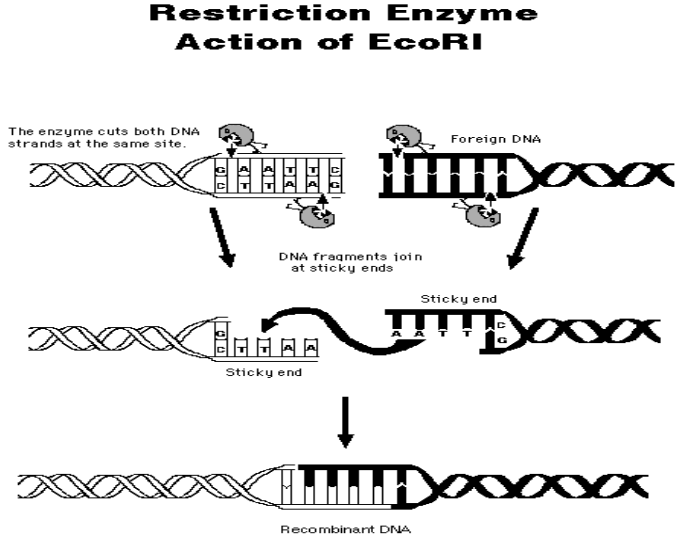

Fig. 1. Restriction enzyme action using EcoRI (GAATTG) overhang enzyme.

Multiple problems happening accompany to biological laboratory if ruled out problems outside of it like find gene, primers, diagnosis and classify mutations, etc. The main of these problems related to laboratory are [2].

1) Few or no transforming that causes restriction enzyme(s) didn't cleave completely.
2) The digested DNA ran as a smear on an agarose gel this causes the restriction enzyme(s) is bound to the substrate DNA.
3) Incomplete restriction enzyme digestion will cause cleavage is blocked by methylation, Inhibition by PCR components, too few units of enzyme used, addition may lead to presence of slow sites, DNA is contaminated with an inhibitor, etc.
4) Extra bands in the gel that may cause if larger bands than expected are seen in the gel, this may indicate binding of the enzyme(s) to the substrate and partial restriction enzyme digest.

That means needed a integrated CAT methodology for replacement a mutated gene (which caused diseases) via predicting of restriction enzymes that done by employing programs, websites and software (supported bioinformatics tools) can produce integrated therapy methodology before going to the laboratory to avoid its problems.

2 Related Work

Colosimo A, Guida V, Antonucci, et al**.** [2007], reveals gene therapy has been proposed as a definitive cure for β-thalassemia applied a gene targeting approach, based on the introduction of small DNA fragments (SDF) into erythroid progenitor cells, to specifically modify the β-globin gene sequence at codon 39. The strategy was first tested in normal individuals by delivering mutant SDF that were able to produce the β39 (C=>T) mutation. Secondly, wild-type SDF were electroporated into target cells of β39/β39. β-thalassemic patients to correct the endogenous mutation in both cases, gene modification was assayed by allele-specific polymerase chain reaction of DNA and mRNA, by restriction fragment length polymorphism analysis and by direct sequencing. Unfortunately, the number of corrected cells remaining from each experiment was insufficient to carry out protein studies by MALDI-TOF analysis. More importantly, improvements in delivery approaches and/or design of SDF are clearly required to yield sufficient quantities of corrected cells for a significant therapeutic benefit. [3].

Owen T. M. Chan, MD, PhD, Kenneth D., et al**.** [2010], suggested methods that assay hemoglobin β-globin chain variants can have limited clinical sensitivity when applied techniques identify only a predefined panel of mutations. Even sequence-based assays may be limited depending on which gene regions are investigated. They sought to develop a clinically practical yet inclusive molecular assay to identify β-globin mutations in multicultural populations. The paper highlights the β-globin mutation detection assay (β-GMDA), an extensive gene sequencing assay. The polymerase chain reaction (PCR) primers are located to encompass virtually all hemoglobin β locus (HBB) mutations. In addition, this assay is able to detect, by gap PCR, a common large deletion (Δ619 base pair), which would be missed by sequencing alone. We describe our 5-year experience with the β-GMDA and indicate its capability for detecting homozygous, heterozygous, and compound heterozygous sequence changes, including previously unknown HBB variants. The β-GMDA offers superior sensitivity and ease of use with comprehensive detection of HBB mutations that result in β-globin chain variants [4].

The drawbacks of these methods implementing the therapy of mutations gene caused diseases done within laboratory and as referring to in section (1) there are multiple problems associated to the lab. The motivation overcome those drawbacks in previous techniques to reach a new CAT methodology, can replace a mutated gene by predict the restriction enzymes based on bioinformatics tools and software then later can go to biological laboratory to implement the plan of replacement mutated gene via CAT methodology after classifying mutations in protein (i.e. not in gene only), because "two sequences may have big differences in DNA sequence but have similar protein" [5, 6].

3 Proposed of CAT Methodology

Start with finding Homology normal gene, it must be represented to the environment by checkup GC% content (satisfying 38% or greater), after that diagnoses and classify there is mutation at that gene causes disease and protein yield him [5, 6, 7], then comes the other tasks. The algorithm of CAT methodology as PCR application shown follow:

Input: DNA template includes sequence of gene target, and sequence of mutated gene for the Person holds disease.
Output: Replacement gene of the Person holds disease via predicting of restriction enzymes.
BEGIN
Step 1: Find DNA sequence as FASTA file (DNA template) using NCBI.
Step 2: Find gene (Homology Normal), i.e. target sequence using certain (software, program, package, etc) and save it as FASTA file.
Step 3: If GC% content >= 38% then Go to step 4: continue
Else Go to step 1: to find another DNA sequence.
Step 4: Diagnosis and Classifying is there mutation at Person's gene and at protein too using ClustalW within BioEdit
Step 5: If there is malignant mutations then:
a) Determine the primers of normal gene via package, program, etc to avoid them when used the restriction enzymes.
b) Predict restriction enzymes using program of Analyze Sequence
c) While (there is malignant mutation) do:
i. Selecting the adjacent enzymes (around the mutilation which needed replacement).
ii. Obtaining the foreign (Homology Normal) gene sequence to replace a mutated gene for Person.
iii. Expected replacement mutated gene using restriction enzymes which predicted in (i) above.
Step 6: Else There isn't risk (no need to replacement gene).
END

4 EXPERIMENTAL RESULTS

A. Programs and Software

Implement the proposed CAT methodology for replacement and applied on mutated TP53 gene (Tumor protein P53) which caused more than 50% types of human cancers shown in the following steps:

1) Fig. 2 shows searching a NCBI for nucleotide (DNA sequence) via the gene name to obtain the DNA template (TP53 gene) then saved as FASTA format.

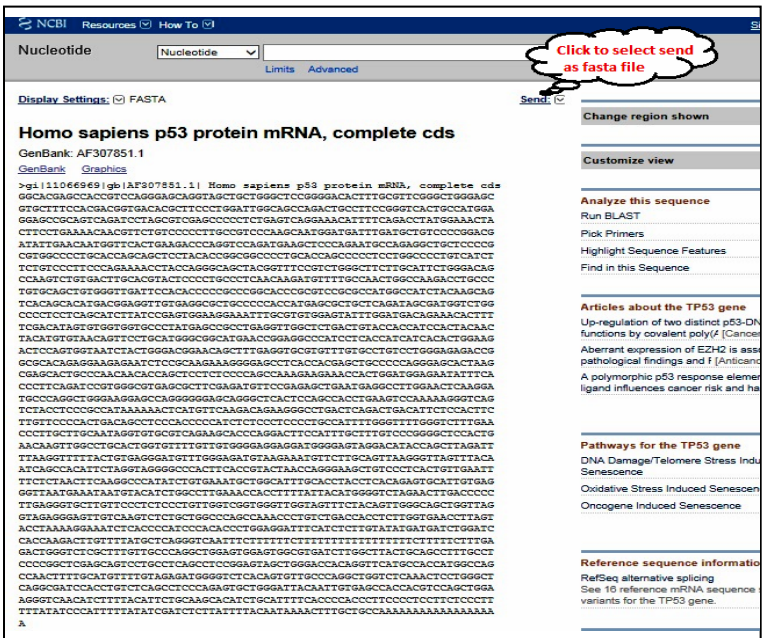

Fig. 2. How the Normal homology TP53 gene can obtain as FASTA format.

2) Find TP53 Gene (Homology Normal Gene) using GENSCAN Web Server at MIT within the gene sequence which obtained at (1) above, will paste as shown in Fig. 3, A; this package can be extracted the normal TP53 gene (target sequence) and save it (in FASTA file) as shown in Fig. 3, B [8].

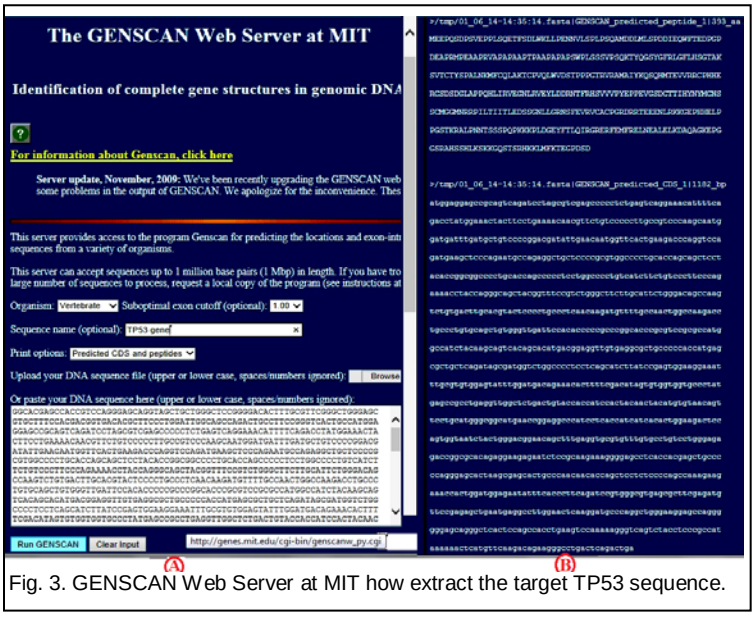

Fig. 3. GENSCAN Web Server at MIT how extract the target TP53 sequence.

3) Using the BioEdit package ver. 7.2.0 to calculate the GC% content for the normal gene of TP53 (FASTA file) which obtained in step (2) by select sequence option→ Nucleic Acid→ Nucleotide Composition, that will give the GC% whether equal or greater than 38%, and find it equal to 54.55% so will continue with step (4), otherwise will return to step (1) to find another TP53 sequence related to environment.

4) The FASTA file which obtained in step (3) will use in ClustalW to display result of alignment, i.e. diagnosis there is malignant mutations by comparing the normal TP53 gene sequence with one sequence (or more than one) TP53 gene sequences for persons at the same time. That is done using BioEdit by selecting Accessory Application→ ClustalW Multiple Alignment→ Run ClustalW, then obtained the result whether there is malignant mutation or not, and in this example will see mutation at codon 349 in TP53 gene for the Person comparing with normal TP53 gene as shown in Fig. 4.

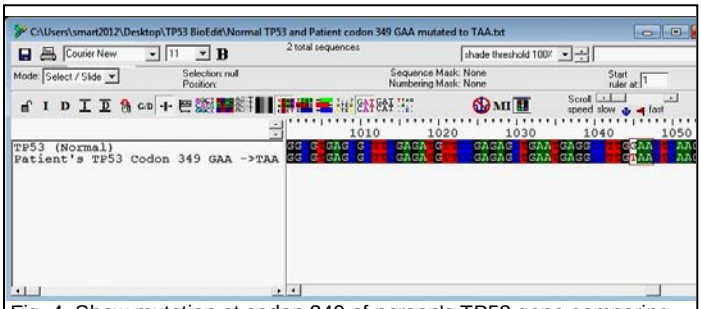

Fig. 4. Show mutation at codon 349 of person's TP53 gene comparing with normal TP53 gene

That is not enough as referring to in section (2), so needed to transform the normal homology TP53 gene to tumor protein P53 as well as TP53 gene for Person to the tumor protein P53 then using the same tool ClustalW at BioEdit package for diagnosis there is malignant mutations (in level of proteins) as done in step (4) or not (means no risk). Fig. 5 shows there is malignant mutations at codon 349 (GAA→ TAA), i.e. will find the codon 349 transformation from (E) to (stop) at Person's P53 gene. The classify of this malignant mutation which discovered may cause lung cancer, head or neck cancers based on the database (UMD_Cell_line_2010) of TP53 website as modern and comprehensive database (URL http://p53.free.fr/Database/p53_MUT_MAT.html [9]).

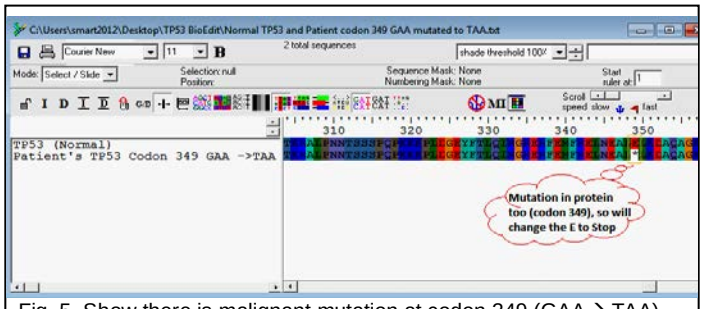

Fig. 5. Show there is malignant mutation at codon 349 (GAA→ TAA).

5) In this step from suggested CAT methodology there is more than stage as follow:
   a) Find the primers at normal TP53 gene (target sequence) that is done using BioEdit by selecting normal TP53 gene→ BLAST NCBI→ then select Primer-BLAST→ paste the FASTA sequence of normal TP53 gene, finally will obtain the primers determines in (10), some of them shown in Fig. 6.

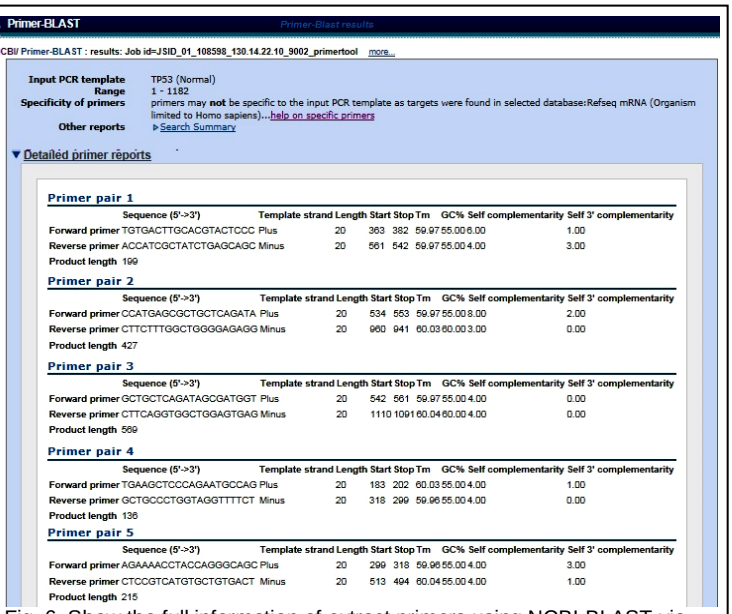

Fig. 6. Show the full information of extract primers using NCBI-BLAST via BioEdit.

   b) Predict restriction enzymes via program of Analyze Sequence to select adjacent enzymes (around each mutation of mutations at TP53 gene), can obtain by open Analyze Sequence → paste the FASTA sequence (target sequence, i.e. normal TP53) at this program → click Submit → click Digest that will reveal all enzymes as shown in Fig. 7 [10].

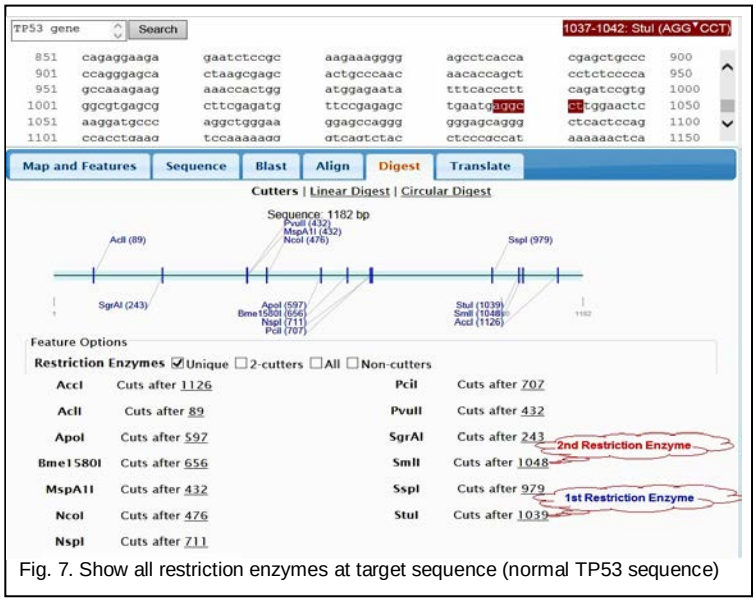

Fig. 7. Show all restriction enzymes at target sequence (normal TP53 sequence)

   c) Multiple steps required to implement continuity while there is mutation in TP53 gene caused diseases as follow:
      i. The program shows the restriction enzymes which are around (adjacent) the mutation at codon 349 in TP53 gene are the Eco147I (StuI) restriction enzyme recognizes AGG^CCT sites (Blund end) and cuts best at 37°C in B buffer (Isoschizomers: AatI, PceI, SseBI, StuI). While the second restriction enzyme recognizes is SmoI (SmlI), where C^TYRAG sites (overhang) and cuts best at 55°C in Tango buffer (Isoschizomers: SmlI) [11].
      ii. In this stage will get the foreign TP53 sequence (AGGCCTTGGAACTCAAG), i.e. that foreign TP53 sequence must be (without mutation) and not primer will use for replacement the mutated TP53 sequence of the Person (AGGCCTTGTAACTCAAG) as shown the mutation at codon 349 (in red), while restriction enzymes (in blue).
      iii. Fig. 8 shows an expected procedure in biologic lab for replacement PT53 gene of Person which has mutation at codon 349 with foreign TP53 sequence (without mutation) before going to the laboratory.

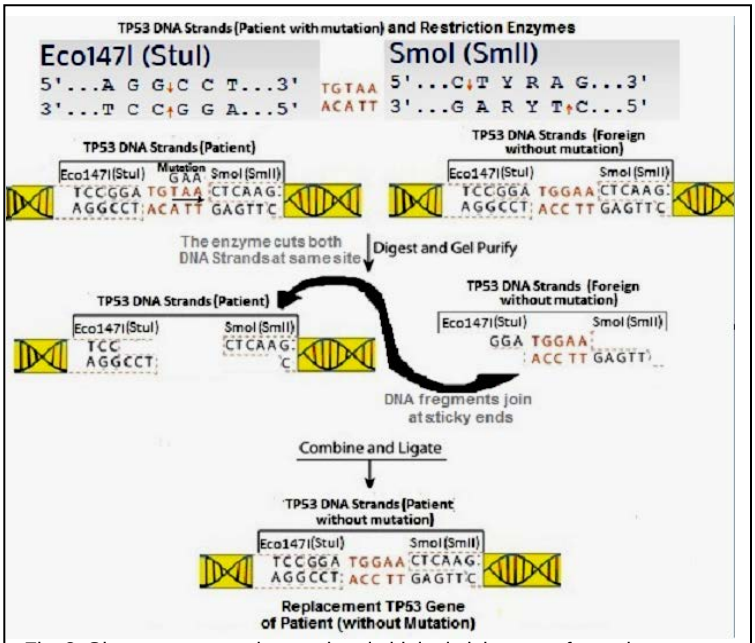

Fig. 8. Show an expected procedure in biologic laboratory for replacement PT53 gene by foreign sequence

B. Discussion of Results

Table 1 shows comparing the results of proposed CAT methodology for replacement mutated gene with other methods.

5 Conclustions And Future Work

The proposed CAT methodology for replacement mutated genes via predicting the restriction enzymes shows the following conclusions:

A. This proposed of CAT will diagnosis the mutations at Person's gene caused disease and protein of that gene based

TABLE 1
REVEALS THE COMPARISON OF PROPOSED CAT METHODOLOGY WITH OTHER TECHNIQUES OR METHODS

Proposed CAT methodology	Method of Colosimo A, et al. [3]	Method of Owen T. M. Chan, et al. [4]
Integrity CAT methodology out the lab starting with finding DNA template, target gene, mutations in gene and its protein, primers; finally replacement mutation at Person's gene has disease (cancer or pre-cancer).	Focus on doing these operations within lab.	Focus on doing these operations within lab.
Predict the restriction enzymes via bioinformatics tools, technique, software, programs, etc.	Not applicable.	Not applicable.
This methodology of CAT friendly, i.e. can use by researcher, molecular biologic, Biomedical, etc.	Limited to molecular biologic.	Limited to molecular biologic.
This CAT methodology will give us full view before going to the lab to avoid problems within lab.	Can't give us this benefit.	Can't give us this benefit.

on collecting datasets (the normal TP53 gene and tumor protein P53) related to environment, i.e. GC% content >= 38% at this case the prediction is more expressive for the region and will reveal impact this results at environment.

B. The proposed CAT methodology offers friendly diagnosis for malignant mutations at gene causes disease and predicting restriction enzymes finally replacement mutated gene for Person has cancer as referring to in Table 1.

C. Can use this CAT methodology by researcher or any other person interested who needed a therapy malignant mutations gene and its' protein caused diseases.

D. That CAT methodology reveals the finding of target gene from DNA template, detecting primers; predict restriction enzymes, etc, i.e. integrated CAT methodology before going to biologic lab will reduce the problems associative to laboratory as referring to in section 1.

As future work needs to create database based on this proposed CAT methodology for all mutations (codons) genes caused diseases as TP53 (tumor protein P53) which caused 50% of human cancers and other important genes like breast cancer genes (BRCA1 and BRCA2), etc. This database will allow mining and flexible retrieving the results related to replace mutated genes using predicting restriction enzymes.

ACKNOWLEDGMENT

Thank to my wife (DR. NEMA SILAH ABDAL KAREEM) that helped me in the maturation of the idea of research and advices at the medical side, which has to do with their competences.

REFERENCES

[1] NCBI, " PCR", URL: http://www.ncbi.nlm.nih.gov/genome/probe/doc/TechPCR.shtml.

[2] New England Biolabs (2014), "Restriction Enzyme Troubleshooting Guide", © Copyright. https://www.neb.com/tools-and-resources/troubleshooting-guides/restriction-enzyme-troubleshooting-guide.

[3] Colosimo A, Guida V, Antonucci, and others (2007), "Sequence-specific modification of α□ -thalassemia locus by small DNA fragments in human erythroid progenitor cells", the hematology journal, Haematologica; 92:129-130.

[4] Owen T. M. Chan, MD, PhD, Kenneth D. Westover, and others (2010), " Comprehensive and Efficient *HBB* Mutation Analysis for Detection of -Hemoglobinopathies in a Pan-Ethnic Population ", © American Society for Clinical Pathology, DOI: 10.1309/AJCP7HQ2KWGHECIO. Pages 700-707.

[5] Ayad Ghany Ismaeel, Anar Auda Ablahad (2013), "Novel Method for Mutational Disease Prediction using Bioinformatics Techniques and Backpropagation Algorithm", IRACST- Engineering Science and Technology: An International Journal Vol. 3, No. 1, (ESTIJ), pages 150-156.

[6] Ayad Ghany Ismaeel, Anar Auda Ablahad (2013), "Enhancement of a Novel Method for Mutational Disease Prediction using Bioinformatics Techniques and Backpropagation Algorithm", INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCH, VOLUME 4, ISSUE 6, JUNE. Pages 1169-1173.

[7] Ayad Ghany Ismaeel (2013), "New Approach for Prediction Pre-cancer via Detecting Mutated in Tumor Protein P53", INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCH, VOLUME 4, ISSUE 10, OCT. Pages 463-468.

[8] GENSCAN Web Server at MIT to find gene under URL: http://genes.mit.edu/GENSCAN.html

[9] France database of TP53 gene (2010), can accessed at URL:http://p53.free.fr/Database/p53_MUT_MAT.html

[10] Analyze Sequence under addgene web server, the URL: http://www.addgene.org/tools/analyze/bd572822b12fc8847c35e013a0f4989d9e9997d3/default/

[11] Thermo Scientific "Molecular Biology", the URL: http://www.thermoscientificbio.com/restriction-enzymes/eco147i-stui/

BIOGRAPHY

Ayad Ghany Ismaeel is professor at 2012, had computer science PhD at 2006 from University of Technology, Baghdad-IRAQ, and M.Sc. in computer science at 1987 from National Computer Center NCC - Ministry of Higher Education and Scientific Research, IRAQ.

Dr. Ismaeel is visiting professor and senior lecturer for undergraduate and postgraduate (doctorate, master, and higher diploma) studies in major topics of computer science, information systems, engineering and software engineering at the universities of Baghdad and Kurdistan region, IRAQ. Professor Ismaeel is editorial board member, program committee member and reviewer for different international journals and conferences, having developed a strong record in research since joining the faculty at 1988. His research interests cover the fields of mobile and IP networks, web application, GPS and GIS techniques, distributed systems, information systems, bioinformatics and bio-computing. He has many publications in international and national journals and conferences, supervising M.Sc. and higher diploma degree projects in software engineering and computer science, also participating in different discussion, examination and scientific PhD and M.Sc. committees.