IJPSR (2012), Vol. 3, Issue 10 (Research Article)

INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES AND RESEARCH

Received on 18 June, 2012; received in revised form 14 July, 2012; accepted 23 September, 2012

IDENTIFICATION OF SPECIFIC MUTATIONS IN HUMAN RAS GENE

Mohammed Qumani Ahmed*, Syed Nazar Hussain , Awdah Masoud Al-hazimi , Mohammad Feroze Fazaludeen, Abdelbaset Mohamed Elasbali and Ibrahim Ginawi

MS Pharmacology, Department of Pharmacology, College of Medicine, University of Hail, Hail, KSA

Keywords:
Bioinformatics tools,
Primer designing,
H-Ras,
K-Ras and N-Ras mutations

Correspondence to Author:

Mohammed Qumani Ahmed

MS Pharmacology, Department of Pharmacology, College of Medicine, University of Hail, Hail, [KSA]

E-mail: clinical.cology@gmail.com

ABSTRACT

Cancer is a group of disease characterized by unregulated cell growth and spread of cells from site of origin to other sites in body. Two main genetic changes lead to cancer they are inactivation of tumour suppressor gene and activation of proto-oncogene. Ras gene is a proto-oncogene, when this gene activated it stimulates signalling pathway and that causes unregulated proliferation of cells. Ras family is a group of three precursors H-Ras, K-Ras and N-Ras. It was analyzed that more than 30% of cancers occurs due to Ras gene. But most of the Ras related cancers are occur due to mutations in H-Ras gene. In H-Ras gene codons 12, 13 and 61 are hotspots. Bioinformatics tools are help to find out the amino acid sequence from which got the gene sequence of H-Ras. Using the H-Ras gene sequence, six primer pairs were designed to amplify the DNA of patients with the potential mutations. Designed primers can be used for the diagnosis of H-Ras related cancers.

INTRODUCTION: Cancer is an ancient condition and the first instance of it was found with the early Egyptians. Despite this ancient lineage, two modern components, longevity and lifestyle, have both had major impacts on both the type and of number of cancers encountered.

Carcinogenesis is the process by which cancer is generated-is a multi-step mechanism resulting from accumulation of errors in vital regulatory pathways.

History of Cancer: Cancer tends to develop when cells grow out of control. There are many types of cancers and they all start due to abnormal cell growth. The normal body cells divide and die. The cells divide more rapidly until the person becomes an adult and after this stage the cells divide only to replace the dying cells or worn-out cells or to repair injuries. Cancer cells continue to grow and divide to form abnormal cell colonies and give birth to the new abnormal cells, which travel to different parts of the body to spread cancer. Cancer may also develop due to DNA damage (www.cancer.org).

Between 3000-1500 BC the Egyptian papyri gave the description of human cancer, which refers to breast cancer, the specimen of human cancer was found in female skull which dates back to the Bronze Age (1900-1600 BC). The first person to recognise the difference between benign and malignant states was Hippocrates. Bone cancer indicates Osteosarcoma found in mummies at the earliest stage. . In 1932 Louis Leakey found the oldest possible cancer that is hominid malignant tumour.

The disease which is known to be rare today has a very long history (http://www.rare-cancer.org/history-of-cancer.html). Cancer cells usually travel from one part of body to another by which they replace the normal tissue.

The process called metastasis is where the cancer cells move to the different parts of the body such as the liver etc. The metastases of mammary cancer are often found in lungs, bone and brain and which the basis of selective metastasis. The expressed cells of breast cancer contain chemokine receptors such as CXR4. It has been recognized that these receptors are expressed in high levels in such tissues like lungs, bone, brain and circulating cancer cells (Rang, 2005). Most cancers form solid tumours, however, some like leukaemia do not (www.cancer.org).

Biology of cancer: Cancer is a group of diseases which is characterised by the unregulated growth signals and by invasion of cells at site of origin or the different sites in the body. Approximately 85% of cancer is found in epithelial cells and classified as carcinomas. Cancer of different origins have different features like ultraviolet radiation can cause skin and cigarette smoke may cause lungs cancer. The carcinogen which is cancer producing agents will not produce tumour immediately (*Pecorino,* 2005).

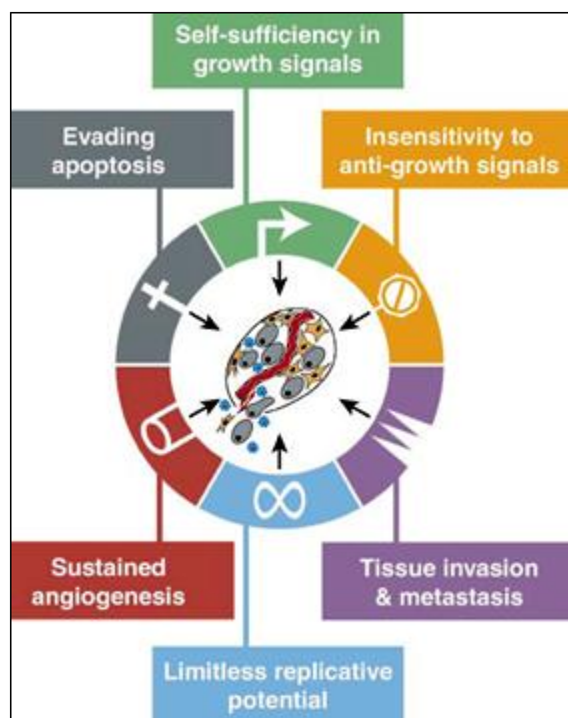

FIG. 1: CAUSES OF CANCER (Hanahan, D. *et al,* 2000)

1. **Tumour suppressor genes:** Tumour suppressor genes break the cell cycle. They are a class of cancer causing genes involved in DNA repair process or mismatch repair. Such types of genes are mutated and acquire more mutation result in tumour suppressor gene and oncogenes leading to increased size of tumour formation. The genes take more time to reveal genes. They use protein which they interact with one another and form cancer (Brighton, 2005). Tumours require angiogenesis to grow in size more than 1-2 mm in diameter. Angiogenesis tumours will strongly correlate with the tumour of metastatic, over-expression of vascular endothelial growth factor with active participation of its receptor which serves to regulate angiogenesis (Oshimoto *et al*, 2006).

   Tumour suppressor genes inhibit cell growth by preventing tumour formation; mutation of such type of a gene will show no longer normal inhibition of cell division and cell growth. The loss of function observed after mutation of the gene which results to inhibit cell growth.

2. **Angiogenesis:** Cancer cell develop growth factor which is responsible for the development or growth of new blood vessels. These new developed blood vessels act as a supplier of oxygen and nutrients to proliferating tumour growth. The blood vessels formed are responsible for the growth of tumour cells that stimulate and proliferation of endothelial cells, the walls of capillaries surrounded by tumour with new capillaries. Formation of these new capillaries is important for tumour growth and shows equal importance in metastasis. Growing capillaries are responsible for angiogenesis for the new formation of tumour, which gives an opening for cancer cells by which the cells can enter into the circulatory system and begin metastatic process resulting in the formation of a tumour (cooper, 2000).

3. **Metastases:** Metastases are principle cause of mortality and morbidity in many cancers and also show problem with cancer therapy. Metastases are formed from primary tumour and reach parts of the body by means of blood vessels or

lymphatic to form secondary tumours. Cancer cell which show metastasise stage tend to display serious genetic changes which inhibit the action of normal regulatory factors and control the tissue site of normal cells. Tumour induces growth factor and shows metastasise because it locally shows a new blood vessel which is more likely. Metastases of mammary cancers are also found in different organs of the body like lung, bone and brain. CXR4 are receptor chemokine expressed by breast cancer. High levels of chemokine are expressed in tissue like lung, bone, brain but not in the kidney (Rang, 2005).

For different cancers the treatments will be applied differently, like surgical removal of cancer tumour is more amenable to the skin than the lungs. Hanahan and Weinberg had defined the hallmark study of cancer in six different parts which are as follows: growth signal autonomy, evasion of growth inhibitors signal, evasion of apoptosis and cell death unlimited reflective potential, angiogenesis, invasion and metastasis which are essential for carcinogenesis. These studies show that future therapies will be designed to target specific pathways depending on the type of cancer treatment required (Pecorino, 2005).

4. **Apoptosis:** Apoptosis is a highly regulated process of cell death which controls the cell numbers and removes damaged cells and therefore plays an important role in tumour suppression. Hence the balance between cell growth, differentiation and apoptosis affects the net number of cells in the body and improper regulation of these results in tumours. Defective apoptosis not only results in carcinogenesis but also influences the effectiveness of conventional therapies that mainly exert their effect by inducing apoptosis. The proteases called caspases plays a central role in apoptosis; it helps to break down cellular components for a neat disposal which is distinguished from necrosis cells, undergoes apoptosis either due to extracellular death factor or by internal physical and chemical results such as DNA damage or oxidation stress. Subsequently two non exclusive molecular pathways, the extrinsic and the intrinsic, may be activated (Pecorino, 2005).

Ras Gene: As protein plays a direct role in human cancer with activating mutation in Ras occurring in approximately 30% of tumours. Ras gene has low molecular weight 21 k Daltons in normal cells. This gene located at chromosome 3p21 (Vander et al, 2007). Ras gene is divided into three types; they are N-Ras, Harvey Ras and Kirsten Ras. It is observed that human tumours are mostly effected with Ras oncogenic mutations, nearly 50% of colorectal cancer and up to90% of pancreatic adenocarcinomas are due to mutation in Ras gene (Midget et al). Ras can be consider as a single gate way to transmits signals to many different pathways, therefore mutation in Ras highly effects on the cell.

Ras have been implicated in the oncogenesis of various tumours and appear to be activated by single point mutations. These point mutations occur in all three Ras genes at codons 12, 13 and 61 with associated amino acid substitution in Ras protein (Michael *et al*, 1990).

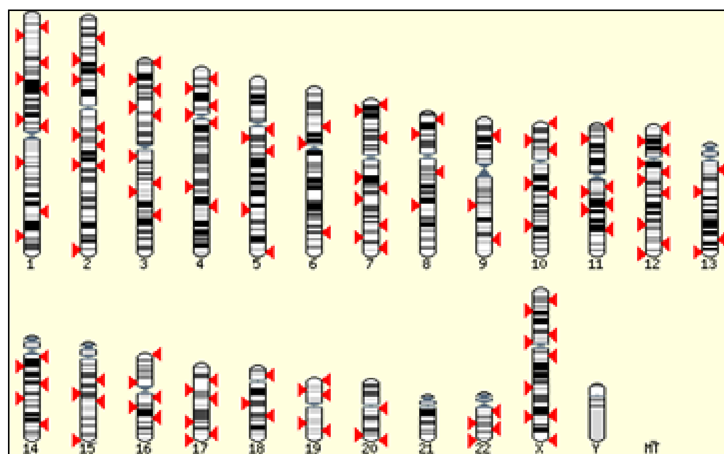

FIG. 2: POSITION OF RAS GTPase IN WHOLE HUMAN GENOME WAS HIGHLIGHTED BY RED ARROW (Source: domain entry taken from ensembl: IPR001806**)**

Nucleic Acid Codes: Sequences are expected to be represented in the standard IUB/IUPAC amino acid and nucleic acid codes, with these exceptions: lower-case letters are accepted and are mapped into upper-case; a single hyphen or dash can be used to represent a gap of indeterminate length; and in amino acid sequences, U and * are acceptable letters (see below). Before submitting a request, any numerical digits in the query sequence should either be removed or replaced by appropriate letter codes (e.g., N for unknown nucleic acid residue or X for unknown amino acid residue).

The nucleic acid codes supported are:

- A --> adenosine M --> A C
- C --> cytidine S --> G C
- G --> guanine W --> A T
- T --> thymidine B --> G T C
- U --> uridine D --> G A T
- R --> G A (purine) H --> A C T
- Y --> T C (pyrimidine) V --> G C A

Amino Acid Codes:

A alanine P proline

B aspartate or asparagine Q glutamine

C cystine R arginine

D aspartate S serine

E glutamate T threonine

F phénylalanine U selenocysteine

G glycine V valine

H histidine W tryptophan

I isoleucine Y tyrosine

K lysine Z glutamate or glutamine

L leucine X any

M methionine * translation stop

N asparagine - gap of intermediate length

Bioinformatics: The application of computer technology to the management and organization of biological information is called as bioinformatics. The potential application of bioinformatics can be applied from single cell to whole ecosystems. Mainly it is used to identify genomic and proteomics activities by involving computers, software tools and databases.

The science of Bioinformatics, which is the melding of molecular biology with computer science, is essential to the use of genomic information in understanding human diseases and in the identification of new molecular targets for drug discovery. In recognition of this, many universities, government institutions and pharmaceutical firms have formed bioinformatics groups, consisting of computational biologists and bioinformatics computer scientists. Such groups will be key to unravelling the mass of information generated by large scale sequencing efforts underway in laboratories around the world.

Bioinformatics is an interdisciplinary research area that is the interface between the biological and computational sciences. The ultimate goal of bioinformatics is to uncover the wealth of biological information hidden in the mass of data and obtain a clearer insight into the fundamental biology of organisms. This new knowledge could have profound impacts on fields as varied as human health, agriculture, the environment and biotechnology. DNA sequence There are three central biological processes around which bioinformatics tools must be;

i) **Determines protein sequence**

ii) **Protein sequence determines protein structure**

iii) **Protein structure determines protein function**

Bioinformatics tools can be used in biological processes like in sequence alignment using a tool called Basic Local Alignment Search Tool (BLAST) (Altschul et al, 1990) and FASTA, the easiest way to use BLAST is through web at the website of National Centre for Biotechnology Information (NCBI) at www.ncbi.nlm.nih.gov/BLAST.

In the similar way protein sequences can also be analyzed, using the protein – sequence database entries and sequence comparison to align proteins at their common regions other sophisticated methods to analyse protein sequences are by design searching using www.expasy.com and CLUSTALW. This can be helpful in prediction of three dimensional structures of proteins; this can be useful in designing primer and probes for PCR and also in constructing phylogenetic tree.

TABLE 1: BLAST AND FASTA PROGRAMS FOR SEQUENCE COMPARISONS.

PROGRAM	COMPARES
FASTA	A nucleotide sequence against a nucleotide sequence database or amino acid sequence against a protein sequence database
TFASTA	Amino acid sequence against a nucleotide sequence database translated in all six reading frames
BALST P	An amino acid sequence against a protein database
BLAST N	A nucleotide sequence against a nucleotide sequence database
BLAST X	A nucleotide sequence translated in all six reading frame against a protein sequence database
EST BLAST	A cDna/EST sequence against Cdna/EST sequence databases
T BLAST N	An amino acid sequence against a nucleotide sequence database translated in all six reading frame

This bioinformatics tools can also be used for primer design some really useful online packages are Gene Fisher, invitrogen, primer 3 which use the parameters which can and be set manually and calculate primer in very less time. These packages can also be used to design multiple primers in PCR which can be used to amplify lots of products in a single reaction process. This is beneficial in molecular diagnosis of clinical disorders. The problems in primer design packages is that they use quality of program that differ in algorithms example a package which is available for free may provide different result than one which is commercial primer due to use of different algorithms (http://bibiserv.techfak.uni-bielefeld.de/genefisher/help/wwwgfdoc.html).

Table 2: Some useful databases and websites used in bioinformatics:

Database	Website
Genome maps	1. (http://compbio.ornl.gov/channel/) 2. (http://www3.ncbi.nilm.nih.gov/entrez/query.fcgi?df=Genome 3. http://www.cephb.fr/bio/ceph-genethon-map.html) 4. (http://lpg.nci.nih.gov/CHLC/)
Nucleotide sequence database	1. (http://lpg.nci.nih.gov/CHLC/) 2. (http://www.ebi.ac.uk/). 3. (http://www.ddbj.nig.ac.jp/fromddbj-e.html).
General databases	1. Gene card, Human gene proteins and disease (http://nciarray.nci.nih.gov/cards/index) 2. OMIM (online mendelian inheritance in man) contains gene and genetic disorders) (http://www.ncbi.nlm.nih.gov/Omim/). 3. Kyoto Encyclopaedia of genes and genomes (KEGG) (http://www.genome.ad.jp/kegg/) 4. Mouse Genome Informatics(MGI) (http://www.informatics.jax.org/)
Protein Database	1. SWISS PROT (http://www.ddbj.nig.ac.jp/fromddbj-e.html). 2. Sanger centre (http://www.sanger.ac.uk/) 3. PFAM (http://genomic.sanger.ac.uk/123D/123D.shtml) 4. PROSITE (http://www.expasy.ch/prosite/)
Signalling pathways	1. G –Protein coupled receptors (http://www.gcrdb.uthscsa.edu). 2. Protein kinase resource (http://www.sdsc.edu/projects/Kinases/pk_home.html) 3. Cell signalling database ((http://geo.nihs.go.jp/csndb/)
Immunologic database	1. Cytokine family cdna database(http://cytokine.medic.kumamoto-u.ac.jp/) 2. Cytokine family database(http://crf.medic.kumamoto-u.ac.jp/) 3. CD guides(http://www.ncbi.nlm.nih.gov/prow/guide/)
Gene expression database	1. (http://www.ncbi.nlm.nih.gov/prow/guide/) 2. National centre for genome resource (NCGR)(http://www.ncgr.org/)

Southern blotting: Blotting method first developed by Ed Southern in mid 1970's to detecting complementary fragments of DNA or RNA sequence by using agarose gel. Southern developed this technique that the reason it is known as Southern blotting (Desmond S.T. Nicholls). In this procedure, agarose gel is mounted on a filter paper wick which dips into a reservoir containing transfer buffer. The hybridisation membrane insert between the gel and a stack of paper towels. The absorbent paper serves to draw the transfer buffer through the gel by capillary action. In this due to the buffer flow the DNA molecules are

carried out of the gel and immobilized on the membrane. For efficient southern blotting, gel pre-treatment is important. Large DNA fragments require a longer time transfer time than short fragments (Primrose S.B, 2001). When the fragments have been transferred from the gel and bound to the filter, it becomes a replica of the gel. A radioactive probe hybridised the filter in a similar way to colony or plaque filters. After hybridisation and washing, the filter is exposed to X-ray film and an autoradiograph prepared, which provides information on the structure of clone (Desmond S.T. Nicholls).

Although Southern blotting is a very simple technique, it has many applications. The technique of Southern blotting is one of the most central methods used in molecular biology. It has been applied to detect Restriction Fragment Length Polymorphism (RFLP), Variable Number of Tandem Repeat Polymorphism (VNTR) is known as finger printing and in molecular markers.

1. **PolymaRase chain reaction (PCR):** Polymerase chain reaction is broadly held as one of the most important inventions of the 20th century in molecular biology. The production of large quantities of a specific DNA from a complex DNA template in a sample enzymatic reaction is called polymerase chain reaction. It is a new technique which amplifies exact piece of DNA fragment with high efficiency (Perl et al, 1990). Kary Mullis invented polymerase chain reaction in 1983, when he was working with CETUS CORPORATION (California), one of the first biotechnology company. For this he is credited with noble prize in chemistry in 1993. This technique is widely adopted because it is simple to change in practical method.

   PCR proceeds with three major steps which are controlled by temperature. The first step is denaturation in which the double stranded template DNA is denatured by heating at 94 C to separate the complementary single strands. The second step is called annealing where the reaction is rapidly cooled to 54°C to allow the oligonucleotide primers to hybridize to the template. During this step the thermo stable DNA polymerase will be active to some extent and start to extend the primers. The last stage of PCR is extension where new DNA complementary strands synthesis, perform at 72°C and this temperature is ideal for polymerase to work. This synthesis of DNA proceed s from both of the primers until the new strands have been extended along and beyond the target DNA to be amplified as this system is taken through successive cycles of denaturation, annealing and extension all new strands act as template and so there will be an exponential increase in amount of DNA produced (Walker *et al*, 2000). After 20 cycles the theoretical yield from each original template molecule is about 10^6 molecules, and about 10^9 molecules after 30 cycles (Colin et al, 2001).

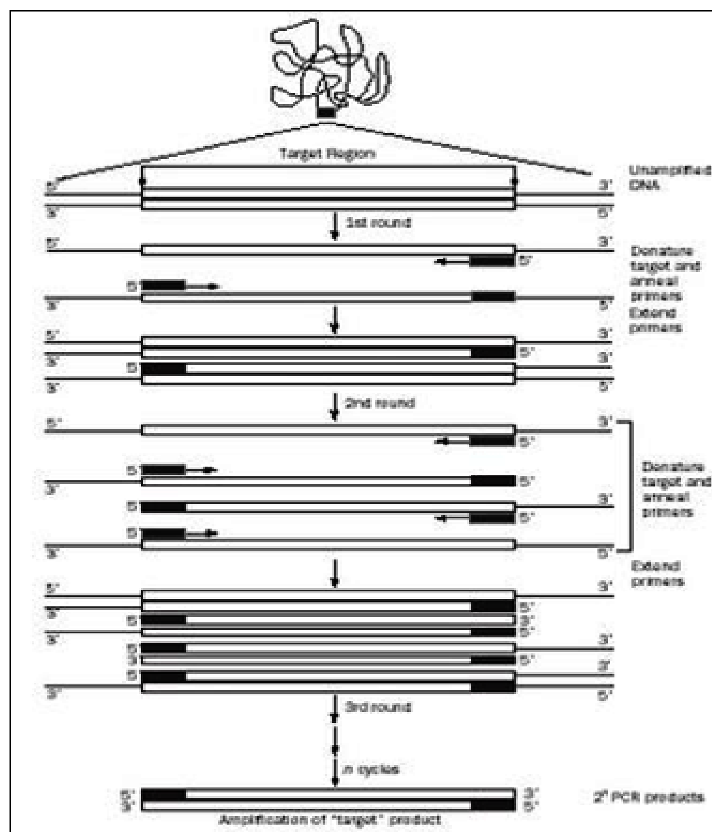

FIG. 3: STAGES IN POLYMERASE CHAIN REACTION
(www.chemistryexplained.com)

2. **Modification of PCR techniques:** The successful use of polymerase chain reaction has driven researchers to develop other amplification techniques similar to PCR these are based on either thermal cycling or non thermal cycling methods. Some of these modified methods are Nested polymerase chain reaction, touchdown PCR, Asymmetric PCR, RT –PCR, quantitative PCR, colony PCR and RACE –PCR.

Reverse transcriptase (RT) PCR: One of the most useful PCR applications is RT-PCR. This method is nothing but a normal PCR technique preceded by transcription by reverse transcriptase. In the method it can be possible to analyse 1kb or more at a time. cDNA is DNA copy synthesized from mRNA. The enzyme reverse transcriptase an RNA-dependant DNA polymerase isolated from retrovirus .this method is used to amplify, isolate or identify a known sequence from cell or tissue the reverse transcriptase synthesizes a complementary DNA on m RNA template. It has many disadvantages like it has poor precision, low sensitivity low resolution there is only size based discrimination only Ethidium Bromide which is used for staining is not quantitative and the process is not automated.

Real time PCR (RT-PCR): Real Time PCR monitors the fluorescence emitted during the reaction as an amplicon production during each of the PCR cycle when compared to conventional quantitative PCR method this quantification is based on the fluorescent reporter .The signal increase is directly proportional amount of PCR product produced. (http://dorakmt.tripod.com/genetics/realtime.html), commonly used fluorescent based DNA detection method are direct labelling of ds-DNA SYBR GREEN 1or hybridization of Taq Man probes .

Asymmetric PCR: It is a PCR in which one strand of DNA is amplified more than the other strand. The product is single stranded DNA; this is because of unequal primer concentration. The primer with lower concentration is incorporated in double stranded DNA and one with higher concentration continues primer synthesis. This mostly used when sequencing hybridization probing one of the complementary strands is required (http://viroligo.okstate.edu /enterdata/PCRs.html).

Q-RT-PCR: It stands for quantitative reverse transcription-PCR. It measures quantity of PCR product thus is an indirect method for measuring start product of PCR.

3. **Applications of PCR:** Polymerase chain reaction is used for broad variety of experiments and as a research and diagnostic tool used for the improvement in human health and to standard of life. PCR gives multiple copies of a specific sequence which helps us to detect presence of viruses, bacteria and other genetically sequenced materials therefore PCR is widely being used in fields of medical research and clinical medicine and in detection of crime by method called genetic finger printing.

Genetic Finger Printing: It is technique which can identify a person by comparing his DNA with a sample DNA, blood of some unknown person which can be collected from a crime spot is compared with the blood, semen, hair or saliva of suspect. DNA sample is broken into single strand and amplified using PCR and run over gel electrophoresis this is called DNA finger print this than compared with the suspects since there is very little possibility that two person have same DNA sequence, this test can also be used for paternity testing by comparing a genetic finger prints of parents and child (Wright et al, 2005).

Detection of hereditary disease: Since PCR can amplify small fragments of DNA, the gene on which variations occurs can be amplified and mutation can be detected, though not many detection kits are available, a detection kit for diagnosis of cystic fibrosis (Friedman et al, 1991) is available in market.

PCR technique can also be used for cloning genes (Smith et al, 1991), mutagenesis and comparison of gene expression.

Detection of infectious disease: PCR can be used to detect viral and bacterial disease (Belak et al, 1993) by making a PCR of the genetic material of the pathogenic organism (DNA, RNA)can be amplified and infectious disease can be diagnosed before occurrence of actual symptoms ,which helps physicians to treat the patient in early stages of the disease.

4. **Limitation of Polymerase Chain Reaction:** Besides its wide application polymerase chain reaction has a few problems associated with it, like polymerase errors, which is due to use of Taq polymerase which lacks a 3’-5’exoneuclease activity which results in error in about 1 to 10,000 bases this can be solved by using other polymerases which have exoneuclease activity examples of such polymerases are KOD DNA polymerase, which is a polymer of Thermococcus Kodakarensis, VENT

which is derived from Thermococcus litoralis, pfu DNA polymerase extracted from pyrococcus furiosus.

Since polymerase chain reaction is enzyme dependant process, enzyme can get exhausted and denatured leading to degradation of enzymes active sites in polymerization Enzymes can also be degraded by presence of inhibitors such as porphyrins and haem which can contaminate and affect the PCR.

Another problem is size limitation, i.e. Polymerase chain reaction works efficiently only in about two thousand base pairs but above that size the enzyme gets degraded and there will be no enough time for polymerisation, this can be solved by slow heating and using high heat resistant enzymes, another potential problem of PCR can be non-specific binding of primer due to sequence duplication and non specific binding this can be prevented by using hot start polymerase during low temperatures of reaction where the active site is blocked by antibody or chemical which comes out on when reaction temperature reaches 95°C (denaturation) another way of preventing is by use of methods like nested PCR and touchdown PCR (Wilson *et al*, 2000).

In PCR the efficiency is just 60%-80%, as shown in figure the by the time PCR reaches 35th cycle the theoretical result is 68 billion copies but the number of copies formed are just 40 -55 billions.

5. **Thermostable DNA polymerases:** Taq polymerase is enzyme extracted from *Thermus Aquaticus* found in hot springs in Yellowstone National park first describe by Brock and freez1969. It is stable at high temperature. Its derivative called stoffel fragment has higher thermal stability and sensitive to magnesium concentration. Higher stability of Taq polymerase makes it useful particularly useful for amplifying GC regions if higher denaturation temperatures are required (Stoffel et al, 1988).

It is important to think about which type of experiment you wish to perform and then to select the suitable enzyme. A short summary of properties and source of some proofreading enzymes is provided (Michael et al, 2006). All these limitation like degradation of dNTPs, enzymes; reactant depletion; inhibition by end products like pyrophosphates and competition for reactants by non specific products lead to effect called 'Plateau Effect' which is shrink in exponential rate of product in PCR process(Gelfand, et al ,1990).

TABLE 3. OTHER THERMOS TABLE POLYMERASES (J.M. WALKER ET AL, 2000).

Thermos table polymerases	Description
Pfu DNA polymerase	Isolated from marine bacterium ***Pyrococcus furiosus***. • It has proof reading ability • It can incorporate radiolabel led nucleotides and analogues helpful in making radio labeled gene probes and performing techniques such as cycle sequencing.
***Thermococcus litoralis* (Vent™)**	***Thermoccous litoralis*** • It is found in deep ocean floors. • Capable of extending templates of more than 12 base pairs. • It has proof reading ability and has high fidelity. • Derivatives have high thermo stability (Deep-vent™, derivative lacking exo-).
***Thermotoga maritime* (UITma™)**	• Derived from the hyperthermophilic gram negative eubacterium ***Thermotoga martima***. • Problem with this is that 3'-5'exoneuclese activity cause primer degradation of primers under initials sub optimal can be overcome by using its exoneuclease forms(-ex o).
Thermus Thermophillus	• It can carry out reverse transcriptase reactions in presence of Manganese which enables use to carryout RNA –PCR.

Primer design: The primer is most important feature of PCR since the results of PCR are dependent on the specificity of primer designed they should be complementary to sequence of target DNA. These primers designed should not be complementary to each other and should not bind to each other and form dimer since it effects the amplification. Therefore following features should be taken into consideration.

Primer sequence-design of 3'end of primer molecule is important for results of PCR, therefore this end should not have 3 or more G or C bases at this position since it leads to non specific annealing of primer and a 3'thymidine nucleotide since it is leads to mis priming of nucleotides.

Primer Length-Primer length should not be less than 18 to 30 bases, it may cause non specific PCR products.

Design of 5'-end of the primer: The 5'-end of the primer should overlap with 5'end of the gene of interest and should have a restriction site (should have same sticky end as the restriction enzyme, 5'extension site often NCOI (CCATGG) are chosen because the ATG in this site can be used to create start codon, 5'extension to restriction site. Restriction enzyme cleave DNA much less efficiently at end of primer therefore a 5'extension of restriction site within 2-10 bases increases efficiency of cleavage, start codon and the overlap between the primer and gene of interest should be long enough.

Design of 3' end primer: The 3'end primer should overlap with the DNA strand complementary to the 3'end of the gene of interest it should have restriction site it should provide the same sticky end to the second of the restriction enzymes, it should have a 5'extension site to restriction site

1. **Melting temperature (Tm) of primer:** The specificity of PCR depends strongly on melting temperature of the primer good results are obtained when melting temperatures if melting temperatures for both primer almost similar (within a range of 2°C to 4°C) and if it is above 60°C. A formula to calculate melting temperature has been developed based on above considerations. GC content- The GC content of the primer should be in between 40 % to 60%. The sequence and length of each primer will decide its Tm (Newton et al, 1994).

$$Tm = 2°C*(A+T) + 4°C*(C+G)$$

2. **Bioinformatics in Primer Design:** Since it is tedious process to design a primer, bioinformatics tools such as Gene runner, Gene fisher and Oligo makes this process straightforward since it takes all the parameter like primer length, product size, GC content etc therefore design of primer by bioinformatics resources makes the initial process of primer selection simpler and lot more time saving.

METHOD:

1. **To obtain Ras gene sequence using expasy: EXPASY** proteomics server(www.expasy.org), the SWISSPROT and TrEMBL database , from which all the protein parameters can be analysed so RAS gene has been done.

   Front page of the server help us to find all information about protein's function, structure, post transcriptional changes and different variants. This is also helpful in finding sequences of the different protein, here SRS is used which is highlighted on the web page to retrieve the protein sequence.

   This is the proteomics database developed by Swiss Institute of Bioinformatics (SIB) with collaboration of European Molecular Biology Laboratory (EMBL). This database especially TrEMBL contains data which developed in dry lab (computational method).

http://expasy.org/

Site Map | Search ExPASy | Contact us

Search Swiss-Prot/TrEMBL for human ras gene Go Clear

ExPASy Proteomics Server

The ExPASy (**Ex**pert **P**rotein **A**nalysis **Sy**stem) proteomics server of the Swiss Institute of Bioinformatics (SIB) is dedicated to the analysis of protein sequences and structures as well as 2-D PAGE (Disclaimer / References).

[Announcements] [Job opening] [Mirror Sites]

Databases

- UniProt Knowledgebase (Swiss-Prot and TrEMBL) - Protein knowledgebase
- PROSITE - Protein families and domains
- SWISS-2DPAGE - Two-dimensional polyacrylamide gel electrophoresis
- ENZYME - Enzyme nomenclature
- SWISS-MODEL Repository - Automatically generated protein models
- Links to many other molecular biology databases

Tools and software packages

- Proteomics and sequence analysis tools
  - Identification and characterization (Aldente, FindMod, Popitam, Phenyx, pI/Mw, ProtParam...)
  - DNA -> Protein
  - Similarity searches (BLAST...)
  - Pattern and profile searches (ScanProsite...)
  - Post-translational modification and topology prediction
  - Primary structure analysis
  - Secondary and tertiary structure tools (Swiss-PdbViewer...)
  - Alignment and Phylogenetic analysis
- ImageMaster / Melanie - Software for 2-D PAGE analysis
- MSight - Mass Spectrometry Imager
- Roche Applied Science's Biochemical Pathways

Education and services

- The ExPASy FTP server
- Swiss-Shop - automatically obtain (by email) new sequence entries relevant to your field(s) of interest
- Vital-IT - The HPC Center for Life Sciences
- e-Proxemis - Bioinformatics Learning Portal for Proteomics
- Master's degree in Proteomics and Bioinformatics
- Proteomics Core Facility (previously SWISS-2DSERVICE) - get your 2-D Gels performed according to Swiss standards

Documentation

- What's New on ExPASy
- SWISS-FLASH electronic bulletins
- Swiss-Prot documents
- How to create HTML links to ExPASy
- Complete table of available documents

Links to lists of molecular biology resources

- ExPASy Life Science Directory - The ExPASy list of biomolecular servers

Links to some major molecular biology servers

- European Bioinformatics Institute (EBI)

FIG. 4: SCREENSHOT OF EXPASY PROTEOMICS SERVER CREATED AND MANAGED BY SWISS INSTITUTE OF BIOINFORMATICS (SIB), PROVIDING VARIOUS LINKS TO MOLECULAR BIOLOGY DATABASES, PROTEOMICS TOOLS AND SOFTWARES AND TO MANY OF BIOINFORMATICS RESOURCES, (http://expasy.org/).

ExPASy Home page | Site Map | Search ExPASy | Contact us | PROSITE | Proteomics tools

Search Swiss-Prot/TrEMBL for human ras gene Go Clear

swissprot

Swiss-Prot
Protein knowledgebase
TrEMBL
Computer-annotated supplement to Swiss-Prot

UniProt
the universal protein resource

The UniProt Knowledgebase consists of:

- **UniProtKB/Swiss-Prot**; a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc.), a minimal level of redundancy and high level of integration with other databases [More details / References / Linking to Swiss-Prot / User manual / Recent changes / Disclaimer].
- **UniProtKB/TrEMBL**; a computer-annotated supplement of Swiss-Prot that contains all the translations of EMBL nucleotide sequence entries not yet integrated in Swiss-Prot.

These databases are developed by the Swiss-Prot groups at SIB and at EBI.

UniProt Knowledgebase Release 12.1 consists of:
UniProtKB/Swiss-Prot Release 54.1 of 21-Aug-2007: 277883 entries (More statistics)
UniProtKB/TrEMBL Release 37.1 of 21-Aug-2007: 4754787 entries (More statistics)

> Swiss-Prot headlines
More than 18'500 phosphorylation sites identified by mass spectrometry in UniProtKB/Swiss-Prot (Read more...)

Access to the UniProt Knowledgebase

- SRS - Access to UniProtKB/Swiss-Prot, UniProtKB/TrEMBL and other databases using the Sequence Retrieval System
- Full text search in the UniProt Knowledgebase
- Advanced search in the UniProt Knowledgebase by description, gene name and organism (can be used to create html links to UniProt Knowledgebase queries)
- Taxonomy browser (NEWT)
- BLAST similarity search
- by description or identification (any word in the DE, OS, OG, GN and ID lines)
- by citation (RL line; UniProtKB/Swiss-Prot only)
- Retrieve a list of UniProtKB entries
- Randomly retrieve a UniProtKB entry
- UniProtKB Sequence/Annotation Version Database new

FIG. 5: SWISS PROT AND TREMBL PAGE (Screenshot of the SWISS-PROT and TrEMBL web page, which was developed to manage the protein's data which was annotated both by wet-lab and dry-lab procedures, i.e., electronically.) http://expasy.org/sprot/

Ras association domain-containing protein 4. {GENE: Name=RASSF4; ORFNames=AD037} - Homo sapiens (Human)
RASF5_HUMAN (Q8WWW0)
Ras association domain-containing family protein 5 (New ras effector 1) (Regulator for cell adhesion and polarization enriched in lymphoid tissues) (RAPL). {GENE: Name=RASSF5; Synonyms=NORE1, RAPL} - Homo sapiens (Human)
RASF6_HUMAN (Q6ZTQ3)
Ras association domain-containing protein 6. {GENE: Name=RASSF6} - Homo sapiens (Human)
RASF7_HUMAN (Q02833)
Ras association domain-containing protein 7 (HRAS1-related cluster protein 1). {GENE: Name=RASSF7; Synonyms=C11orf13, HRC1} - Homo sapiens (Human)
RASF8_HUMAN (Q8NHQ8)
Ras association domain-containing protein 8 (Carcinoma-associated protein HOJ-1). {GENE: Name=RASSF8; Synonyms=C12orf2} - Homo sapiens (Human)
RASH_HUMAN (P01112)
GTPase HRas precursor (Transforming protein p21) (p21ras) (H-Ras-1) (c-H-ras) (Ha-Ras). {GENE: Name=HRAS; Synonyms=HRAS1} - Homo sapiens (Human)
RASK_HUMAN (P01116)
GTPase KRas precursor (K-Ras 2) (Ki-Ras) (c-K-ras) (c-Ki-ras). {GENE: Name=KRAS; Synonyms=KRAS2, RASK2} - Homo sapiens (Human)
RASL1_HUMAN (O95294)
RasGAP-activating-like protein 1. {GENE: Name=RASAL1; Synonyms=RASAL} - Homo sapiens (Human)
RASL2_HUMAN (O43374)
Ras GTPase-activating protein 4 (RasGAP-activating-like protein 2) (Calcium-promoted Ras inactivator). {GENE: Name=RASA4; Synonyms=CAPRI, GAPL, KIAA0538} - Homo sapiens (Human)
RASM_HUMAN (O14807)
Ras-related protein M-Ras precursor (Ras-related protein R-Ras3). {GENE: Name=MRAS; Synonyms=RRAS3} - Homo sapiens (Human)
RASN_HUMAN (P01111)
GTPase NRas precursor (Transforming protein N-Ras). {GENE: Name=NRAS; Synonyms=HRAS1} - Homo sapiens (Human)
RB11A_HUMAN (P62491)
Ras-related protein Rab-11A (Rab-11) (YL8). {GENE: Name=RAB11A; Synonyms=RAB11} - Homo sapiens (Human)
RB11B_HUMAN (Q15907)
Ras-related protein Rab-11B (GTP-binding protein YPT3). {GENE: Name=RAB11B; Synonyms=YPT3} - Homo sapiens (Human)
RB22A_HUMAN (Q9UL26)
Ras-related protein Rab-22A (Rab-22). {GENE: Name=RAB22A; Synonyms=RAB22} - Homo sapiens (Human)
RB27A_HUMAN (P51159)
Ras-related protein Rab-27A (Rab-27) (GTP-binding protein Ram). {GENE: Name=RAB27A; Synonyms=RAB27} - Homo sapiens (Human)
RB27B_HUMAN (O00194)
Ras-related protein Rab-27B (C25KG). {GENE: Name=RAB27B} - Homo sapiens (Human)
RB33A_HUMAN (Q14088)
Ras-related protein Rab-33A (Small GTP-binding protein S10). {GENE: Name=RAB33A; Synonyms=RABS10} - Homo sapiens (Human)
RB33B_HUMAN (Q9H082)
Ras-related protein Rab-33B. {GENE: Name=RAB33B} - Homo sapiens (Human)
RB39A_HUMAN (Q14964)

FIG. 6: SCREEN SHOT OF THE WEB PAGE SHOWING THE RESULT OF THE QUERY AS HUMAN RAS GENE AGAINST SWISS-PROT AND TREMBL DATABASE. (need somemore here) http://www.expasy.org/cgi-bin/sprot-search-de?human%20Ras%20gene

When Ras with accession number P01112 clicked and new web page with all the information regarding the Ras human gene is shown. By viewing this page we can say that Ras gene is sequence of 189 amino acids, it has all the information about the domains and regions in Ras gene. When the chain length of the amino acid is clicked, it shows the sequence of Ras gene has entire information about the whole sequence of all 189 amino acids. The next page also consist of list all the variant sequences of mutant Ras gene at different positions on amino acid sequence of Ras gene, this also shows the change of amino acid. This can be helpful in finding mutation on Ras genes in different type of cancer and calculating the regions on which these mutations occurs.

FIG. 7: SEQUENCE RETRIEVAL SYSTEM IS THE USER FRIENDLY WEB INTERFACE DESIGNED TO SEARCH LIST OF DATABASE BASED ON NECESSITY. Just a click on start button, will allow to choose the database and search can be performed using wildcard pattern to meet the need. (http://expasy.org/srs5/)

SRS (Sequence Retrieval System) provides data according to the need. By the help of wild card pattern search, link to different database to retrieve the data required, likewise RAS gene of Homo sapiens was retrieved from SWISSPORT database , using accession number as P01112.

2. **To obtain homology sequence of HRAS:** BLAST(Basic Local Alignment Search Tool) used to find the similar or identical sequence provided either in FASTA or raw format.

The sequence of amino acid which is obtained earlier from Expasy is used to get base pair sequence with help of BLAST. Firstly on blast home page click on t-blast which is responsible for translated nucleotide vs protein query data base. This can be seen as highlighted in figure below.

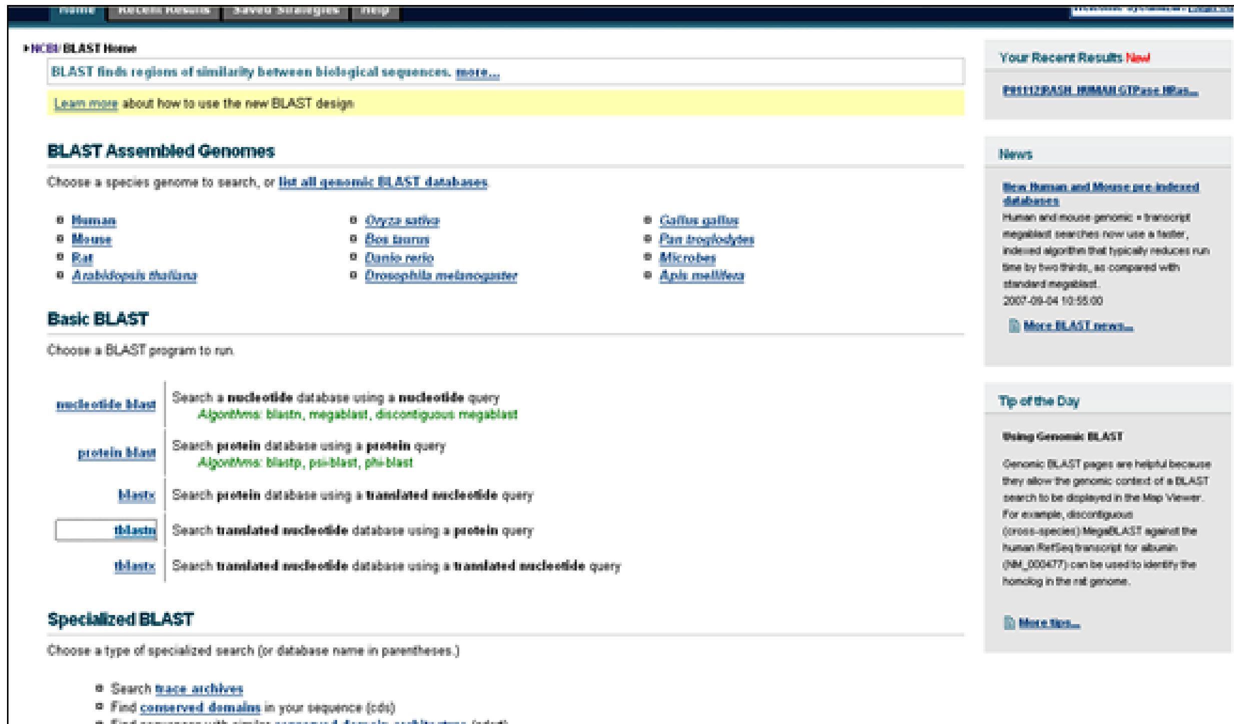

FIG. 8: BLAST HOME PAGE SHOWING TRANSLATED NUCLEOTIDE DATABASE VS PROTEIN QUERY. http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi

When this is done , a query box is obtained in which the amino acid sequence enter, which is from expasy in fasta format is copied and paste into this box and clicked on BLAST which lower left side of the page.

FIG. 9: BLAST PAGE WITH AMINO ACID SEQUENCE TO OBTAIN BASE PAIR

Efficiency of the BLAST search can be improved by selecting their appropriate database.

FIG. 10: THIS PAGE SHOW ALL RELATED HOMO SAPIENS GENE

BLAST search was executed against 'nr' database, in order to find similar protein in other organism. BLAST result showing 105 hits. In order to retrieve the complete data about HRAS gene, HRAS gene "AF493916.1", which shows the complete information like start codon , exons, introns, promoters etc. from which sequence can be downloaded in FASTA format or can copy the FASTA format from the web page.

RESULTS:

Results of Swiss-prot: From Swiss-prot database detailed information about H-Ras gene was found as in the figure 3.1., shows the H-Ras genes accession number P01112 and its synonymous names as transforming protein p21, p21Ras, H-Ras-1, c-H-Ras and Ha-Ras. It also has the information about taxonomy, source and description about amino acids.

FIG. 11: SCREEN SHOT OF THE RESULT PAGE OF P01112 SEARCH IN SWISS-PROT AND TREMBL DATABASE. This page contains all information like, accession number – p01112, taxonomy and studies carried out with regard to the gene. General information of function, related diseases and for structural information links to PDB database and feature table showing the functional aspects of the protein that includes mutations too. This whole page was split into two more figures (figure 3.2 – feature table and figure 3.3 – fasta format of p01112 (HRAS gene) http://www.expasy.org/uniprot/P01112

H Ras gene made up of 189 amino acids and it has molecular weight 21298 Daltons. It also has all the information about variants and mutagens of H Ras gene. It also showing that amino acid 187 to 189 are in immature form.

UniProtKB/Swiss-Prot: RASH_HUMAN (P01112)

GTPase HRas precursor (Transforming protein p21) (p21ras) (H-Ras-1(c-H-ras) (Ha-Ras)
Homo sapiens (Human).

Feature Table:

```
FT   CHAIN        1    186     GTPase HRas.
FT   PROPEP     187    189     Removed in mature form.
FT   NP_BIND     10     17     GTP.
FT   NP_BIND     57     61     GTP.
FT   NP_BIND    116    119     GTP.
FT   REGION     166    185     Hypervariable region.
FT   MOTIF       32     40     Effector region.
FT   MOD_RES    186    186     Cysteine methyl ester.
FT   LIPID      181    181     S-palmitoyl cysteine.
FT   LIPID      184    184     S-palmitoyl cysteine.
FT   LIPID      186    186     S-farnesyl cysteine.
FT   VARIANT     12     12     G -> A (in Costello syndrome).
FT   VARIANT     12     12     G -> S (in Costello syndrome and OSCC).
FT   VARIANT     12     12     G -> V (in Costello syndrome and bladder
FT                             carcinoma; constitutively activated;
FT                             interacts and recruits PLCE1 to plasma
FT                             membrane; loss of interaction with and
FT                             recruitment to plasma membrane of PLCE1
FT                             when associated with F-32; loss of
FT                             interaction with PLCE1 when associated
FT                             with G-26, F-32 and S-35; no effect on
FT                             interaction with PLCE1 when associated
FT                             with A-29, G-34, G-37, N-38 and C-39).
FT   VARIANT     13     13     G -> C (in Costello syndrome).
FT   VARIANT     13     13     G -> D (in Costello syndrome).
FT   VARIANT     61     61     Q -> L (in melanoma).
FT   MUTAGEN     17     17     S->N: Dominant negative. Prevents PLCE1
FT                             EGF-induced recruitment to plasma
FT                             membrane.
FT   MUTAGEN     26     26     N->G: Loss of interaction with PLCE1;
FT                             when associated with V-12.
FT   MUTAGEN     29     29     V->A: No effect on interaction with
FT                             PLCE1; when associated with V-12.
FT   MUTAGEN     32     32     Y->F: Loss of interaction and recruitment
```

FIG. 12: FEATURE TABLE SHOWING THE FUNCTIONS OF THE SPECIFIC RESIDUES LIKE, NUCLEOTIDE BINDING SITES FROM 10 – 17TH , 57 – 61ST AND 116 – 119TH AMINO ACID RESIDUES AND POSSIBLE MUTATIONS WERE ALSO NOTED FROM 12TH RESIDUE, THAT LEADS TO MANY OF THE CARCINOMA'S. http://www.expasy.org/cgi-bin/ft_viewer.pl?P01112

Sequence information

Length: **189 AA** [This is the length of the unprocessed precursor]	Molecular weight: **21298 Da** [This is the MW of the unprocessed precursor]	CRC64: **EE6DC2D933E2856A** [This is a checksum on the sequence]

```
        10         20         30         40         50         60
MTEYKLVVVG AGGVGKSALT IQLIQNHFVD EYDPTIEDSY RKQVVIDGET CLLDILDTAG

        70         80         90        100        110        120
QEEYSAMRDQ YMRTGEGFLC VFAINNTKSF EDIHQYREQI KRVKDSDDVP MVLVGNKCDL

       130        140        150        160        170        180
AARTVESRQA QDLARSYGIP YIETSAKTRQ GVEDAFYTLV REIRQHKLRK LNPPDESGPG

CMSCKCVLS
```

P01112 in FASTA format

FIG. 13: THIS PART OF THE SCREEN SHOT SHOWS THE INFORMATION RELATED TO THE H-RAS AMINO ACID SEQUENCE, SUCH AS LENGTH AS 189 AA AND APPROXIMATE MOLECULAR WEIGHT AS 21298 DA, WHICH CAN BE SAVED AS FASTA FORMAT, FROM THE LINK AVAILABLE WHICH IS HIGHLIGHTED IN ROYAL BLUE COLOR LETTERS. From which the fasta format of the Hras sequence can be later used for further analysis, same format of the sequence can be obtained from NCBI, which connects to many different databases. http://www.expasy.org/uniprot/P01112, while clicking on the P01112 in FASTA format, sequence is retrieved in FASTA format as showing as in fig 3.4.

```
>gi|20147724|gb|AF493916.1| Homo sapiens Ras family small GTP binding protein H-Ras (HRAS) mRNA, complete cds
ATGACGGAATATAAGCTGGTGGTGGTGGGCGCCGGCGGTGTGGGCAAGAGTGCGCTGACCATCCAGCTGA
TCCAGAACCATTTTGTGGACGAATACGACCCCACTATAGAGGATTCCTACCGGAAGCAGGTGGTCATTGA
TGGGGAGACGTGCCTGTTGGACATCCTGGATACCGCCGGCCAGGAGGAGTACAGCGCCATGCGGGACCAG
TACATGCGCACCGGGGAGGGCTTCCTGTGTGTGTTTGCCATCAACAACACCAAGTCTTTTGAGGACATCC
ACCAGTACAGGGAGCAGATCAAACGGGTGAAGGACTCGGATGACGTGCCCATGGTGCTGGTGGGGAACAA
GTGTGACCTGGCTGCACGCACTGTGGAATCTCGGCAGGCTCAGGACCTCGCCCGAAGCTACGGCATCCCC
TACATCGAGACCTCGGCCAAGACCCGGCAGGGAGTGGAGGATGCCTTCTACACGTTGGTGCGTGAGATCC
GGCAGCACAAGCTGCGGAAGCTGAACCCTCCTGATGAGAGTGGCCCCGGCTGCATGAGCTGCAAGTGTGT
GCTCTCCTGA
```

FIG. 14: FASTA FORMAT OF HRAS, WHOSE ACCESSION IS AF493916.1 IN GenBank. Fasta format is same with different database, other than their accession number, which is specific to the concern gene or protein.

3. **Mutations for H-Ras gene:** H-Ras gene shows number of mutations at different sites but three main codons 12, 13 and 61 which are called hot spot for H-Ras gene. Apart from that codon 11, 17, 18, 20, 22, 59, 81, 83 and 117 also shows less number of mutations at these positions. But in different cancer there is a particular type of codon play a major role.

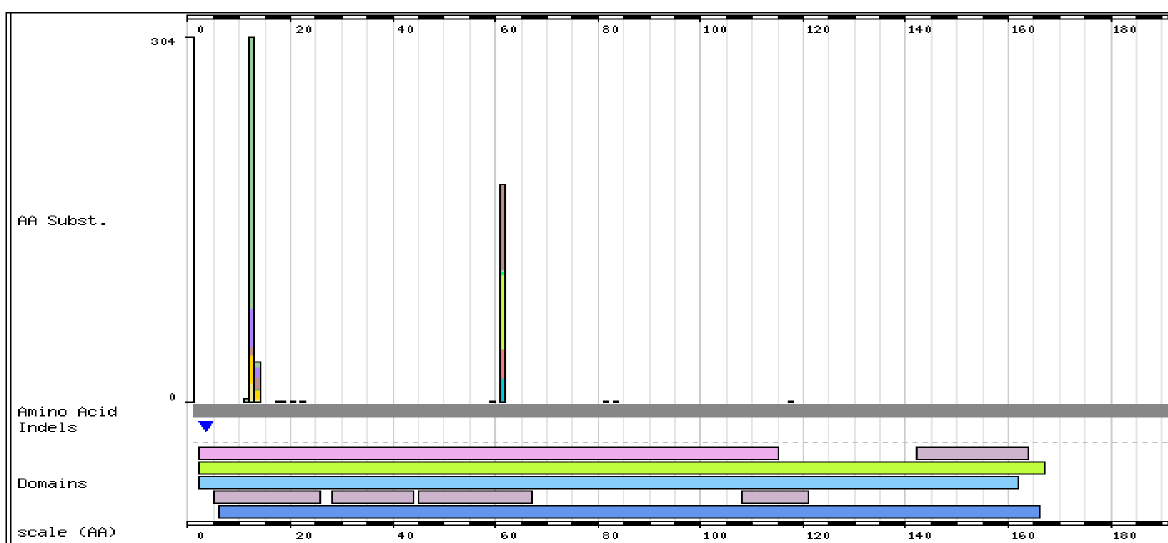

FIG. 15: SOMATIC MUTATION, THAT OCCURS IN HRAS GENE, LEADING TO VARIOUS TYPES OF CANCERS. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=bygene&ln=HRAS&start=1&end=190&coords=AA%3AAA. Figure, taken from Sanger database, shows the hotspot for single base substitution in H-Ras gene.

Mutations have been reported at various positions along the gene. Codons with highest percentage of mutations are 12, 13 and 61 as shown in figure 3.7.

Substitutions	
Position	Mutation(n)
11	p.A11A(2) p.A11S(1)
12	p.G12A(2) p.G12C(14) p.G12D(23) p.G12R(7) p.G12S(32) p.G12V(226)
13	p.G13C(3) p.G13D(7) p.G13G(1) p.G13R(10) p.G13S(8) p.G13V(4)
17	p.S17G(1)
18	p.A18T(1)
20	p.T20I(1)
22	p.Q22*(1)
59	p.A59T(1)
61	p.Q61E(1) p.Q61H(13) p.Q61H(6) p.Q61K(24) p.Q61L(62) p.Q61P(3) p.Q61Q(1) p.Q61R(69) p.Q61R(2)
81	p.V81M(1)
83	p.A83D(1)
117	p.K117E(1)

FIGURE 17: THIS DATA WAS TAKEN FROM SANGER'S INSTITUTE, IN WHICH MUTATIONS IN HUMAN GENEOME WAS CATEGORIZED IN USER – FRIENDLY MANNER TO UNDERSTAND EASILY. Data shows that the major number of mutation occurs at 12th and 13th amino acid residue, leading to many malignant cancers.

4. **Different type of somatic mutations in human Ras gene:**

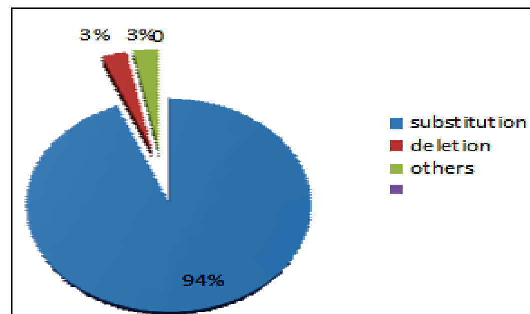

FIG. 18: MAJOR RATIO OF CANCER CAUSED BY SUBSTITUTION AND DELETION MUTATION WAS SHOWN AS PIE CHART OF HRAS GENE, AGAIN, THESE DATA WAS GENERATED FROM THE ANALYSIS OF DIFFERENT CANCERS CAUSED MUTATION FROM THE DATA AVAILABLE FROM THE SANGER'S INSTITUTE DATABASE

The data from Sanger shows that more than 94% of the somatic mutations are single base substitutions. Most of the substitutions are missense. H Ras gene also effected with deletion and others mutations but it is less than the 3%. There is no insertion complex and fusion mutations seen in H-Ras gene.

TABLE 5: TUMOUR SITE DISTRIBUTION OF H-RAS GENE SOMATIC MUTATIONS

Tumor Site	Number of Mutations Recorded	Percentage of Total Mutations
Salivary gland	24	17%
Urinary tract	133	12%
Cervix	23	9%
Upper aero digestive tract	64	8%
Soft tissue	32	7%
Prostate	29	6%
Skin	91	5%
Stomach	14	4%
Thyroid	93	4%
Pituitary	2	2%
Bone	3	2%
Adrenal gland	1	1%
Breast	3	1%
Endometrium	4	1%
Lung	7	1%
Oesophagus	1	1%
TOTAL	527	

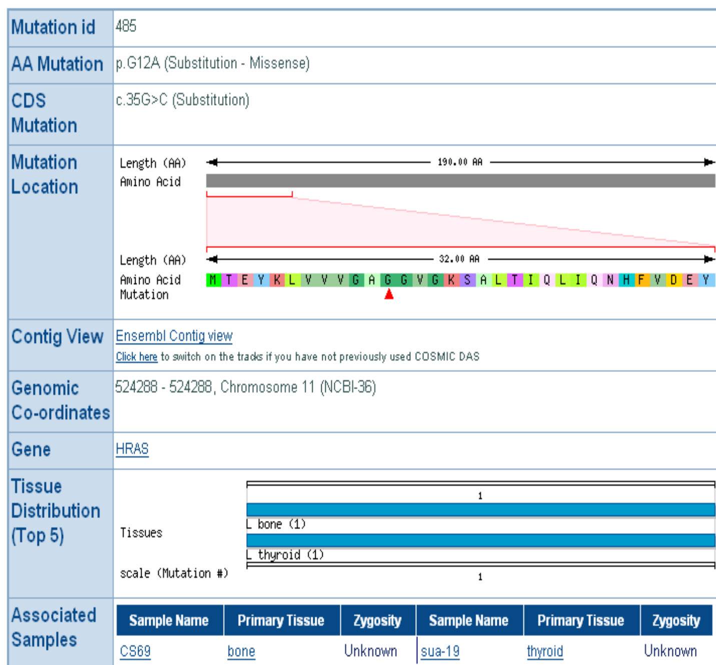

FIG. 19: SUBSTITUTION MUTATION AT POSITION 12 OF THE H-RAS PROTEIN, I.E., G > C, EFFECT OF MISSENSE MUTATIONS WAS NOTED IN PRIMARY TISSUES LIKE BONE AND THYROID. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=mut_summary&id=485

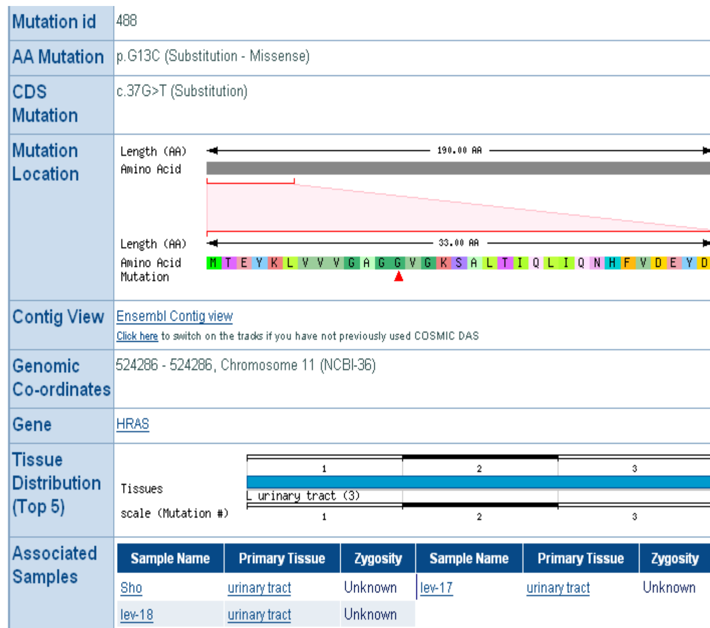

FIG. 20: SUBSTITUTION MUTATION AT POSITION 13 OF THE H-RAS PROTEIN, I.E., G > T, EFFECT OF MISSENSE MUTATIONS WAS NOTED IN URINARY TRACT. SITE FO MUTATION WAS HIGHLIGHTED USING RED ARROW. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=mut_summary&id=488

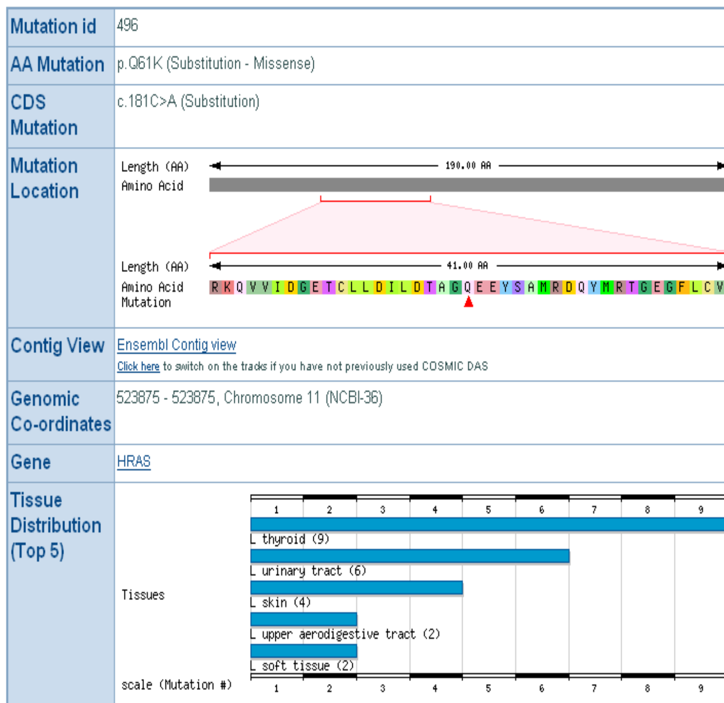

FIG. 21: SUBSTITUTION MUTATION AT 61ST POSITION OF THE H-RAS PROTEIN, I.E., C > A, EFFECT OF MISSENSE MUTATIONS WAS NOTED IN PRIMARY TISSUES LIKE THYROID, URINARY TRACT, SKIN, DIGESTIVE TRACT AND SOFT TISSUES. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=mut_summary&id=497

5. **Important H-Ras cancers:**

a. **Salivary Gland Cancer:** The glands are found in and around mouth and throat, major salivary glands are parotid, submandibular and sublingual glands. Approximately 550 new cases are diagnosed in UK each year, this causing any age group but most common over 50 years. It was found that more than 17% of somatic substitution mutation H-Ras mutations occur at codon 12, where Arginine changes in to Glycine. Salivary glands are the first common position of H-Ras mutations (American cancer society, 2000).

b. **Urinary Tract:** It was founded that over 12% of somatic H-Ras mutations leads to the cancer of urinary tract. At codon 12 the G substitute T which leads to amino acid changes from glycine to valine by which H-Ras oncogene causes urinary tract cancer. It also seen that in some cases codon 13 is also actively involved in urinary cancer.

c. **Cervix:** Cervix is the lower narrow portion of the uterus. In the UK approximately 3000 women are diagnosed with cervical cancer each year. It is found that more than 9% of H-Ras mutation leads to the cancer in cervix. This cancer also affected by point substitution mutations at codon 12.

d. **Upper Aerodigestive Tract:** Patient initially with upper digestive tract cancer have an increased risk of second primary cancer (Han-Kwang et al, 2006). Cancer of upper digestive tract constitutes approximately 4% of all malignancies (Muir C et al, 1995). In that H-Ras substitution mutations are at codon 12 and 13 shown more than 8%.

e. **Soft Tissue:** The term soft tissue refers to tissue that connects, support, or surrounds other structure and organs of the body. All the H-Ras mutated tumours in soft tissue associated with K-Ras mutated genes (Yoo J. et al, 1999). Approximately 7% of H-Ras mutations were found in soft tissue sarcoma. In soft tissue H-Ras mutations occur due to changes at codon 12 from G to T which converted glycine to valine in amino acid sequence.

6. **Primer design:** Following the identification of major sites for H-Ras gene in various type of cancer, appropriate primers are designed using primer3 plus tool.which is capable of amplifying the suspected mutations. In primer design length measured in number of bases and melting temperature (Tm) is measured in °C. The primer from 5' to 3' is called the left or forward primer and the primer from 3' to 5' is known as right or reverse primer. In primer3 plus forward primer highlighted with blue color and reverse primer with yellow color. Primer3 plus also provide information about product size in base pair.

Primers designed to specific H-Ras gene muations:

Primer for Codons 12 and 13: when primer3 plus program was run to get primers for codons 12 and 13 by choosing specified target regoin, here 35 to 40 regoins were selcted in neucliotide sequence. Following primers got as a left and right primers.

Primer3Plus
pick primers from a DNA sequence
Primer3Manager | Help | About | Source Code
< Back

Pair 1:
Left Primer 1: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: CGGAATATAAGCTGGTGGTG
Start: 5 Length: 20 bp Tm: 58.1 °C GC: 50.0 % ANY: 4.0 SELF: 0.0
Right Primer 1: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: ATGGCAAACACACACAGGAA
Start: 251 Length: 20 bp Tm: 60.0 °C GC: 45.0 % ANY: 3.0 SELF: 0.0
Product Size: 247 bp Pair Any: 4.0 Pair End: 1.0
Send to Primer3Manager | Reset Form

```
  1 ATGACGGAAT ATAAGCTGGT GGTGGTGGGC GCCGGCGGTG TGGGCAAGAG
 51 TGCGCTGACC ATCCAGCTGA TCCAGAACCA TTTTGTGGAC GAATACGACC
101 CCACTATAGA GGATTCCTAC CGGAAGCAGG TGGTCATTGA TGGGGAGACG
151 TGCCTGTTGG ACATCCTGGA TACCGCCGGC CAGGAGGAGT ACAGCGCCAT
201 GCGGGACCAG TACATGCGCA CCGGGGAGGG CTTCCTGTGT GTGTTTGCCA
251 TCAACAACAC CAAGTCTTTT GAGGACATCC ACCAGTACAG GGAGCAGATC
301 AAACGGGTGA AGGACTCGGA TGACGTGCCC ATGGTGCTGG TGGGGAACAA
351 GTGTGACCTG GCTGCACGCA CTGTGGAATC TCGGCAGGCT CAGGACCTCG
401 CCCGAAGCTA CGGCATCCCC TACATCGAGA CCTCGGCCAA GACCCGGCAG
451 GGAGTGGAGG ATGCCTTCTA CACGTTGGTG CGTGAGATCC GGCAGCACAA
501 GCTGCGGAAG CTGAACCCTC CTGATGAGAG TGGCCCCGGC TGCATGAGCT
551 GCAAGTGTGT GCTCTCCTGA
```

Select all Primers

Pair 2:
Left Primer 2: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: ACGGAATATAAGCTGGTGGTG
Start: 4 Length: 21 bp Tm: 59.0 °C GC: 47.6 % ANY: 4.0 SELF: 0.0
Right Primer 2: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: ATGGCAAACACACACAGGAA
Start: 251 Length: 20 bp Tm: 60.0 °C GC: 45.0 % ANY: 3.0 SELF: 0.0
Product Size: 248 bp Pair Any: 4.0 Pair End: 1.0
Send to Primer3Manager | Reset Form

Pair 3:
Left Primer 3: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: CGGAATATAAGCTGGTGGTG
Start: 5 Length: 20 bp Tm: 58.1 °C GC: 50.0 % ANY: 4.0 SELF: 0.0
Right Primer 3: gi|47117697:189-758 Homo sapiens v-Ha-ras
Sequence: CGGTATCCAGGATGTCCAAC
Start: 175 Length: 20 bp Tm: 60.2 °C GC: 55.0 % ANY: 7.0 SELF: 2.0
Product Size: 171 bp Pair Any: 4.0 Pair End: 2.0
Send to Primer3Manager | Reset Form

FIG. 22: PRIMER OF CODON 12&13 H-RAS GENE FROM PRIMER3 PLUS.http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

From above result 3 best primers pairs are chosen by taking into consideration length, Tm and GC content of the primers which are best among them.

Pair 1	Length	Tm	GC%
Left primer: CGGAATATAAGCTGGTGGTG	20	58.1	50.0
Right primer: ATGGCAAACACACACAGGAA	20	60.0	45.0
Product size: 247 bp			
Pair 2			
Left primer: ACGGAATATAAGCTGGTGGTG	21	59.0	47.6
Right primer: ATGGCAAACACACACAGGAA	20	60.0	45.0
Product size: 248 bp			
Pair 3			
Left primer: CGGAATATAAGCTGGTGGTG	20	58.1	50.0
Right primer: CGGTATCCAGGATGTCCAAC	20	60.2	55.0
Product size: 171 bp			

These primers which are designed for codon 12&13 are used in different cancer like upper aerodigestive tract, thyroid, urinary tract, skin, salivary gland and bone cancer by potentially preventing amino acid substitution at codon 12 and 13.

Primer for codon 61: To design primers for codon 61 from primer3 plus target region selected as 180 to 185 in human H-Ras gene nucleotide sequence. Following result got from which 3 best right and left primers are chosen depending upon their length Tm and GC content.

FIG 3.13: PRIMER OF CODON 61 H-RAS GENE FROM PRIMER3 PLUS. http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

Pair 1	Length	Tm	GC%
Left primer: GTTGGACATCCTGGATACCG	20	60.2	55.0
Right primer: CCTGCCGAGATTCCACAGT	19	61.2	57.9
Product size: 233 bp			
Pair 2			
Left primer: GTTGGACATCCTGGATACCG	20	60.2	55.0
Right primer: GCCGAGATTCCACAGTGC	18	60.4	61.1
Product size: 230 bp			
Pair 3			
Left primer: GTTGGACATCCTGGATACCG	20	60.2	55.0
Right primer: CTGAGCCTGCCGAGATTC	18	59.6	61.1
Product size: 230 bp			

The above primers which are designed for codon 61 are useful in thyroid, skin, cervix, urinary tract, soft tissue, upper aerodigestive tract and prostate cancers.

DISCUSSION: Cancer is a group of diseases in which cells divide and grow without respect to normal limit and spread to other areas of body. Cancer is said to have affected about one in three people, the world wide incidence is about 10 million cases in 10 years, therefore study and effective cure to this disease is essential. Cancer is a disease that to which a person belonging any age group may be susceptible, but there is more risk to the higher age group. Among the deaths that are happening, Cancer is a disease that claims more than 13% of these deaths and worrying factor is that this percentage is increasing alarmingly each passing year.

Initially cancer treatment is carried out by using heparin in various doses. Can is a primarily a disease of regulation of tissue growth. In order for a normal cell to transform into a cancer cell, genes which regulate cell growth and differentiation must be altered. Genetic changes can occur at many levels, from add or deletion of entire chromosomes to mutation affecting a single DNA nucleotide. The major cause behind the cancer is the abnormality in the genetic material of the cell. Genetic abnormalities found in cancer typically affect two classes of gene oncogene and tumour suppressor genes.

An oncogene is a modified gene or a set of nucleotides that codes for a protein and is believed to cause cancer. Usually, a single oncogene is not enough to turn a normal cell into a cancer cell, and many mutations in a number of different genes may be regulated to make a cell cancerous (Watzinger F *et al*, 1999). A tumour suppressor gene is a gene that reduces the probability that a cell in a multi-cellular organism will turn into a tumour cell. A mutation or deletion of tumour suppressor gene will increase the formation of a tumour. Tumour suppressor genes inhibit cell growth by preventing tumour formation; mutation of such type of a gene will show no longer normal inhibition of cell division and cell growth. The loss of function observed after mutation of the gene which results to inhibit cell growth.

Ras gene is a proto-oncogene. Ras family proto-oncogenes are H-Ras, K-Ras and N-Ras. Ras products are involved in kinase signalling pathway that control the transcription of genes, which regulate cell growth and differentiation. In human cancer Ras protein plays a direct role , Ras gene has low molecular weight 21 k Daltons in normal human cell, this gene located at chromosome 3p21(Vander et al, 2007). To switch ‘on’ the pathway Ras protein must bind to GTP in the cell and to switch ‘off’ pathway the Ras protein break up the GTP molecule. Alteration in the Ras gene can change the Ras protein so that there is no break up and release of the GTP.

Then there is no longer switch 'off', it causes the pathway to get stuck at switch 'on' position. It leads to the cell growth and proliferation, which causes cancer.

Mutant Ras has seen in cancers of many parts of body including pancreas (90%), colon (50%), lung (30%), thyroid (50%), bladder (6%) and ovarian (15%). H-Ras mainly effects colon, thyroid, lung and bladder cancer which are major areas of all among the Ras gene. It means amongst all in the Ras family H-Ras play a major role in causing cancer. The official name of this gene is "v-Ha-ras-Harrey rat sarcoma viral oncogene homolog". The H-Ras gene provides instructions for making a protein that is involved in cell division, through a process known signal transduction.

Use of computer application to solve biological problem is called bioinformatics, bioinformatics is defined as application of computer databases and computational method management of biological information, is essential for almost every aspect of data management in modern biology (Kaminki. N et al, 2000). A common thread in projects in bioinformatics and computational biology is the use of mathematical tools to extract useful information from data produced by high throughput biological techniques such as genome sequencing.

PCR has revolutionized the field of molecular biology. It has enabled researchers to perform experiments easily that previously had been unthinkable with southern blotting. Before PCR was developed, molecular biologist had to use southern blotting, northern blotting and western blotting which are very laborious and time consuming methods to identify, clone and purity DNA sequence. A southern blot is a method routinely used in molecular biology to check for presence of a DNA sequence in a DNA sample. PCR is extremely efficient and sensitive; it can make millions or billions of copies of any specific sequence of DNA, even when the sequence is in a complex mixture. The first step to design primer, the gene of interest usually has to be amplified from genomic or vector DNA by PCR.

The present study is to investigate mutations at various positions in H-Ras gene in different cancers and to design primer for main mutations codon using PRIMER3 PLUS program for diagnosis of those particular cancers. I analyzed that most of the cancers occurs due to mutations at codon 12, 13 and 61 of H-Ras gene and primers were designed for these mutations.

I used Expasy to get back information about the H-Ras gene help to retrieve an amino acid sequence. This program helped to obtain molecular weight, synonyms and FASTA format of gene amino acid sequence. The only drawback of this program is that it presented lot of similar proteins information that takes to find out interest of gene from huge information. But it highlighted the specific sites to get easier to certain extent to identify information about the gene of interest.

BLAST (Basic Local Alignment Search Tool) it is a set of search programs designed for the Windows platform and is used to perform fast similarity searches regardless of whether the query is for protein or DNA. Amino acid sequenced which is obtained in the FASTA format is used to get nucleotide sequence in BLAST program; tblastn compares a protein query sequence against a nucleotide sequence database. Once the result is generated in BLAST program it easier to get the nucleotide sequence in FASTA format.

Sanger is a database to get information about H-Ras mutations and it is easy to use it. I found this program is helpful in finding hotspots, effect of mutations in different cancers types, position in amino acid, mutation type etc. Using this program I found that codon 11, 12, 13,17, 18,20, 22, 59, 61, 81, 83 and 117 are sites where mutation take place in H-Ras gene. But it is clearly described about codon 12, 13, 61 which are hot spots. It has been seen that substitution mutations (94%) are much more when compare to deletion, fusion, complex and other mutations (www.sanger.ac.uk).

Sanger database also present the information about the type of cancers which are effected with H-Ras gene and the detail study about various cancer causing regions , salivary gland 17%, urinary tract 12%, cervix 9%, upper aerodigestive tract 8%, soft tissue7% and skin 5% are mutated with H-Ras gene. Vast majority of H-Ras mutations occur in salivary gland at codon 12 due to changes in amino acid arginine to glycine.

In bladder cancer 18% of tumours are affected with H-Ras substitution mutations, here glycine changes into the valine amino acid due the substitution of G with T (GGC to GTC). In about 9% of cervix cancer mutated H-Ras gene at codon 12 and 13. It has seen that at codon 61 amino acid glutamine changes to glutamate due to substitution of C to G which causes the cervix cancer. Changes in the amino acid in H-Ras from glutamine to histidine at codon 61 due substitution of G to T causes upper digestive tract, skin and prostate cancer.

After analyzing above data decided to design primers for 3 main codons of H-Ras gene which are called hot spots of gene, they are codon 12, 13 and 61. I observed that codon 12 and 13 generating same primers because of the side by side position. Three primers are design for 12&13 and three for codon 61 by using PRIMER3 PLUS tool.

It is tedious process to design a primer, bioinformatics tools such as Gene runner, Gene fisher and Oligo makes this process straightforward since it takes all the parameter like primer length, product size, GC content etc therefore design of primer by bioinformatics resources makes the initial process of primer selection simpler and lot more time saving. Also keep in mind to design primer, primer should be 17-28 base in length , base composition should be 40-60% of G+C and temperature is between 55-65**°.** Besides its wide application polymerase chain reaction has a few problems associated with it, like polymerase errors, which is due to use of Taq polymerase which lacks a 3'-5'exoneuclease activity which results in error in about 1 to 10,000 bases this can be solved by using other polymerases which have exoneuclease activity.

Primer3 plus is an updated task oriented version of primer3. Primer3 plus pick primers PCR reactions by taking into consideration the melting temperature, GC content, length, primer dimer possibilities and product size. This program is user friendly and easy to understand. Primer3 plus home page is containing a box in which nucleotide sequence was paste, it also has many parameters that are changed as per convenience to get appropriate primer. In this, targets are given to design primer for codon 12& 13 is 35, 40 and for codon 61 is 180,185. Targets are given as on sequence which nucleotide mutations are take place.

In primer3 plus when primers are designed each pair of primer shown in a separate box , left primer in light blue colour and right primer in a yellow colour in a very pleasant way.

From that for each codon 3 primer pairs are selected out of 5 from primer3 plus result by analyzing length, GC percentage and Tm value of the primer. Apart from that one primer is chosen as a excellent for 12 & 13 codon, left primer sequence is 'CGGAATATAAGCTGGTGGTG' and right primer sequence is

'ATGGCAAACACACACAGGAA'. The product size is 247 base pairs and temperature different is 2°C (58.1°C for left primer and 60.0°C for right primer), GC content is 50.0%, 45.0% and position of left primer is at 5 and right is 251. Left and right primers are visibly highlighted with purple and yellow colour in primer3 plus.

Most of the H-Ras mutations were found at codon 61, the primers which are designed from primer3 plus are selected and analyzed to find out 3 best primers. From three primers one excellent primer left sequence 20 base pair is 'GTTGGACATCCTGGATACCG' and the right with 19 base pair sequence is 'CCTGCCGAGATTCCACAGT'. The product size is 233 base pairs, temperature difference is within 2°C (left primer 60.2°C and right primer is 61.2°C), position of left primer is 156 and right is 388 and the GC content is 55.0% for left 57.9 for right primer.

H-Ras gene mutations are seen in vast majority in different cancers mainly in the thyroid, cervix, upper digestive, prostate, skin, urinary tract, salivary gland etc. In HRAS most common mutation replaces the amino acid glycine with the amino acid serine at position 12(Gly 12 Ser). Somatic mutation in H-RAS gene in bladder cells has been associated with bladder cancer.

Mutations in the HRAS gene also have been associated with the progression of bladder cancer and an increased risk of tumour recurrence after treatment. Somatic mutations in the HRAS gene are probably involved in the development of several other types of cancer.

These mutations lead to an HRAS protein that is always active and can direct cell to grow and divide without control. Recent studies suggest that HRAS mutation may be common in thyroid and kidney cancers. The HRAS protein may produced at high levels in other types of cancer cells. So the primers which are designed in this project might be helpful in diagnostic test of cancers, but PCR is not carry out and proved that these primers can be used effectively until they are not valid primers.

Future Aspects: In cancer diagnosis biopsies and MRI scan are widely being used but it is requires high skill and has fewer chance to get accurate results because cancer tumours are surrounded with vast majority of normal cells. Providing more suitable tests than the biopsies would greatly increase the sensitivity of detection of mutations in cancer patient cells. A present PCR is not the most popular method being used to detect cancer. The hope is that in future primers are designed to improve the sensitivity of diagnosis test which would establish much more effective than the standard procedure. Future studies are required to understand other Ras family precursors and designing suitable primers by updated bioinformatics tools. Hopefully, in the near future with increased knowledge and understanding it will help to prevent and treatment of cancer.

REFERENCES:

1. Altschul, S.F. Gish, W. Myers, E.W. Lipman, D.J. 1990. Basic Local Allignment Tools. J.M. Biol. 215, 403-410.
2. Belak, S. Ballagi, A. 1993. Application of polymerase chain reaction in veterinary diagnostic virology. Veterinary Research communication. Vol 17, 55-72.
3. Brighton, D., Wood, M., Johnston, S. R.D., Ford, H., Ross, P.J. 2005. The Royal Marsden Hospital Handbook Of Cancer Chemothereapy. Edinburgh: Elsevier Churchill Livingstone, 03-14.
4. Colin ratledge and Bjorn Kristiansen , 2001. ”Basic Biotechnology” second edition, Cambridge university press.
5. Cooper, G.M. 2000. The Cell a Molecular Approach. 2nd. ed. ASM press.
6. Desmond ,S.Nicholl,T 2002. An Introduction to Genetic Engineering , second edition,Cambridge university press.
7. Fried man, K.J. Highsmith Jr, W.E. Silverman, L.M. 1991. Detecting Multiple Cystic fibrosis by Polymerase Chain Reaction mediated site –directed mutagenesis. Clinical chemistry. Vol 37, 753-755.
8. Gelfand DH and White T J(1990) “Thermostable DNA polymerases .” 129-141 in: PCR Protocols (Innis, Gelfand, Sinsky and White, Eds); Academic Press New York.
9. Hanahan, D and Weinberg, R.A. 2000. The Hallmarks of Cancer. Cell. 100 :57-70.
10. Han-Kwang. 2006. International Journal of Cancer.
11. Innis, M.A. Gelfand, D.H. Sninsky, J.J. White, T.J. 1990. PCR Protocols. A Guide to Methods and Applications.
12. Kaminki, N. (2000)‘Bioinformatics –A Users Perspective’ cell mol bio vol 23 705 -711.
13. McPherson, M & Moller, S. 2006. PCR Second Edition.
14. Muir, C. Weiland, L. 1995. Information and Statistics Division. Scottish Health Service, Edinburgh.
15. Newton, C.R. Graham, A. 1994. PCR Introduction to Bioinformatics.
16. Oshimoto, Y.M., Kinuya, S., Kawashima, A., Nishii, R., Yokoyama, K., Kawai, K., 2006. Radiaoiodinated VEGF to image tumor angiogenesis in a LS180 tumor xenograft model. Nuclear Medicine and Biology 33 (2006). 963–969.
17. Pecorino, L. 2005. Molecular Biology of Cancer. Oxford University press. New York.
18. Perl, A. Feher, J. 1990. Polymerase chain reaction in clinical diagnosis. 131(13). 671-5.
19. Rang. H.P and Dale M.M 2003. Pharmacology 5th. Ed. Churchill Livingstone, Edinburg: 693-707.
20. Satoskar, S.R. Bhandarkar, S.D. Ainapure, S.S. Satoskar, R.R. 1999. Pharmacology and pharmacotherapeutics. Sixteenth ed. Popular Prakashan. Mumbai.
21. Vander, L. Wayden, D.J. Adams. 2007. Biochemica et Biophysica Acta.
22. Watzinger , F. Lion,T.1999. RAS family , Atlas Genet Cytogenet Oncol Haematol.
23. Wright, A. Wynford, D.T. 2005. Polymerase Chain Reaction. Miracle or Mirage. *The Pathology Journal*. Vol 162(2), 99-117.
24. Wilson, K. and Walker, J. 2005. Principles and techniques of biochemistry and molecular biology. 6th, ed. Cambridge university press.
25. Yoo, J. Robinson, R.A. Lee, J.Y. 1999. Department of Pathology. Catholic University. St. Vincent Hospital, Suwon, South Korea.
26. www.cancer.org
27. http://www.rare-cancer.org/history-of-cancer.html
28. http://bibiserv.techfak.uni-bielefeld.de/genefisher/help/wwwgfdoc.html
29. http://dorakmt.tripod.com/genetics/realtime.html
30. http://viroligo.okstate.edu/enterdata/PCRs.html
31. www.expasy.org
32. http://www.answers.com/topic/polymerase-chain-reaction?cat=health
33. http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi
34. http://www.sanger.ac.uk/
35. http://AtlasGeneticsOncology.org/Deep/Ras.html
36. http://sgc.utoronto.ca/SGC-WebPages/StructureDescription/img/2FV8_400x400.png
37. http://cache.eb.com/eb/image?id=22477&rendTypeId=4
38. http://expasy.org/sprot/
39. http://www.expasy.org/uniprot/P01112
40. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=bygene&ln=HRAS&start=1&end=190&coords=AA%3AAA
41. http://www.sanger.ac.uk/perl/genetics/CGP/cosmic?action=mut_summary&id
42. http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

How to cite this article:
Ahmed MQ, Hussain SN, Al-hazimi AM, Fazaludeen MF and Elasbali AM: Identification of Specific Mutations in Human Ras Gene. *Int J Pharm Sci Res*. 3(10); 3893-3914.