ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ

Βιολογία

Εργασία στην Βιολογία - Γονίδιο RAS

Επιβλέπουσα: **Αναστασιάδου Έμα**, Ερευνήτρια (βαθμίδα Δ'), ΙΙΒΕΑΑ

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2013

Βιολογία

Εργασία στην Βιολογία - Γονίδιο RAS

Αηδονόπουλος Σ.Ορφέας
Κωνσταντόπουλος Γ.Δημήτριος
Μπεγέτης Ι. Νικόλαος

ΑΜ: ΠΙΒ099 / ΠΙΒ0112 / ΠΙΒ0111

ΕΠΙΒΛΕΠΟΥΣΑ :

Αναστασιάδου Έμα, Ερευνήτρια (βαθμίδα Δ'), ΙΙΒΕΑΑ

Λέξεις-Κλειδιά :
οικογένεια RAS, KRAS, πρωτεϊνη GTPase, ανθρώπινα ογκογονίδια, σηματοδοσία κυττάρου

Biology

Assignment in Biology - RAS Gene

Aidonopoulos S.Orfeas
Begetis I. Nikolaos
Konstantopoulos G.Dimitrios

A.N.: PIV099 / PIV0111 / PIV0112

SUPERVISOR:

Anastasiadou Ema, Investigator (Lecturer Level), BRFAA

Keywords:
RAS family, KRAS, GTPase protein, human-oncogene, cell signaling

Table Of Contents

List Of Figures

CHAPTER 1

Initially, in this chapter we are going to provide some information about our assignment subject so that the reader of our report can easily follow the text flow. Subsequently, there will be provided the answers to the assignment's demands.

1.1 Introduction

Biology is a natural science concerned with the study of life and living organisms, including their structure, function, growth, evolution, distribution, and taxonomy. Biology has many subdisciplines unified by five so-called axioms of modern biology:

- ◇ Cells, which are the basic unit of life
- ◇ Genes, which are the basic unit of heredity
- ◇ New species and inherited traits, which are the product of evolution
- ◇ An organism, who regulates its internal environment to maintain a stable and constant condition
- ◇ Living organisms consume and transform energy

In this assignment we mostly deal with the second axiom, which in particular has to do with the subdiscipline of biology called molecular biology.

Molecular biology studies the complex interactions among biological molecules; This field overlaps with other areas of biology and chemistry, particularly thought with genetics and biochemistry.

Having given the above two definitions it would be recommended in here to state our team's specialization associated to these fields, the field of Bioinformatics. Bioinformatics is a branch between biology, computer science, mathematics, and engineering that develops

and improves upon methods for storing, retrieving, and analyzing biological data. It develops software tools to generate useful biological knowledge.

Bioinformatics deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modeling and simulation, discrete mathematics, control and system theory, circuit theory, and statistics.

The primary goal of bioinformatics is to increase the understanding of biological processes. What sets it apart from other approaches, however, is its focus on developing and applying computationally intensive techniques to achieve this goal, like in pattern recognition, data mining, machine learning algorithms, and visualization. Major research efforts in the field include sequence alignment, gene finding, genome assembly, drug design, drug discovery, protein structure alignment, protein structure prediction, prediction of gene expression and protein-protein interactions, genome-wide association studies and the modeling of evolution.

In this assignment our purpose is to provide as much useful information as it is requested by using Bioinformatic tools, i.e. we are responsible for the RAS gene speculation and more specifically for the KRAS gene, the KRAS promoter and the protein that it encodes.

1.2 RAS Gene

Ras (Figure 1.1) is the name given to a family of related proteins found inside cells, including human cells. All Ras protein family members belong to a class of protein called small GTPase, and are involved in transmitting signals within cells (cellular signal transduction). Ras is the prototypical member of the Ras superfamily of proteins, which are all related in 3D structure and regulate diverse cell behaviours. Proteopedia[1] provides us with a lot of useful information. In our case a 3D model of a subfamily of RAS, HRAS is analysed in depth[2]. But to continue with RAS, the name 'RAS' is an abbreviation of 'Rat sarcoma', reflecting the way the first members of the protein family were discovered. The name RAS is also used to refer to the family of genes encoding those proteins.

When RAS is 'switched on' by incoming signals, it subsequently switches on other proteins, which ultimately turn on genes involved in cell growth, differentiation and survival. As a result, mutations in RAS genes can lead to the production of permanently activated RAS proteins. This can cause unintended and overactive signalling inside the cell, even in the absence of incoming signals.

Because these signals result in cell growth and division, overactive RAS signaling can ultimately lead to cancer. RAS is the most common oncogene in human cancer - mutations that permanently activate RAS are found in 20-25% of all human tumors and up to 90% in certain types of cancer (e.g. pancreatic cancer). For this reason, RAS inhibitors are being studied as a treatment for cancer, and other diseases with RAS overexpression.

[1] http://www.proteopedia.org/wiki/index.php/Main_Page
[2] http://www.proteopedia.org/wiki/index.php/Ras
[4] http://en.wikipedia.org/wiki/File:Hras_surface_colored_by_conservation.png

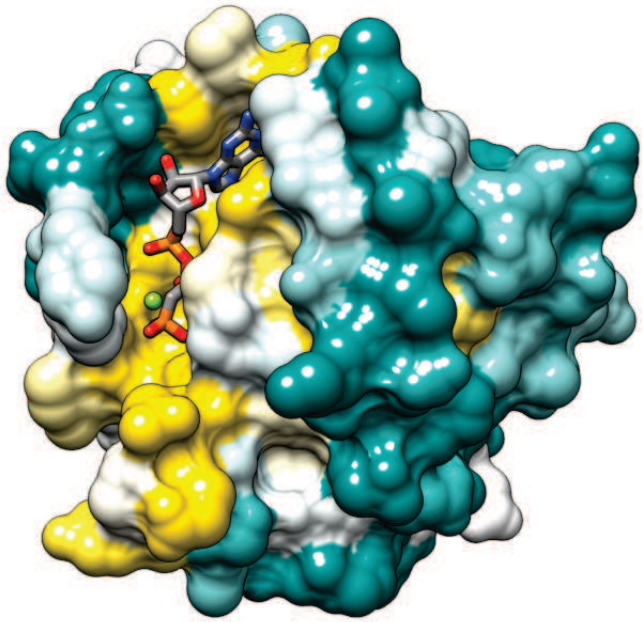


Figure 1.1: *MH-Ras structure PDB 121p, surface colored by conservation in Pfam seed alignment: gold, most conserved; dark cyan, least conserved.*[4]

The RAS superfamily is a protein superfamily of small GTPases, which are all related, to a degree, to the RAS protein subfamily (the key human members of which are KRAS, NRAS, and HRAS).

There are more than a hundred proteins in the RAS superfamily. Based on structure, sequence and function, the RAS superfamily is divided into eight main families, each of which is further divided into subfamilies: Ras, Rad, Rab, Rap, Ran, Rho, Rheb, Rit, and Arf. Miro is a recent contributor to the superfamily.

Each subfamily shares the common core G domain, which provides essential GTPase and nucleotide exchange activity.

The surrounding sequence helps determine the functional specificity of the small GTPase, for example the 'Insert Loop', common to the Rho subfamily, specifically contributes to binding to effector proteins such as IQGAP and WASP.

The RAS family is generally responsible for cell proliferation, Rho for cell morphology, Ran for nuclear transport and Rab and Arf for vesicle transport. Figure 1.2 shows some general details on the RAS gene family.

In order to be more complete in our report about RAS, and more exclusively KRAS, we are also providing a little information about the GTPase class of proteins and Chromosome 12.

GTPases are a large family of hydrolase enzymes that can bind and hydrolyze guanosine triphosphate (GTP). The GTP binding and hydrolysis takes place in the highly conserved G

[5]http://pfam.sanger.ac.uk/family/PF00071?tab=pdbBlock

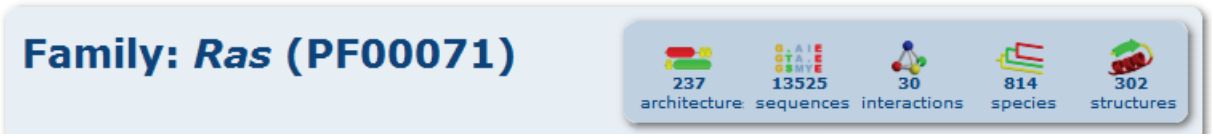


Figure 1.2: *Family: Ras (PF00071) details*[5]

domain common to all GTPases. RAS belongs to small GTPases, which have a molecular weight of about 21 kilodaltons and generally serve as molecular switches for a variety of cellular signaling events.

Chromosome 12(Figure 1.3) is one of the 23 pairs of chromosomes in humans. People normally have two copies of this chromosome. Chromosome 12[6] spans about 143 million base pairs (the building material of DNA) and represents between 4 and 4.5 percent of the total DNA in cells. 8 Identifying genes on each chromosome is an active area of genetic research. Because researchers use different approaches to predict the number of genes on each chromosome, the estimated number of genes varies. Chromosome 12 likely contains between 1,000 and 1,300 genes. It also contains the Homeobox C gene cluster. Some of the most known genes located on chromosome 12 are: ACVRL1, CBX5, COL2A1, HPD, KCNA1, KERA, LRRK2, MMAB, MYO1A, NANOG, PAH, PPP1R12A, PTPN11 and **KRAS**.

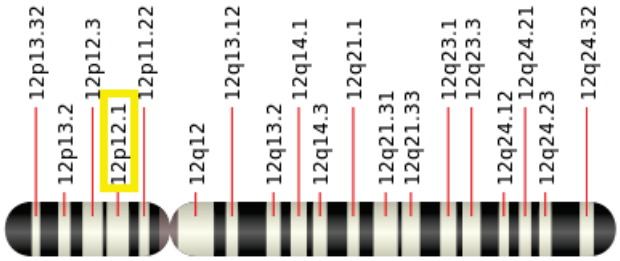


Figure 1.3: *Chromosome 12 - KRAS region*

1.3 KRAS

As indicated in the previous section Ras is a protein family whose members belong to a class of protein called small GTPase. Ras protein subfamily which is found of human genome has as members the three genes KRAS, NRAS, and HRAS.

In this report we decided to analyse and apply all the requisite given by the assignment on KRAS gene.

So to begin with, GTPase KRAS (Figure 1.4) also known as V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog and KRAS, is a protein that in humans is encoded by the KRAS

[6]http://ghr.nlm.nih.gov/chromosome/12

gene[1][2].

The protein product of the normal KRAS gene performs an essential function in normal tissue signaling, and the mutation of a KRAS gene is an essential step in the development of many cancers.[3] Like other members of the Ras family, the KRAS protein is a GTPase and is an early player in many signal transduction pathways. KRAS is usually tethered to cell membranes because of the presence of an isoprenyl group on its C-terminus.

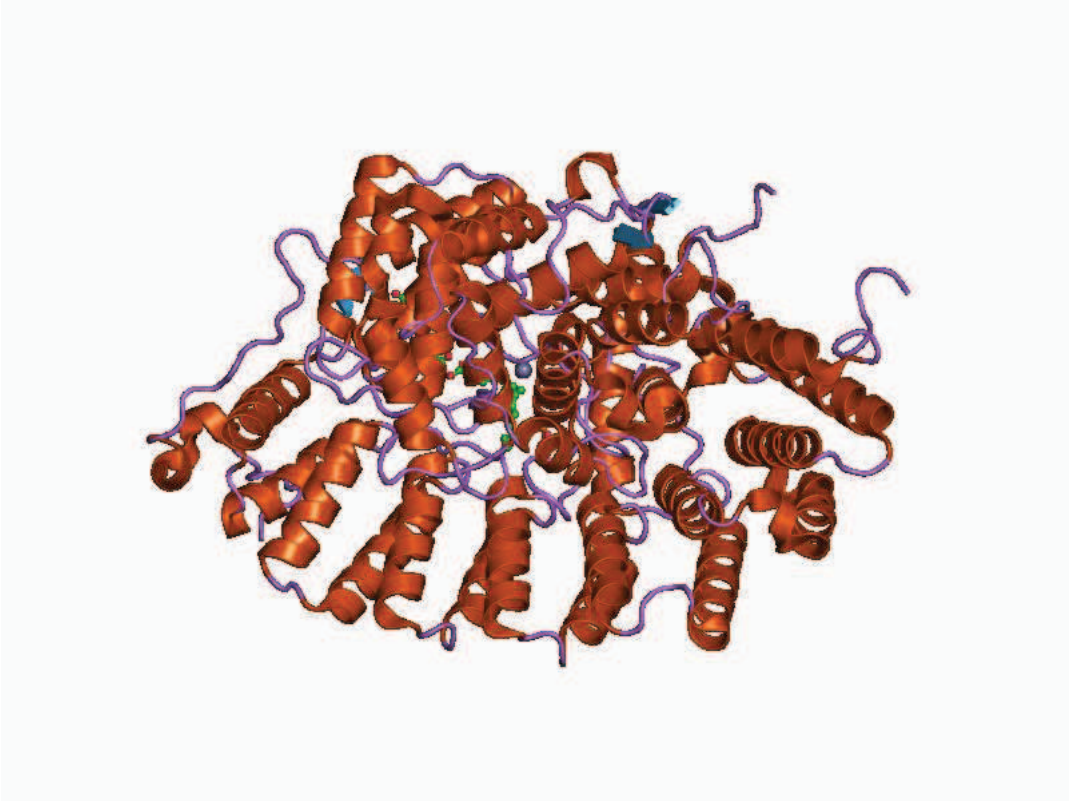


Figure 1.4: *V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog*[7]

1.3.1 KRAS Functionality

KRAS acts as a molecular on/off switch. Once it is turned on it recruits and activates proteins necessary for the propagation of growth factor and other receptors' signal, such as c-Raf and PI 3-kinase. KRAS binds to GTP in the active state and possesses an intrinsic enzymatic activity which cleaves the terminal phosphate of the nucleotide converting it to GDP. Upon conversion of GTP to GDP, KRAS is turned off. The rate of conversion is usually slow but can be sped up dramatically by an accessory protein of the GTPase-activating protein (GAP) class, for example RasGAP. In turn KRAS can bind to proteins of the Guanine Nucleotide

[7] http://en.wikipedia.org/wiki/File:PDB_1d8d_EBI.jpg

Exchange Factor (GEF) class, for example SOS1, which forces the release of bound nucleotide. Subsequently, KRAS binds GTP present in the cytosol and the GEF is released from ras-GTP. For more clues over KRAS functionality we can explore the tissue-specific pattern of mRNA expression. Figure 1.5 shows the results of this quest, and some explanation if provided in its caption.

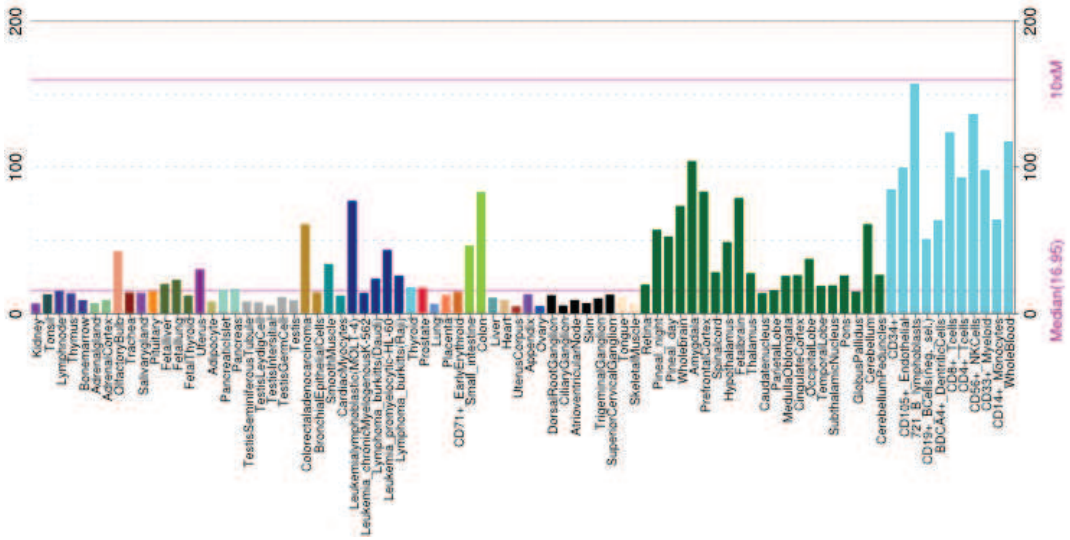


Figure 1.5: *KRAS Gene Expression/activity chart. The tissue-specific pattern of mRNA expression can indicate important clues about gene function. High-density oligonucleotide arrays offer the opportunity to examine patterns of gene expression on a genome scale. Toward this end, we have designed custom arrays that interrogate the expression of the vast majority of protein-encoding human and mouse genes and have used them to profile a panel of 79 human and 61 mouse tissues. The resulting data set provides the expression patterns for thousands of predicted genes, as well as known and poorly characterized genes, from mice and humans. We have explored this data set for global trends in gene expression, evaluated commonly used lines of evidence in gene prediction methodologies, and investigated patterns indicative of chromosomal organization of transcription. We describe hundreds of regions of correlated transcription and show that some are subject to both tissue and parental allele-specific expression, suggesting a link between spatial expression and imprinting*[9].

Symbol:	KRAS
Description:	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
Accessions:	3845 (NCBI Gene) ENSG00000133703 (Ensembl) P01116 (UniProt) 190070 (OMIM) 37990 (HomoloGene)
Aliases:	C-K-RAS, K-RAS2A, K-RAS2B, K-RAS4A, K-RAS4B, KI-RAS, KRAS1, KRAS2, NS, NS3, RASK2
Genome Location:	chr12:25357723-25403870 (hg19)
Function:	**Molecular Function** GTPase activity (GO:0003924) protein binding (GO:0005515) GTP binding (GO:0005525) GMP binding (GO:0019002) GDP binding (GO:0019003) LRR domain binding (GO:0030275) **Biological Process** MAPK cascade (GO:0000165) activation of MAPKK activity (GO:0000186) positive regulation of protein phosphorylation (GO:0001934) epidermal growth factor receptor signaling pathway (GO:0007173) small GTPase mediated signal transduction (GO:0007264) Ras protein signal transduction (GO:0007265) axon guidance (GO:0007411) blood coagulation (GO:0007596) positive regulation of cell proliferation (GO:0008284) insulin receptor signaling pathway (GO:0008286) visual learning (GO:0008542) fibroblast growth factor receptor signaling pathway (GO:0008543) positive regulation of gene expression (GO:0010628) cytokine-mediated signaling pathway (GO:0019221) actin cytoskeleton organization (GO:0030036) regulation of synaptic transmission, GABAergic (GO:0032228) positive regulation of Rac protein signal transduction (GO:0035022) social behavior (GO:0035176) positive regulation of MAP kinase activity (GO:0043406) negative regulation of neuron apoptotic process (GO:0043524) nerve growth factor receptor signaling pathway (GO:0048011) regulation of long-term neuronal synaptic plasticity (GO:0048169) leukocyte migration (GO:0050900) positive regulation of nitric-oxide synthase activity (GO:0051000) positive regulation of NF-kappaB transcription factor activity (GO:0051092) striated muscle cell differentiation (GO:0051146) response to glucocorticoid stimulus (GO:0051384) response to mineralocorticoid stimulus (GO:0051385) **Cellular Component** mitochondrion (GO:0005739) plasma membrane (GO:0005886) membrane raft (GO:0045121)
Interpro:	Mitochondrial Rho-like (IPR013684) Protein synthesis factor, GTP-binding (IPR000795) Ran GTPase (IPR002041) Small GTP-binding protein domain (IPR005225) Small GTPase superfamily (IPR001806) Small GTPase superfamily, ARF/SAR type (IPR006689) Small GTPase superfamily, Rab type (IPR003579) Small GTPase superfamily, Ras type (IPR020849) Small GTPase superfamily, Rho type (IPR003578)
Transcripts:	NM_004985.3 NM_033360.2 ENST00000256078 ENST00000311936 ENST00000556131 ENST00000557334
Proteins:	NP_004976.2 NP_203524.1 ENSP00000256078 ENSP00000308495 ENSP00000451856 ENSP00000452512
Reporters:	**HG-U133_Plus_2** 1559203_s_at 1559204_x_at 204009_s_at 204010_s_at 214352_s_at **HG-U95Av2** 1940_at 32159_at

Figure 1.6: *KRAS Gene Identifiers*

Aidonopoulos S.Orfeas
Begetis I.Nikolaos
Konstantopoulos G.Dimitrios

1.3.2 KRAS Clinical Significance

This proto-oncogene is a Kirsten ras oncogene homolog from the mammalian ras gene family. A single amino acid substitution, and in particular a single nucleotide substitution, is responsible for an activating mutation. The transforming protein that results is implicated in various malignancies, including lung adenocarcinoma, mucinous adenoma, ductal carcinoma of the pancreas and colorectal carcinoma.

Several germline KRAS mutations have been found to be associated with Noonan syndrome and cardio-facio-cutaneous syndrome. Somatic KRAS mutations are found at high rates in leukemias, colon cancer, pancreatic cancer and lung cancer.

Colorectal cancer: The chronological order of mutations is important in the impact of KRAS mutations in regard to colorectal cancer, with a primary KRAS mutation generally leading to a self-limiting hyperplastic or borderline lesion, but if occurring after a previous APC mutation it often progresses to cancer.

KRAS mutation is predictive of a very poor response to panitumumab (Vectibix•) and cetuximab (Erbitux•) therapy in colorectal cancer. Currently, the most reliable way to predict whether a colorectal cancer patient will respond to one of the EGFR-inhibiting drugs is to test for certain "activating" mutations in the gene that encodes KRAS, which occur in 40% of colorectal cancers. Studies show patients whose tumors express the mutated version of the KRAS gene will not respond to cetuximab or panitumumab.

Although presence of the wild-type (or normal) KRAS gene does not guarantee that these drugs will work, a number of large studies have shown that cetuximab has significant efficacy in mCRC patients with KRAS wild-type tumors. In the Phase III CRYSTAL study, published in 2009, patients with the wild-type KRAS gene treated with Erbitux plus chemotherapy showed a response rate of up to 59% compared to those treated with chemotherapy alone. Patients with the KRAS wild-type gene also showed a 32% decreased risk of disease progression compared to patients receiving chemotherapy alone.

Lung cancer: Whether a patient is positive or negative for a mutation in the epidermal growth factor receptor (EGFR) will predict how patients will respond to certain EGFR drugs such as erlotinib (Tarceva). EGFR positive patients have an impressive 60% response rate to erlotinib. However, KRAS positivity and EGFR positivity are generally mutually exclusive. Lung cancer patients who are KRAS positive have a low response rate to Tarceva estimated at 5% or less.

1.4 KRAS Promoters

Now that all the needed information is referred in the previous chapters we move on to the requests of our assignment. In this section we are providing information about the KRAS promoters and all the additional data that are not mentioned above, like the position in the

[9] http://plugins.biogps.org/data_chart/data_chart.cgi?id=3845

genome, the chromosome and the region in genome that it outspreads, and the exons that it includes.

At first we looked for KRAS gene in Ensembl genome browser[10]. There, the first description of KRAS, as an addition to all that was previously indicated, was that v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:6407] is a gene located in the **Chromosome 12**(Figure 1.7) at **six exon positions: 25,357,723-25,403,870** and in the **reverse(-1) strand** with the INSDC coordinates at the **human genome** (Figure 1.8)[11] set as **chromosome:GRCh37:CM000674.1 from 25357723 to 25403870**. Exons, having been referred earlier, are any nucleotide sequence encoded by a gene that remains present within the final mature RNA product of that gene after introns have been removed by RNA splicing. The term exon refers to both the DNA sequence within a gene and to the corresponding sequence in RNA transcripts. In RNA splicing, introns are removed and exons are covalently joined to one another as part of generating the mature messenger RNA or noncoding RNA product of a gene. Figure 1.9 shows the exon positions in KRAS[12]

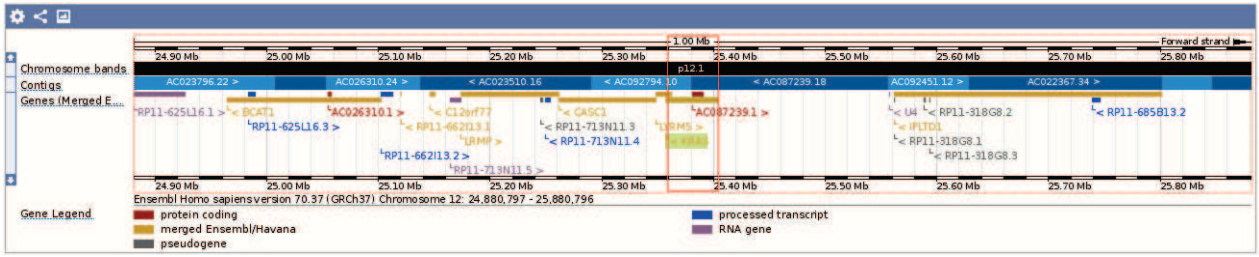


Figure 1.7: *KRAS location in Chromosome 12.1*

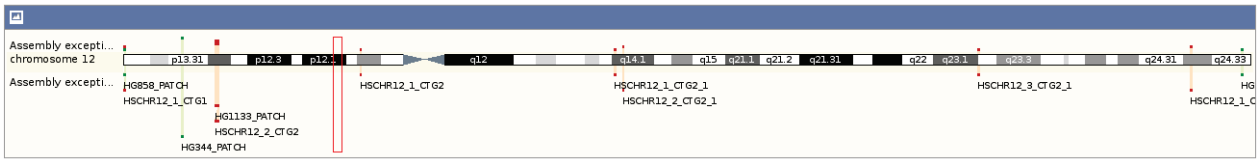


Figure 1.8: *KRAS location in Human Genome region of interest*

KRAS gene has four promoters as shown in Figure 1.10 and some useful information about each of them is provided in Figures 2.1, 2.2, 2.3 and 2.4.

In general a promoter(Figure 1.11) is a region of DNA that initiates transcription of a particular gene. Promoters are located near the genes they transcribe, on the same strand and upstream on the DNA (towards the 3' region of the anti-sense strand, also called template

[10]http://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000133703;r=12:25357723-25403870

[11]http://www.ensembl.org/Homo_sapiens/Location/View?db=core;g=ENSG00000133703;r=12:25357723-25403870

[12]http://atlasgeneticsoncology.org/Genes/KRASID91.html

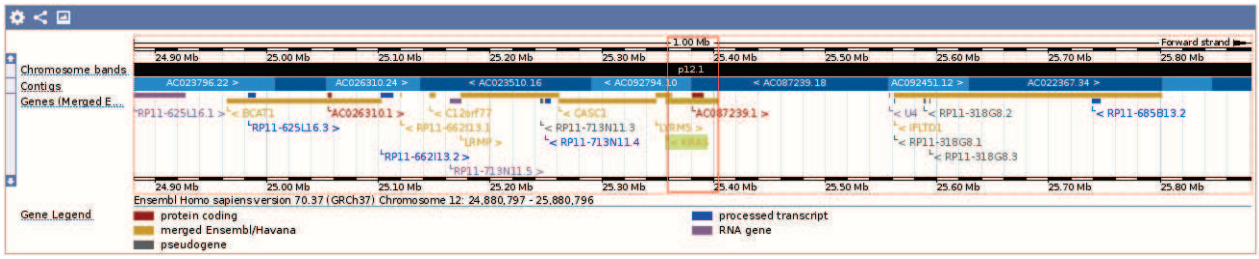


Figure 1.9: *KRAS six exons. KRAS splicing variants alternative splicing of K-ras precursor mRNA leads to the two transcripts which differ by the ex- or inclusion of Exon 4a; Exons that encode protein are shown as black boxes, untranslated exons as white boxes; the upstream untranslated exon is indicated as Exon -1*

Mark ALL | 1 to 4 of 4 | Page 1 | GO

	Gene ID	Gene Name	Species	Chromosome Location	Strand	Promoter ID	Quality	Transcription Start Site
☑	7209	KRAS2	human, Homo sapiens	12p12.1	-	10423	2: known	25298596
☑	7209	KRAS2	human, Homo sapiens	12p12.1	-	10424	2: known	25304037
☑	7209	KRAS2	human, Homo sapiens	12p12.1	-	10425	3.1: refseq,predicted	25304607
☑	7209	KRAS2	human, Homo sapiens	12p12.1	-	116525	3.2: refseq	25262391

Marked promoter sequence from -700 to 299 relative to transcription start site.

Figure 1.10: *KRAS Gene Promoters as exported from the Transcriptional Regulatory Element Database.*

strand and non-coding strand). Promoters can be about 100•1000 base pairs long. Promoters contain specific DNA sequences and response elements that provide a secure initial binding site for RNA polymerase and for proteins called transcription factors that recruit RNA polymerase. They are typically immediately adjacent to the gene in question, positions in the promoter are designated relative to the transcriptional start site, where transcription of DNA begins for a particular gene (i.e., positions upstream are negative numbers counting back from -1, for example -100 is a position 100 base pairs upstream), as happens in the case of KRAS gene.

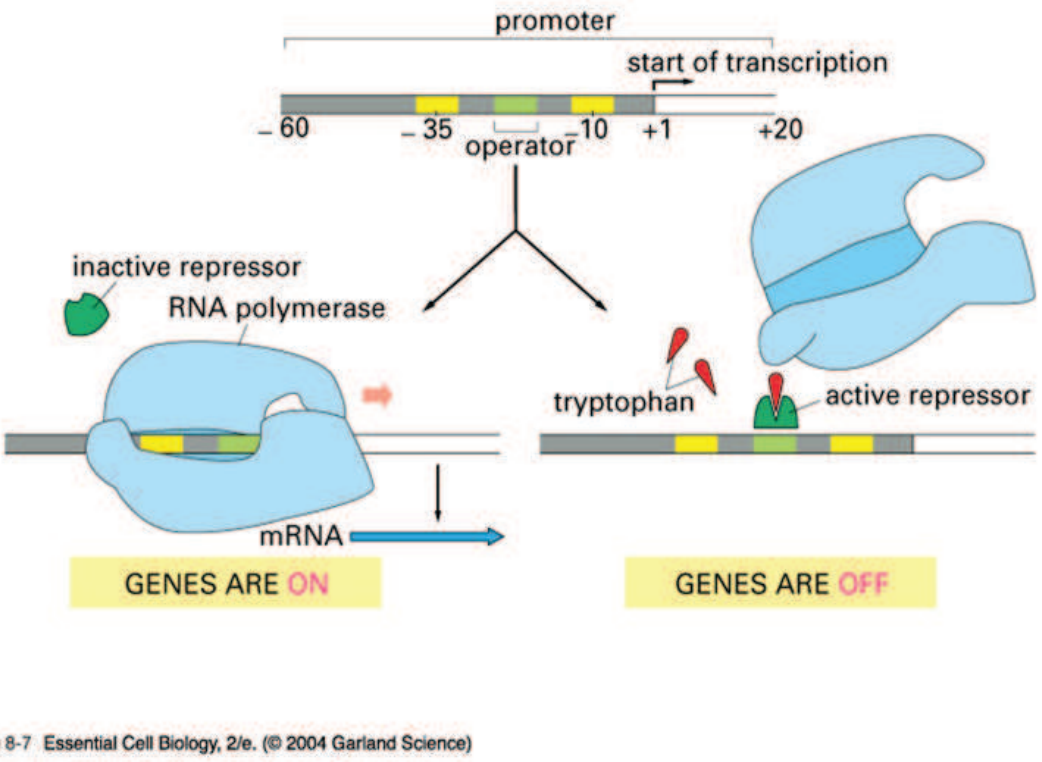


Figure 1.11: *Gene Promoter. RIGHT: Gene is "off". There is a repressor who binds to the operator of the promoter and the RNA polymerase from binding to the promoter. LEFT: The gene "on". Lactose inhibites the repressor, allowing the RNA polymerase to bind with the promoter, and express the genes.*

Moreover, as it was requested in the assignment we check in the KRAS promoter for regulatory areas such as TATAA box and binding regions, but the results will be provided at the end of this case study because a lot information should be explained in next sections.

1.5 KRAS Isoforms

Gene isoforms are mRNAs that are produced from the same locus but are different in their transcription start sites (TSSs), protein coding DNA sequences (CDSs) and/or untranslated regions (UTRs), potentially altering gene function.

Cis-regulatory elements in the promoter contain sequences recognized by transcription factors and the basal transcription machinery. So the location of the TSS is important for understanding the biogenesis of specific isoforms.The idea that different binding partners

confer different functional properties has been well studied in tissue-specific gene regulation. For example, the same transcription factor (TF) can direct gene expression in different tissues simply by binding with different TFs in each tissue. Isoforms harboring changes in the CDS have been the most thoroughly characterized because they commonly give rise to proteins with different functional properties. UTRs regulate the levels of primary transcript in numerous ways: transcript stability, folding and turnover, as well as translation ef?ciency. UTRs are often the target of miRNA, which typically downregulate transcript expression by triggering degradation or halting translation.

The gene isoforms can be sequenced by Whole Transcriptome Shotgun Sequencing (RNA-Seq). Recently some progress has been made to characterize known isoforms of regeneration associated genes (RAGs) using RNA-Seq, which is important in understanding the isoform diversity in the CNS.

KRAS gene has four transcription factors from which the two (at least, because only these two have a known protein coding, the two others protein coding is still putative) have more than one isoforms, as we found from the CCDS database[13],[14] of NCBI. KRAS's isoforms are shown in Figure 1.12 and more particularly in Figures 2.5 and 2.6, 2.7, 2.8, 2.9 and 2.10.

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
KRAS-001	ENST00000311936	5765	ENSP00000308495	188	Protein coding	CCDS8702
KRAS-002	ENST00000556131	1696	ENSP00000451856	43	Protein coding	-
KRAS-003	ENST00000557334	1052	ENSP00000452512	75	Protein coding	-
KRAS-004	ENST00000256078	1119	ENSP00000256078	189	Protein coding	CCDS8703

Figure 1.12: *KRAS Isoforms*.

1.6 KRAS Conserved Regions

In biology, conserved sequences are similar or identical sequences that occur within nucleic acid sequences (such as RNA and DNA sequences), protein sequences, protein structures or polymeric carbohydrates across species (orthologous sequences) or within different molecules produced by the same organism (paralogous sequences). In the case of cross species conservation, this indicates that a particular sequence may have been maintained by evolution despite speciation. The further back up the phylogenetic tree a particular conserved sequence may occur the more highly conserved it is said to be. Since sequence information is normally transmitted from parents to progeny by genes, a conserved sequence implies that there is a conserved gene[15].

[13]http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS8702
[14]http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS8703
[15]http://www.biology-online.org/biology-forum/about15854.html

It is widely believed that mutation in a "highly conserved" region leads to a non-viable life form, or a form that is eliminated through natural selection.

KRAS found in human gene when it is compared with other species as shown in Figure 1.13 has various conserved regions. Moreover, in the miscellaneous chapter one more figure(Figure 2.11) is presented that shows the conserved area by comparing the KRAS gene in human genome with that of mouse, dog, frog, etc.

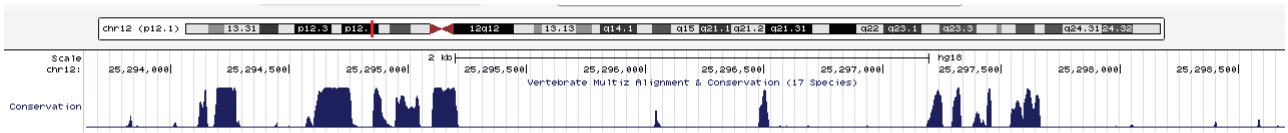


Figure 1.13: *KRAS Conserved Regions.*

1.7 KRAS Phylogenetic Tree

A phylogenetic tree is a tree diagram which depicts the lines of evolutionary descent of different•species, organisms, or genes from a common ancestor. The taxa joined together in the tree are implied to have descended from a common ancestor.

The descendants derived from a same ancestor are represented on each node of an evolutionary tree. Moreover, we can study how a particular gene or a sequence evolves in different species. When the species are together on a node we say that they have a "conserved gene" or a "conserved sequence". Phylogenetic trees can be exploited in several scientific biological fields, like bioinformatics, medical informatics, comparative genomics, etc.

KRAS gene's phylogenetic tree can be shown in details or more abstractly. Figure 1.14 shows the KRAS phylogenetic tree and Figures 2.12 and 2.13 show the phylogenetic tree with paralogs and in this Ensembl link[16] you can find the whole gene tree image for all organisms.

[16]http://www.ensembl.org/Multi/GeneTree?gt=ENSGT00690000101785

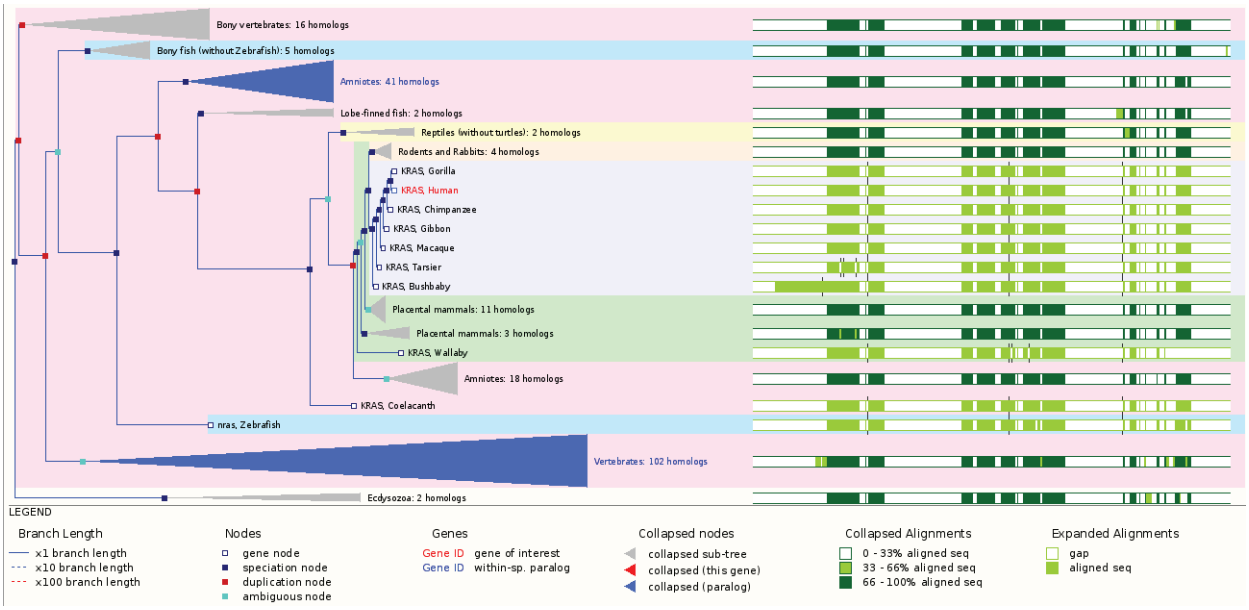


Figure 1.14: *KRAS Phylogenetic Tree.*

1.8 KRAS Proteins

KRAS is a GTPase protein, more specifically a G protein and plays a very important role in cell signaling. It was first discovered in the early 1980s and changed the understanding of molecular biology of cancer. The RAS mutation is responsible for the 20-30% of the development of all cancers. The protein sequence length is 189 AA and the molecular weight is 21,565 Da. KRAS plays an essential role in many signal transduction pathways, at least 4 pathways, such as RAS Map Kinase Path, as a molecular on/off switcher. KRAS is "on" when it is on its GTP bound form, when the protein has a GTP bound on it. A signal will be sent to the nucleus and cell division will be activated. In the "off" mode, the GDP bound form, the GTP bound is broken (typical GTPase reaction), after an intrinsic enzymatic activity when GTP is converted to GDP(Figure 1.15). When a mutation occurs, KRas is permanently in the "on" mode as the enzymatic activity doesn't occurs anymore and the cell cycle is constantly running, which results in the development of tumors and in cases cancer.

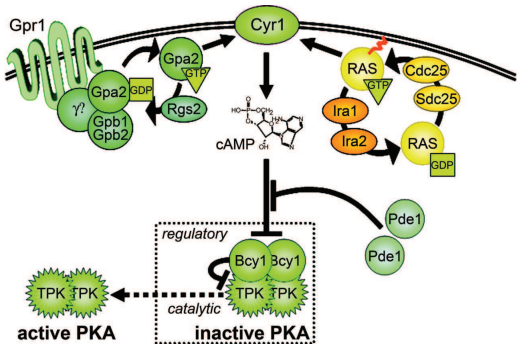


Figure 1.15: *KRAS Protein GTP-GDP switch*

KRAS protein is more likely to be found in mitochondria, on the plasma membrane and on the lipid raft.

KRAS gene functional regions have two aminoacid and polypeptide sequences as indicated in UniProt knowledge Base[17] and Enslembl browser (Figure 1.16).

Last modified July 21, 1986. Version 1.
Checksum: 973547B2E11C2C81

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

        70         80         90        100        110        120
QEEYSAMRDQ YMRTGEGFLC VFAINNTKSF EDIHHYREQI KRVKDSEDVP MVLVGNKCDL

       130        140        150        160        170        180
PSRTVDTKQA QDLARSYGIP FIETSAKTRQ RVEDAFYTLV REIRQYRLKK ISKEEKTPGC


VKIKKCIIM
```

Checksum: B1B6D189BB259861

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

        70         80         90        100        110        120
QEEYSAMRDQ YMRTGEGFLC VFAINNTKSF EDIHHYREQI KRVKDSEDVP MVLVGNKCDL

       130        140        150        160        170        180
PSRTVDTKQA QDLARSYGIP FIETSAKTRQ GVDDAFYTLV REIRKHKEKM SKDGKKKKKK


SKTKCVIM
```

Figure 1.16: *KRAS Polypeptide Sequences.*

In addition to the above shown in Figure 1.16 we get the following:

[17] http://www.uniprot.org/uniprot/P01116

1. **KRAS protein initiator is methionine:** M in position 1
2. **KRAS protein precursor or propeptide**: positions 187-189, sequence IIM
3. **KRAS protein interaction with DNA, GTP binding sites:**
   - ◇ Position 10-17, sequence GAGGVGKS
   - ◇ Position 57-61, sequence DTAGQ
   - ◇ Position 116-119, sequence NKCD
4. **Hypervariable region (region that are highly polymorphic):** positions 166-185, sequence YRLKKISKEEKTPGCVKIKK
5. **Conserved sequence motif of biological significance:** positions 32-40, sequence YDPTIEDSY

Furthermore, KRAS gene phosphorylation regions are shown in Figure 1.17.

Position	Species	ACC#	Gene ID	MW (DA)	Site	Group	Id	Residue Mod Type	(+/-)7 Sequence
KRas	2	human	P01116	3845	21,656	14326300	T2	Phosphorylation	MtEYKLVVV
KRas	4	human	P01116	3845	21,656	14326301	Y4	Phosphorylation	MTEyKLVVVGA
KRas	39	human	P01116	3845	21,656	18261833	S39	Phosphorylation	yDPTIEDsYRKQVVI
KRas	64	human	P01116	3845	21,656	5840743	Y64	Phosphorylation	TAGQEEySAMRDQY
KRas	148	human	P01116	3845	21,656	-	T148	Phosphorylation	FIETSAKtRQRVEDA
Kras	157	human	P01116	3845	21,656	-	T157	Phosphorylation	QRVEDAFyTLVREIR
Kras	158	human	P01116	3845	21,656	-	T158	Phosphorylation	QRVEDAFYtLVREIR

Figure 1.17: *KRAS Phosphorylation.*

Dimerization: It is known that the self-association of proteins to form dimers and higher-order oligomers is a very common phenomenon. Recent structural and biophysical studies show that protein dimerization or oligomerization is a key factor in the regulation of proteins such as enzymes, ion channels, receptors and transcription factors. In addition, self-association can help to minimize genome size, while maintaining the advantages of modular complex formation. **Oligomerization**, however, can also have deleterious consequences when nonnative oligomers associated with pathogenic states are generated.Specific protein dimerization is integral to biological function, structure and control, and must be under substantial selection pressure to be maintained with such frequency throughout biology[4].

As indicated in UniProt[18] there exist 13 secondary structure regions:

- ◇ Position 3-9, sequence: EYKLVVV, Beta Strand
- ◇ Position 16-25, sequence: KSALTIQLIQ, Helix
- ◇ Position 38-46, sequence: DSYRKQVVI, Beta Strand

[18]http://www.uniprot.org/uniprot/P01116

- ◇ Position 49-57, sequence: ETCLLDILD, Beta Strand
- ◇ Position 65-74, sequence: SAMRDQYMRT, Helix
- ◇ Position 76-83, sequence: EGFLCVFA, Beta Strand
- ◇ Position 87-104, sequence: TKSFEDIHHYREQIKRVK, Helix
- ◇ Position 111-116, sequence: MVLVGN, Beta Strand
- ◇ Position 120-122, sequence: LPS, Beta Strand
- ◇ Position 127-137, sequence: TKQAQDLARSY, Helix
- ◇ Position 141-143, sequence: FIE, Beta Strand
- ◇ Position 146-148, sequence: AKT, Turn
- ◇ Position 152-164, sequence: VEDAFYTLVREIR, Helix

See also Figure 1.18.

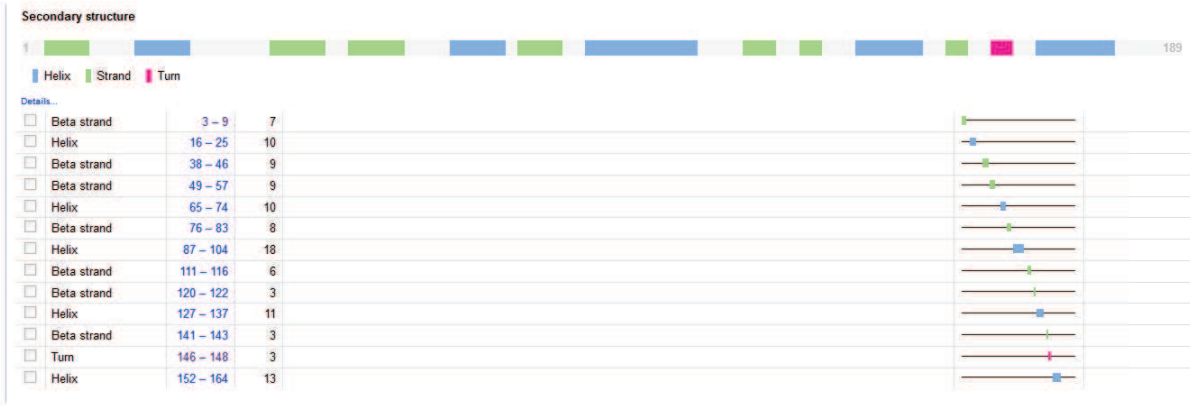


Figure 1.18: *KRAS Dimerization.*

To close the section of KRAS proteins, in the EBI and the UniProt database (link[19]), there can be found all the known information about illnesses occurred by KRAS mutations. Also in the link[20] someone can see the KRAS structure in a 3D shape view.

1.9 KRAS Transcription Factors

In this last section we are going to provide readers with the KRAS paralogues, orthologues and the protein families in general as shown from the Ensembl browser. Then we will continue with the subject of the section providing all the information that is asked in the assignment.

To start with, Homologous sequences are paralogous if they were separated by a gene duplication event: if a gene in an organism is duplicated to occupy two different positions in the same genome, then the two copies are paralogous.

KRAS paralogues[21] are shown in Figure 1.19

Show 10 entries | Show/hide columns | Filter

Type	Ancestral taxonomy	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id
Paralogue (within species)	Sarcopterygii	ENSG00000213281 NRAS No description	Region Comparison Alignment (protein) Alignment (cDNA)	1:115247090-115259515:-1	87	87
Paralogue (within species)	Vertebrata	ENSG00000174775 HRAS No description	Region Comparison Alignment (protein) Alignment (cDNA)	11:532242-537287:-1	87	87
Paralogue (within species)	Vertebrata	ENSG00000187682 ERAS No description	Region Comparison Alignment (protein) Alignment (cDNA)	X:48687283-48688548:1	36	44
Other paralogue (within species)	Opisthokonta	ENSG00000166592 RRAD No description	Region Comparison Alignment (protein) Alignment (cDNA)	16:66955582-66959547:-1	18	29
Other paralogue (within species)	Opisthokonta	ENSG00000152214 RIT2 No description	Region Comparison Alignment (protein) Alignment (cDNA)	18:40323192-40695657:-1	41	47
Other paralogue (within species)	Opisthokonta	ENSG00000143622 RIT1 No description	Region Comparison Alignment (protein) Alignment (cDNA)	1:155867599-155881195:-1	40	46
Other paralogue (within species)	Opisthokonta	ENSG00000164949 GEM No description	Region Comparison Alignment (protein) Alignment (cDNA)	8:95261481-95274578:-1	19	30
Other paralogue (within species)	Opisthokonta	ENSG00000139890 REM2 No description	Region Comparison Alignment (protein) Alignment (cDNA)	14:23352374-23356895:1	14	25
Other paralogue (within species)	Opisthokonta	ENSG00000158186 MRAS No description	Region Comparison Alignment (protein) Alignment (cDNA)	3:138066539-138124375:1	47	51
Other paralogue (within species)	Opisthokonta	ENSG00000088320 REM1 No description	Region Comparison Alignment (protein) Alignment (cDNA)	20:30063096-30072708:1	16	25

Figure 1.19: *KRAS Paralogues.*

On the other hand, Homologous sequences are orthologous if they were separated by a speciation event: when a species diverges into two separate species, the copies of a single gene in the two resulting species are said to be orthologous. Orthologs, or orthologous genes,

[19] www.uniprot.org/docs/humansavar

[20] http://www.phosphosite.org/showStructureAction.do;jsessionid=F02E7C65B8C9813818E91D7540152919?id=1D8D&p
+P%3D185-189&protOrg=12721

[21] http://www.ensembl.org/Homo_sapiens/Gene/Compara_Paralog?db=core;g=ENSG00000133703;r=12:25357723-25403870

are genes in different species that originated by vertical descent from a single gene of the last common ancestor. The term "ortholog" was coined in 1970 by Walter Fitch.

KRAS orthologues[22] are shown in Figure 1.20

Species	Type	dN/dS	Ensembl identifier & gene name	Compare	Location	Target %id	Query %id
Alpaca (*Vicugna pacos*)	1-to-1	n/a	ENSVPAG00000000319 KRAS v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:6407]	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	GeneScaffold_1647:2244309-2278213:-1	99	99
Anole lizard (*Anolis carolinensis*)	1-to-1	n/a	ENSACAG00000016948 KRAS v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:6407]	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	GL344156.1:55258-86337:-1	94	96
Anole lizard (*Anolis carolinensis*)	Possible ortholog	n/a	ENSACAG00000017404 -	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	GL343355.1:167031-170148:-1	88	86
Armadillo (*Dasypus novemcinctus*)	1-to-1	n/a	ENSDNOG00000018538 KRAS v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog [Source:HGNC Symbol;Acc:6407]	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	GeneScaffold_2240:195740-215343:-1	67	67
Bushbaby (*Otolemur garnettii*)	1-to-1	0.13421	ENSOGAG00000014812 KRAS No description	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	GL873576.1:4918803-4955574:-1	77	99
Caenorhabditis elegans (*Caenorhabditis elegans*)	1-to-many	n/a	ZK792.6 let-60 Ras protein let-60 [Source: UniProtKB/Swiss-Prot; acc: P22981]	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	IV:11688204-11691065:-1	77	75
Cat (*Felis catus*)	1-to-1 (apparent)	0.18049	ENSFCAG00000026808 K-RAS Ras p21 [Source: UniProtKB/TrEMBL; acc: QA2858]	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	B4:60615798-60648023:-1	91	90
Chicken (*Gallus gallus*)	1-to-1 (apparent)	n/a	ENSGALG00000014033 KRAS No description	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	1:69357722-69417249:-1	91	90
Chimpanzee (*Pan troglodytes*)	1-to-1	0.00000	ENSPTRG00000004775 KRAS No description	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	12:63267210-63313172:1	100	100
Chinese softshell turtle (*Pelodiscus sinensis*)	1-to-1 (apparent)	n/a	ENSPSIG00000015077 KRAS No description	Region Comparison, Alignment (protein), Alignment (cDNA), Gene Tree (image)	JH209943.1:330111-374079:-1	75	90

Figure 1.20: *KRAS Orthologues.*

KRAS protein families[23] are shown in Figure 1.21

Family ID	Consensus annotation	Other Human transcripts in this family	Multiple alignments
ENSFM00500000269718 (3 genes) (all proteins in family)	RAS PRECURSOR	ENST00000311936 (KRAS-001), ENST00000256078 (KRAS-004)	218 Ensembl members of this family JalView; 432 members of this family JalView
ENSFM00640001118512 (1 gene) (all proteins in family)	UNKNOWN	ENST00000556131 (KRAS-002)	1 Ensembl members of this family JalView; 1 members of this family JalView
ENSFM00640001118513 (1 gene) (all proteins in family)	UNKNOWN	ENST00000557334 (KRAS-003)	1 Ensembl members of this family JalView; 1 members of this family JalView

Figure 1.21: *KRAS Protein Families.*

In continuation we are now going to refer to transcription factors. Transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA by

[22] http://www.ensembl.org/Homo_sapiens/Gene/Compara_Ortholog?db=core;g=ENSG00000133703;r=12:25357723-25403870

[23] http://www.ensembl.org/Homo_sapiens/Gene/Family?db=core;g=ENSG00000133703;r=12:25357723-25403870

the enzyme RNA polymerase. Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA by the action of the correct enzymes. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement. Also unlike DNA replication where DNA is synthesised, transcription does not involve an RNA primer to initiate RNA synthesis.

Transcription proceeds in 5 or 6 steps, each moving like a wave along the DNA.

1. One or more sigma factors initiate transcription of a gene by enabling binding of RNA polymerase to promoter DNA.

2. Helicase enzymes move a transcription bubble, like the slider of a zipper, which splits the double helix DNA molecule into two strands of unpaired DNA nucleotides, by breaking the hydrogen bonds between complementary DNA nucleotides.

3. RNA polymerase adds matching RNA nucleotides that are paired with complementary DNA nucleotides of one DNA strand.

4. RNA sugar-phosphate backbone forms with assistance from RNA polymerase to form an RNA strand.

5. Hydrogen bonds of the untwisted RNA + DNA helix break, freeing the newly synthesized RNA strand.

6. If the cell has a nucleus, the RNA is further processed (addition of a 3'UTR poly-A tail and a 5'UTR cap) and exits to the cytoplasm through the nuclear pore complex.

Transcription(Figure 1.22) is the first step leading to gene expression. The stretch of DNA transcribed into an RNA molecule is called a transcription unit and encodes at least one gene. If the gene transcribed encodes a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either non-coding RNA genes (such as microRNA, lincRNA, etc.) or ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process, or other ribozymes.

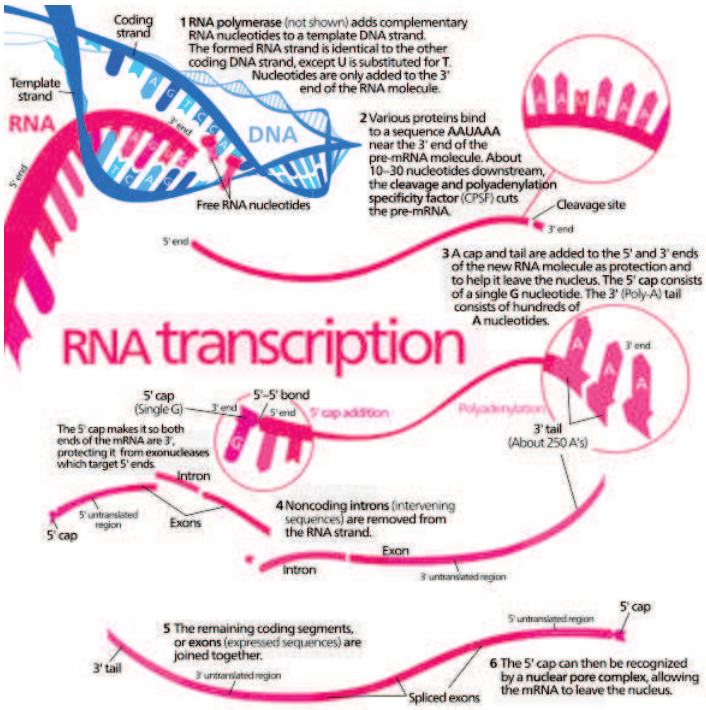


Figure 1.22: *Gene Trancription.*

KRAS gene has four transcripts as shown in Figure 1.23. Moreover, in Figure 1.24 there are depicted the transciption factors that regulate the expression of the KRAS gene (transcription factors). Each one of them can be trussed in the gene and initiate its transcription. Even more, there have been depicted the positions on the gene on which transcription factors are trussed. It is worthy to mention that every one of them can be trussed on more than one position. This is named as transcription binding sites. In Figure 1.25 it is depicted a general view of the transcription factors' binding sites[24].

Show/hide columns | Filter

Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
KRAS-001	ENST00000311936	5765	ENSP00000308495	188	Protein coding	CCDS8702
KRAS-002	ENST00000556131	1696	ENSP00000451856	43	Protein coding	-
KRAS-003	ENST00000557334	1052	ENSP00000452512	75	Protein coding	-
KRAS-004	ENST00000256078	1119	ENSP00000256078	189	Protein coding	CCDS8703

Figure 1.23: *KRAS Trancription.*

[24] http://www.sabiosciences.com/chipqpcrsearch.php?species_id=0&nfactor=n&ninfo=n&ngene=n&B2=Search&src=ge

TRANSCRIPTION FACTOR (NAME)	BINDING SITE Position on chromosome						
p53	25407536						
AP-1	25401885	25411753	25414554	25415632	25422823	25422824	25423753
c-Jun	25396500				25414554		
Elk-1	25401118						
NF-kappaB	25422620						
NF-kappaB1	25401656	25403413	25403589	25408769	25422620		
STAT1beta	25408593	25413335	25414794	25416578	25418839		
STAT1alpha	25408593	25413335	25414794	25416578	25418839		
STAT1	25408593	25413335	25414794	25416578	25418839	25423349	
PPAR-gamma1	25396294	25397112	25399412	25404137			

Figure 1.24: *KRAS Trancription Factors.*

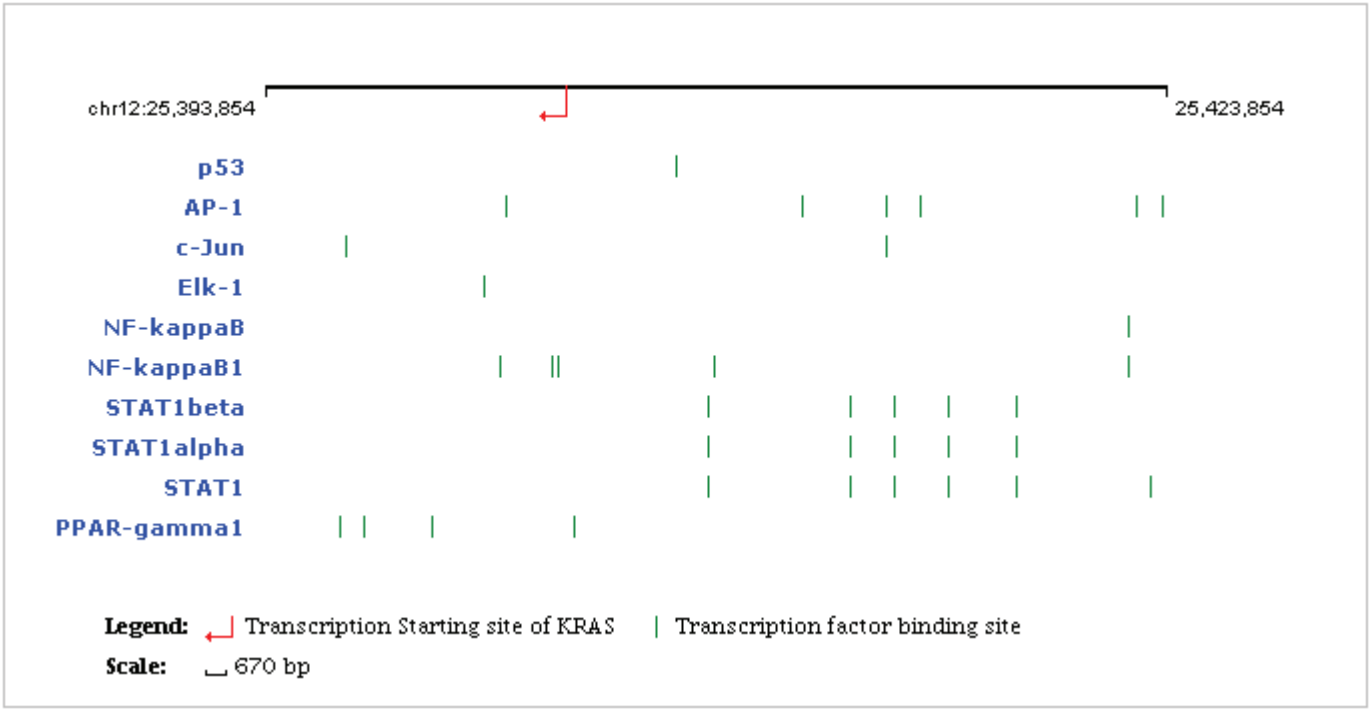


Figure 1.25: *KRAS Trancription Factors Binding Sites.*

Finally, as the last quest, we are asked to refer to the micro molecules of RNA(miRNA) that can involve in the transcripts. More specifically, microRNA (abbr. miRNA) is a small non-coding RNA molecule (ca. 22 nucleotides) found in plants and animals, which functions

in transcriptional and post-transcriptional regulation of gene expression. Encoded by eukaryotic nuclear DNA, miRNAs function via base-pairing with complementary sequences within mRNA molecules, usually resulting in gene silencing via translational repression or target degradation. The human genome may encode over 1000 miRNAs, which may target about 60% of mammalian genes and are abundant in many human cell types.

In our KRAS case study the miRNA that we found to be involved in KRAS transcripts are shown in Figure 1.26. The results illustrated in this figure came from the MiRBase database and there are 62 of them. We also verified them from the GeneCards database[25].

Accession	ID	Start	End	Strand
MI0016049	hsa-mir-3649	1769481	1769546	-
MI0006379	hsa-mir-1244-1	9392063	9392147	-
MI0015833	hsa-mir-4302	26026953	26027012	-
MI0016856	hsa-mir-4494	47757969	47758052	-
MI0006353	hsa-mir-1291	49048227	49048313	-
MI0017334	hsa-mir-4701	49165758	49165820	-
MI0006355	hsa-mir-1293	50627925	50627995	-
MI0017335	hsa-mir-3198-2	54625181	54625260	-
MI0003629	hsa-mir-616	57912946	57913042	-
MI0000750	hsa-mir-26a-2	58218392	58218475	-
MI0016688	hsa-mir-548z	65016289	65016385	-
MI0020351	hsa-mir-6074	66417400	66417506	-
MI0006426	hsa-mir-1279	69666937	69666998	-
MI0016417	hsa-mir-3913-1	69978502	69978603	-
MI0003631	hsa-mir-617	81226312	81226408	-
MI0003632	hsa-mir-618	81329515	81329612	-
MI0016857	hsa-mir-4495	98332834	98332899	-
MI0015834	hsa-mir-4303	98389161	98389226	-
MI0003633	hsa-mir-619	109230684	109230782	-
MI0016057	hsa-mir-3657	112475403	112475519	-
MI0006362	hsa-mir-1302-1	113132839	113132981	-
MI0003634	hsa-mir-620	116586365	116586459	-
MI0016824	hsa-mir-4472-2	116866057	116866123	-
MI0006271	hsa-mir-1178	120151439	120151529	-
MI0016860	hsa-mir-4498	120593238	120593303	-
MI0015832	hsa-mir-4304	123495214	123495275	-

Figure 1.26: *KRAS Trancription miRNAs.*

[25] http://www.genecards.org/cgi-bin/carddisp.pl?gene=KRAS

CHAPTER 2

Miscellaneous

In this chapter we present some more detailed information that are referred in the first chapter. The ordination of this chapter is analogous to the ordination of the first chapter. It is worth noting that all figures shown in the chapter are screenshots from a lot of relevant to what is requested databases and browsers, which are cited in the caption of each figure.

2.1 KRAS Promoters

Accession Number 10423
Gene KRAS2
Species human, Homo sapiens
Chromosome 12
Strand -
Transcription Start Site 25298596
Quality 2: known - stated explicitly in GenBank records
References
- [1] GenBank Nucleotide: NM_004985
  Homo sapiens v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog (KRAS2), transcript variant b, mRNA.

Sequence

```
CTCATTGCAACCTCGGACTCCTATTTTCCCCAGAGATATTTCACACATTA   -651
AAATGTCGTCAAATATTGTTCTTCTTTGCCTCAGTGTTTAAATTTTTATT   -601
TCCCCATGACACAATCCAGCTTTATTTGACACTCATTCTCTCAACTCTCA   -551
TCTGATTCTTACTGTTAATATTTATCCAAGAGAACTACTGCCATGATGCT   -501
TTAAAAGTTTTTCTGTAGCTGTTGCATATTGACTTCTAACACTTAGAGGT   -451
GGGGGTCCACTAGGAAAACTGTAACAATAAGAGTGGAGATAGCTGTCAGC   -401
AACTTTTGTGAGGGTGTGCTACAGGGTGTAGAGCACTGTGAAGTCTCTAC   -351
ATGAGTGAAGTCATGATATGATCCTTTGAGAGCCTTTAGCCGCCGCAGAA   -301
CAGCAGTCTGGCTATTTAGATAGAACAACTTGATTTTAAGATAAAAGAAC   -251
TGTCTATGTAGCATTTATGCATTTTTCTTAAGCGTCGATGGAGGAGTTTG   -201
TAAATGAAGTACAGTTCATTACGATACACGTCTGCAGTCAACTGGAATTT   -151
TCATGATTGAATTTTGTAAGGTATTTTGAAATAATTTTTCATATAAAGGT   -101
GAGTTTGTATTAAAAGGTACTGGTGGAGTATTTGATAGTGTATTAACCTT    -51
ATGTGTGACATGTTCTAATATAGTCACATTTTCATTATTTTTATTATAAG     -1
GCCTGCTGAAAATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGC     49
GTAGGCAAGAGTGCCTTGACGATACAGCTAATTCAGAATCATTTTGTGGA     99
CGAATATGATCCAACAATAGAGGTAAATCTTGTTTTAATATGCATATTAC    149
TGGTGCAGGACCATTCTTTGATACAGATAAAGGTTTCTCTGACCATTTTC    199
ATGAGTACTTATTACAAGATAATTATGCTGAAAGTTAAGTTATCTGAAAT    249
GTACCTTGGGTTTCAAGTTATATGTAACCATTAATATGGGAACTTTACTT    299
```

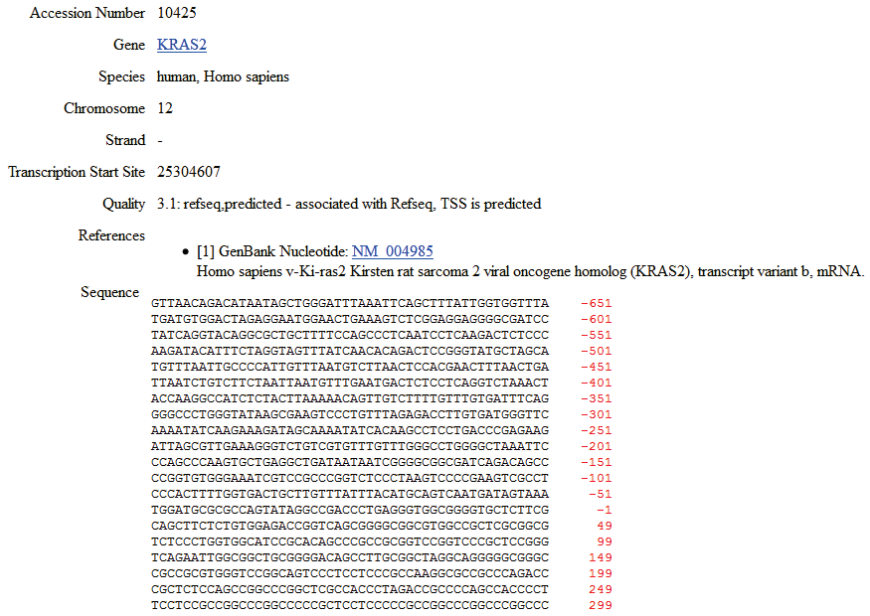
Figure 2.1: *KRAS Gene 1st Promoter. Starts at Transcription Start Site 25304607* [1].

Accession Number 10424
Gene KRAS2
Species human, Homo sapiens
Chromosome 12
Strand -
Transcription Start Site 25304037
Quality 2: known - stated explicitly in GenBank records
References
- [1] GenBank Nucleotide: NM_004985
  Homo sapiens v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog (KRAS2), transcript variant b, mRNA.
- [2] Database of Transcriptional Start Sites: DBTSS:NM_004985
  DBTSS

Sequence

```
CCCGGTCTCCCTAAGTCCCCGAAGTCGCCTCCCACTTTTGGTGACTGCTT   -651
GTTTATTTACATGCAGTCAATGATAGTAAATGGATGCGCGCCAGTATAGG   -601
CCGACCCTGAGGGTGGCGGGGGTGCTCTTCGCAGCTTCTCTGTGGAGACCG   -551
GTCAGCGGGGCGGCGTGGCCCGCTCGCGGCGTCTCCCTGGTGGCATCCGCA   -501
CAGCCCGCCGCGGTCCGGTCCCGCTCCGGGTCAGAATTGGCGGCTGCGGG   -451
GACAGCCTTGCGGCTAGGCAGGGGGCGGGCCGCCGCGTGGGTCCGGCAGT   -401
CCCTCCTCCCGCCAAGGCGCCGCCCAGACCCGCTCTCCAGCCGGCCCGGC   -351
TCGCCACCCTAGACCGCCCCAGCCACCCCTTCCTCCGCCGGCCCGGCCCC   -301
CGCTCCTCCCCCGCCGGCCCGGCCCGGCCCCCTCCTTCTCCCCGCCGGCG   -251
CTCGCTGCCTCCCCCCTCTTCCCTCTTCCCACACCGCCCTCAGCCGCTCCC   -201
TCTCGTACGCCCGTCTGAAGAAGAATCGAGCGCGGAACGCATCGATAGCT   -151
CTGCCCTCTGCGGCCGCCCGGCCCCGAACTCATCGGTGTGCTCGGAGCTC   -101
GATTTTCCTAGGCGGCGGCCGCGGCGGCGGAGGCAGCAGCGGCGGCGGCA    -51
GTGGCGGCGGCGAAGGTGGCGGCGGCTCGGCCAGTACTCCCGGCCCCCGC     -1
CATTTCGGACTGGGAGCGAGCGCGGCGCAGGCACTGAAGGCGGCGGCGGG     49
GCCAGAGGCTCAGCGGCTCCCAGGTGCGGGAGAGAGGTACGGAGCGGACC     99
ACCCCTCCTGGGCCCCTGCCCGGGTCCCGACCCTCTTTGCCGGCGCCGGG    149
CGGGGCCGGCGGCGAGTGAATGAATTAGGGGTCCCCGGAGGGGCGGGTGG    199
GGGGCGCGGGCGCGGGGTCGGGGCGGGCTGGGTGAGAGGGGTCTGCAGGG    249
GGGAGGCGCGCGGACGCGGCGGCGCGGGGAGTGAGGAATGGGCGGTGCGG    299
```

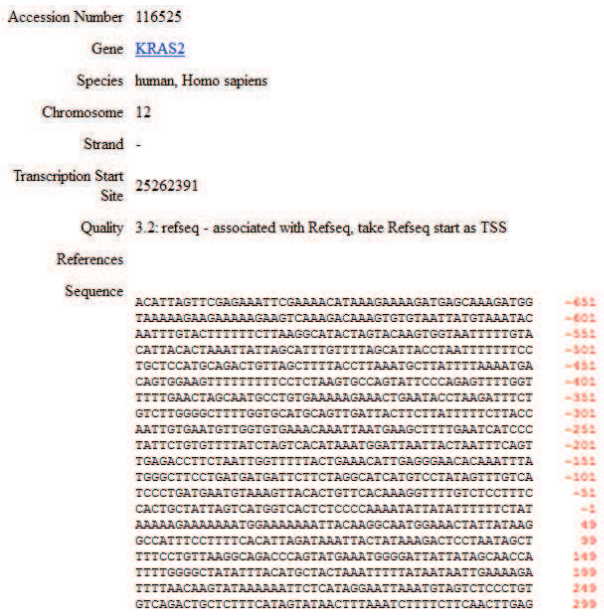
Figure 2.2: *KRAS Gene 2nd Promoter. Starts at Transcription Start Site 25304037*[2].



Accession Number	10425
Gene	KRAS2
Species	human, Homo sapiens
Chromosome	12
Strand	-
Transcription Start Site	25304607
Quality	3.1: refseq,predicted - associated with Refseq, TSS is predicted
References	• [1] GenBank Nucleotide: NM_004985 Homo sapiens v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog (KRAS2), transcript variant b, mRNA.
Sequence	

```
GTTAACAGACATAATAGCTGGGATTTAAATTCAGCTTTATTGGTGGTTTA   -651
TGATGTGGACTAGAGGAATGGAACTGAAAGTCTCGGAGGAGGGGCGATCC   -601
TATCAGGTACAGGCGCTGCTTTTCCAGCCCTCAATCCTCAAGACTCTCCC   -551
AAGATACATTTCTAGGTAGTTTATCAACACAGACTCCGGGTATGCTAGCA   -501
TGTTTAATTGCCCCATTGTTTAATGTCTTAACTCCACGAACTTTAACTGA   -451
TTAATCTGTCTTCTAATTAATGTTTGAATGACTCTCCTCAGGTCTAAACT   -401
ACCAAGGCCATCTCTACTTAAAAACAGTTGTCTTTTGTTTGTGATTTCAG   -351
GGGCCCTGGGTATAAGCGAAGTCCCTGTTTAGAGACCTTGTGATGGGTTC   -301
AAAATATCAAGAAAGATAGCAAAATATCACAAGCCTCCTGACCCGAGAAG   -251
ATTAGCGTTGAAAGGGTCTGTCGTGTTTGTTTGGGCCTGGGGCTAAATTC   -201
CCAGCCCAAGTGCTGAGGCTGATAATAATCGGGGCGGCGATCAGACAGCC   -151
CCGGTGTGGGAAATCGTCCGCCCGGTCTCCCTAAGTCCCCGAAGTCGCCT   -101
CCCACTTTTGGTGACTGCTTGTTTATTTACATGCAGTCAATGATAGTAAA    -51
TGGATGCGCGCCAGTATAGGCCGACCCTGAGGGTGGCGGGGGTGCTCTTCG     -1
CAGCTTCTCTGTGGAGACCGGTCAGCGGGGCGGCGTGGCCGCTCGCGGCG     49
TCTCCCTGGTGGCATCCGCACAGCCCGCCGCGGTCCGGTCCCGCTCCGGG     99
TCAGAATTGGCGGCTGCGGGGACAGCCTTGCGGCTAGGCAGGGGGCGGGC    149
CGCCGCGTGGGTCCGGCAGTCCCTCCTCCCGCCAAGGCGCCGCCCAGACC    199
CGCTCTCCAGCCGGCCCGGCTCGCCACCCTAGACCGCCCCAGCCACCCCT    249
TCCTCCGCCGGCCCGGCCCCCGCTCCTCCCCCGCCGGCCCGGCCCGGCCC    299
```

Figure 2.3: *KRAS Gene 3rd Promoter. Starts at Transcription Start Site 25298596*[3].



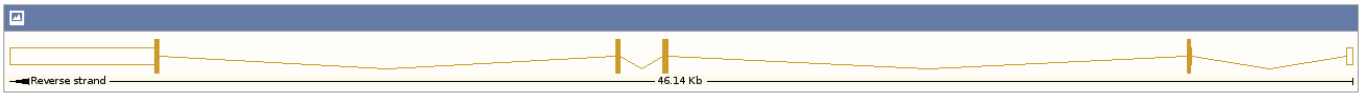
Accession Number	116525
Gene	KRAS2
Species	human, Homo sapiens
Chromosome	12
Strand	-
Transcription Start Site	25262391
Quality	3.2: refseq - associated with Refseq, take Refseq start as TSS
References	
Sequence	

```
ACATTAGTTCGAGAAATTCGAAAACATAAAGAAAAGATGAGCAAAGATGG   -651
TAAAAAGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAATTATGTAAATAC   -601
AATTTGTACTTTTTTCTTAAGGCATACTAGTACAAGTGGTAATTTTTGTA   -551
CATTACACTAAATTATTAGCATTTGTTTTAGCATTACCTAATTTTTTTCC   -501
TGCTCCATGCAGACTGTTAGCTTTTACCTTAAATGCTTATTTTAAAATGA   -451
CAGTGGAAGTTTTTTTTTTCCTCTAAGTGCCAGTATTCCCAGAGTTTTGGT   -401
TTTTGAACTAGCAATGCCTGTGAAAAAGAAACTGAATACCTAAGATTTCT   -351
GTCTTGGGGCTTTTGGTGCATGCAGTTGATTACTTCTTATTTTTCTTACC   -301
AATTGTGAATGTTGGTGTGAAACAAATTAATGAAGCTTTTGAATCATCCC   -251
TATTCTGTGTTTTATCTAGTCACATAAATGGATTAATTACTAATTTCAGT   -201
TGAGACCTTCTAATTGGTTTTTACTGAAACATTGAGGGAACACAAATTTA   -151
TGGGCTTCCTGATGATGATTCTTCTAGGCATCATGTCCTATAGTTTGTCA   -101
TCCCTGATGAATGTAAAGTTACACTGTTCACAAAGGTTTTGTCTCCTTTC    -51
CACTGCTATTAGTCATGGTCACTCTCCCCAAAATATTATATTTTTTCTAT     -1
AAAAAGAAAAAAATGGAAAAAAATTACAAGGCAATGGAAACTATTATAAG     49
GCCATTTCCTTTTCACATTAGATAAATTACTATAAAGACTCCTAATAGCT     99
TTTCCTGTTAAGGCAGACCCAGTATGAAATGGGGATTATTATAGCAACCA    149
TTTTGGGGCTATATTTACATGCTACTAAATTTTTATAATAATTGAAAAGA    199
TTTTAACAAGTATAAAAAATTCTCATAGGAATTAAATGTAGTCTCCCTGT    249
GTCAGACTGCTCTTTCATAGTATAACTTTAAATCTTTTCTTCAACTTGAG    299
```

Figure 2.4: *KRAS Gene 4th Promoter. Starts at Transcription Start Site 25262391*[4].

[4] http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=promInfo&pid=10425
[4] http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=promInfo&pid=10424
[4] http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process= promInfo&pid=10423
[4] http://rulai.cshl.edu/cgi-bin/TRED/tred.cgi?process=promInfo&pid=116525

2.2 KRAS Isoforms



Statistics	**Exons:** 5 **Coding exons:** 4 **Transcript length:** 5,765 bps **Translation length:** 188 residues
CCDS	This transcript is a member of the Human CCDS set: CCDS8702
Ensembl version	ENST00000311936.3
Type	Known protein coding

Figure 2.5: *KRAS Gene 1st transcript isoform. As can be seen from the figure and the details this transcript has more isoforms*[6].

Original	Current	Source	Nucleotide ID	Protein ID	Status in CCDS	Seq. Status	Links
✔	✔	EBI,WTSI	ENST00000311936	ENSP00000308495	Accepted	alive	N P N P
✔	✔	EBI,WTSI	OTTHUMT00000412230	OTTHUMP00000245391	Accepted	alive	N P N P
✔	✔	NCBI	NM_004985.3	NP_004976.2	Accepted	alive	N P N P B

CCDS Sequence Data

Blue highlighting indicates alternate exons.

Red highlighting indicates amino acids encoded across a splice junction.

Mouse over the nucleotide or protein sequence below and click on the highlighted codon or residue to select the pair.

Nucleotide Sequence (567 nt):

```
ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAA
TTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAGCAAGTAGTAATTGA
TGGAGAAACCTGTCTCTTGGATATTCTCGACACAGCAGGTCAAGAGGAGTACAGTGCAATGAGGGACCAG
TACATGAGGACTGGGGAGGGCTTTCTTTGTGTATTTGCCATAAATAATACTAAATCATTTGAAGATATTC
ACCATTATAGAGAACAAATTAAAAGAGTTAAGGACTCTGAAGATGTACCTATGGTCCTAGTAGGAAATAA
ATGTGATTTGCCTTCTAGAACAGTAGACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATTCCT
TTTATTGAAACATCAGCAAAGACAAGACAGGGTGTTGATGATGCCTTCTATACATTAGTTCGAGAAATTC
GAAAACATAAAGAAAAGATGAGCAAAGATGGTAAAAAGAAGAAAAAGAAGTCAAAGACAAAGTGTGTAAT
TATGTAA
```

Translation (188 aa):

```
MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQ
YMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIP
FIETSAKTRQGVDDAFYTLVREIRKHKEKMSKDGKKKKKKSKTKCVIM
```

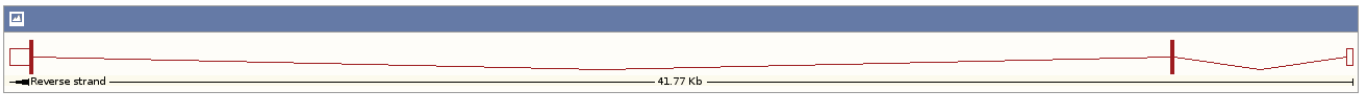
Figure 2.6: *KRAS Gene 1st isoform Nucleotide sequences and Translation. As shown in the figure there are three different Protein that can be altered in the first trancript. And the subsequences of these are highlighted in blue color*[8].

[6]http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000133703;r=12:25357723-25403870;t=ENST00000311936

[8]http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS8702

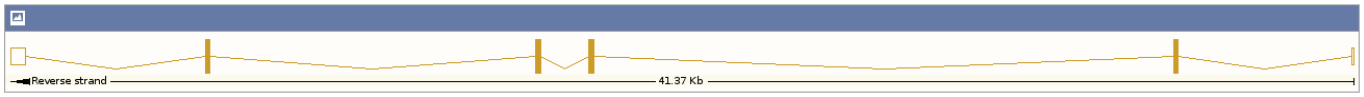
Statistics	**Exons:** 3 **Coding exons:** 2 **Transcript length:** 1,696 bps **Translation length:** 43 residues
Ensembl version	ENST00000556131.1
Type	Putative protein coding

Figure 2.7: *KRAS Gene 2nd transcript isoform*[9].



Statistics	**Exons:** 3 **Coding exons:** 2 **Transcript length:** 1,052 bps **Translation length:** 75 residues
Ensembl version	ENST00000557334.1
Type	Putative protein coding

Figure 2.8: *KRAS Gene 3rd transcript isoform*[10].



Statistics	**Exons:** 6 **Coding exons:** 4 **Transcript length:** 1,119 bps **Translation length:** 189 residues
CCDS	This transcript is a member of the Human CCDS set: CCDS8703
Ensembl version	ENST00000256078.4
Type	Known protein coding

Figure 2.9: *KRAS Gene 4th transcript isoform. As can be seen from the figure and the details this transcript has more isoforms*[12].

[9] http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000133703;r=12:25357723-25403870;t=ENST00000556131

[10] http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000133703;r=12:25357723-25403870;t=ENST00000557334

[12] http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000133703;r=12:25357723-25403870;t=ENST00000256078

Original	Current	Source	Nucleotide ID	Protein ID	Status in CCDS	Seq. Status	Links
✔	✔	EBI,WTSI	ENST00000256078	ENSP00000256078	Accepted	alive	N P N P
✔	✔	EBI,WTSI	OTTHUMT00000412232	OTTHUMP00000245393	Accepted	alive	N P N P
✔	✔	NCBI	NM_033360.2	NP_203524.1	Accepted	alive	N P N P B

CCDS Sequence Data

Blue highlighting indicates alternate exons.

Red highlighting indicates amino acids encoded across a splice junction.

Mouse over the nucleotide or protein sequence below and click on the highlighted codon or residue to select the pair.

Nucleotide Sequence (570 nt):

ATGACTGAATATAAACTTGTGGTAGTTGGAGCTGGTGGCGTAGGCAAGAGTGCCTTGACGATACAGCTAA
TTCAGAATCATTTTGTGGACGAATATGATCCAACAATAGAGGATTCCTACAGGAAGCAAGTAGTAATTGA
TGGAGAAACCTGTCTCTTGGATATTCTCGACACAGCAGGTCAAGAGGAGTACAGTGCAATGAGGGACCAG
TACATGAGGACTGGGGAGGGCTTTCTTTGTGTATTTGCCATAAATAATACTAAATCATTTGAAGATATTC
ACCATTATAGAGAACAAATTAAAAGAGTTAAGGACTCTGAAGATGTACCTATGGTCCTAGTAGGAAATAA
ATGTGATTTGCCTTCTAGAACAGTAGACACAAAACAGGCTCAGGACTTAGCAAGAAGTTATGGAATTCCT
TTTATTGAAACATCAGCAAAGACAAGACAGAGAGTGGAGGATGCTTTTTATACATTGGTGAGGGAGATCC
GACAATACAGATTGAAAAAAATCAGCAAAGAAGAAAAGACTCCTGGCTGTGTGAAAATTAAAAAATGCAT
TATAATGTAA

Translation (189 aa):

MTEYKLVVVGAGGVGKSALTIQLIQNHFVDEYDPTIEDSYRKQVVIDGETCLLDILDTAGQEEYSAMRDQ
YMRTGEGFLCVFAINNTKSFEDIHHYREQIKRVKDSEDVPMVLVGNKCDLPSRTVDTKQAQDLARSYGIP
FIETSAKTRQRVEDAFYTLVREIRQYRLKKISKEEKTPGCVKIKKCIIM

Figure 2.10: *KRAS Gene 4th isoform Nucleotide sequences and Translation. As shown in the figure there are three different Protein that can be altered in the fourth trancript. And the subsequences of these are highlighted in blue color*[14].

[14] http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi?REQUEST=CCDS&DATA=CCDS8703

2.3 KRAS Conserved Regions

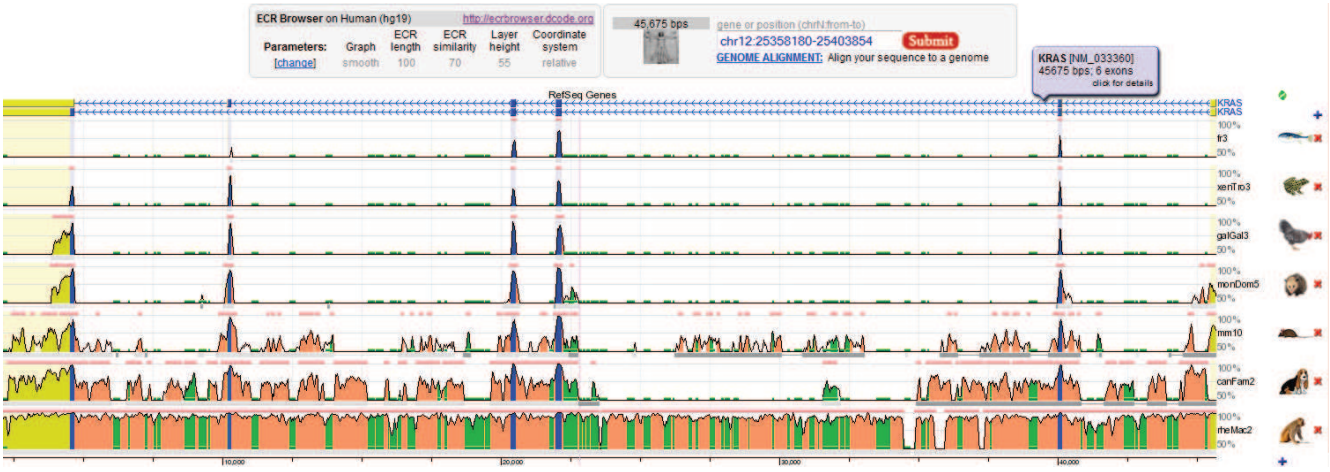


Figure 2.11: *KRAS Conserved Regions by specie kind.*

2.4 KRAS Phylogenetic Tree

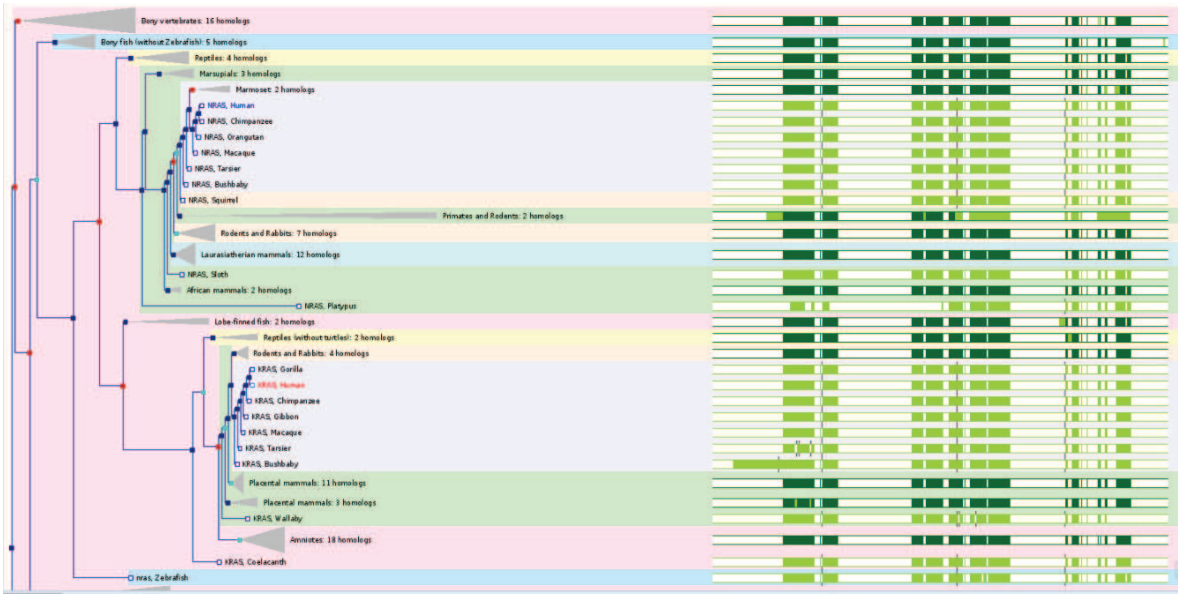


Figure 2.12: *KRAS Phylogenetic Tree Paralogs 1 out of 2.*

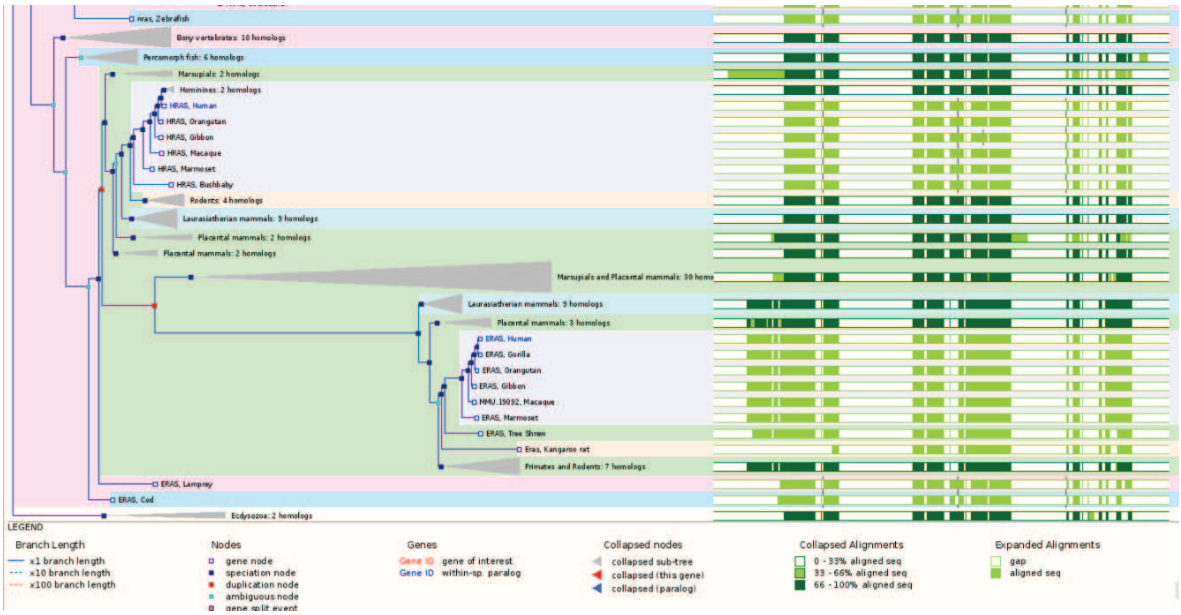


Figure 2.13: *KRAS Phylogenetic Tree Paralogs 2 out of 2.*

CHAPTER 3

Conclusion

Our team which constitutes from Aidonopoulos Orfeas, Begetis Nikolaos and Konstantopoulos Dimitris, was assigned to make a report for the RAS human gene. Because the RAS family constitutes from more than one subfamily genes we decided, given our supervisor's confirmation to investigate and gather information for the KRAS gene.

As we were induced, in every section of our case study we provided some useful information about what we were assigned. At first, we begun with a small introduction giving some definition for our case study. Subsequently, we talked about general information of RAS and KRAS. After that, we continued to separate in parts the KRAS gene and make some research, using well known databases, over the promoters of KRAS. In continuation, we investigated on KRAS isoforms and conserved regions, and with this data we made the KRAS phylogenetic tree.

Then, we provided the most useful information that we found for KRAS proteins and we closed our case study writing about KRAS transcription factors and miRNAs.

Finally, in addition to all the above, we created a miscellaneous chapter were we placed some figures that were more specialized than requested from the assignment.

Bibliography

Bibliography

[1] McGrath JP, Capon DJ, Smith DH, Chen EY, Seeburg PH, Goeddel DV, Levinson AD (1983). "Structure and organization of the human Ki-ras proto-oncogene and a related processed pseudogene". Nature 304 (5926): 501-6. doi:10.1038/304501a0. PMID 6308466.

[2] Popescu NC, Amsbaugh SC, DiPaolo JA, Tronick SR, Aaronson SA, Swan DC (March 1985). "Chromosomal localization of three human ras genes by in situ molecular hybridization". Somat. Cell Mol. Genet. 11 (2): 149-55. doi:10.1007/BF01534703. PMID 3856955.

[3] Kranenburg O (November 2005). "The KRAS oncogene: past, present, and future". Biochim. Biophys. Acta 1756 (2): 81-2. doi:10.1016/j.bbcan.2005.10.001. PMID 16269215.

[4] Neelan J. Marianayagam, Margaret Sunde, Jacqueline M. Matthews, The power of two: protein dimerization in biology, Trends in Biochemical Sciences, Volume 29, Issue 11, November 2004, Pages 618-625, ISSN 0968-0004, 10.1016/j.tibs.2004.09.006. (http://www.sciencedirect.com/science/article/pii/S0968000404002348)