TIU Transactions on Human sciences

# Analysis of an Important Gene and A Brief Study Detect SNPs in The Endometrial Cancer by using UCSC Genome Browser tools as an In-Silico approach

Atreyee Majumder [a], Malavika Bhattacharya [a*]

[a] Department of Biotechnology, Techno India University, EM-4, EM Block, Sector V, Bidhannagar, Kolkata, West Bengal 700091, India.
email: malavika.b@technoindiaeducation.com

## Abstract

There is a genetic variation happened "Single nucleotide polymorphisms", or "SNPs," in human then it's tended to be constitutional variations found in germ cells and other tissues that differ from specific polymorphic regions of Nucleotide position (adenine, thymine, uracil, cytosine, and guanine) in the genetic monitor track sequence. Concession variants, on the other hand, set up 1% of the population wherein the whole genome possesses been duplicated or deleted, replacing one nucleotide in the DNA structure block with another concession variant. Copy number variations (CNVs) occur when a full gene and other significant portion of DNA has been duplicated or deleted. SNPs are generally different from CNVs. We've been able developed the usage of "The UCSC Genome Browser", a publicly available genome cyber-surfer hosted by the University of California, Santa Cruz" (UCSC) for SNP in order to find this variation number. The Bioinformatics department at the UCSC Genome website allows users to set up the Genome Browser Database, browser instruments, download data lines of code, and documentation attestation. The interactive website that provides access to genomic sequence data from many invertebrates, vertebrates, brute species, and key model animals is combined with a sizable collection of aligned annotations that show the methods used to identify SNPs. Using the previously UCSC Genome Browser, a sophisticated online resource for quick and dependable where it show the presentation of any requested genome segment at any scale along with a number of aligned annotations that rapid-fire and dependable representation of any relevant segment of the genome and have together annotations the reflection pathways, researcher scholar has the ability to identify all the SNPs in a gene related to endometrial cancer of http://genome.ucsc.edu has been used as these cross-species homologies, single nucleotide polymorphisms, multiple gene predictions, and cyber-surfer alignments are found here. This cyber-surfer may be used to visualize a multitude of coregistered tracks, including radiation damage data, transposon reprises, multiple gene prognostications, mRNA and expressed sequence label alignments, single nucleotide, sequence-tagged locations, and more. Based on the MySQL database, the widely used Web application UCSC Genome and Table Browser (genome.ucsc.edu) lets you explore a specific variation region of a genome together with several mutation that aligned annotations, or "tracks." A broad variety of data may be annotated, generated, and shown by genes prediction, mRNA and expressed sequence tag alignments (EST), etc., as well as basic single or multiple nucleotide polymorphisms, expression and regulatory data, pairwise and multiple-species alignment BLAST that relates to genomics data, genetic and variance information and more, every one of which is carried out by the UCSC The genome Bioinformatics Group and its external partners. Every individual can add custom data for academic reasons.

**Keywords**: Genome browser and Table browser, UCSC ,Endometrial Cancer genome analysis, in-silico analysis, Bioinformatics, Human Genetic variation, comparative genomics, rs numbers, The Biological Databases, BLAT

## 1. Introduction

Genes are arranged in certain ways to produce SNPs [1]. The precise location of SNPs throughout across the human genome can contain code for gene sequences of data, regardless of whether certain non-coding regions are connected to database applications where disorders like diabetes, cancer, or coronary artery disease have been detected. "UCSC Genome" can find genes associated with complaints.[1] The protein that arises from the degeneracy of these genetic codes is a

consequence of periodic improvements in to the sequence of the amino acids in a gene inside an amino acid sequence that codes, as demonstrated by a "Haplotype Map" in SNP (University of California Santa Cruz Genomics Institute).[2] SNPs detect variances associated with our susceptibility to various kinds of illnesses where our bodies react differently and cause differing degrees of illness.[2] The vast majority SNPs are the result of arranging genes in a certain way.[3] While certain non-coding sections are linked to database applications where diseases like diabetes, cancer, or coronary artery disease have been identified, the position of SNPs in the human genome can include gene coding sequences. UCSC-SNP shows that every now and then, the sequence of amino acids of a gene inside an amino acid sequence that codes changes.[1] the protein that results from these genetic codes' degeneracy. SNPs identify variations in our vulnerability to a variety of diseases where our body responds in various ways and the severity of the sickness varies.[3] Currently under public access, the majority of the human genome has been sequenced, and efforts are underway to fill in the blanks that continues to proceed rapidly.[2] Every 1,000 letters of the code include one SNP—a single letter which might contain a T or a C—which explains why it is called an SNP. Such SNPs were the source of therapies that may demonstrate genetic variances. Due to their location in a region of the genome without an essential purpose, the majority of SNPs in the variation have no effect.[1]

## 2. ORGANIZATION OF UCSC BROWSER:

SNP can be applied in in conjunction via the UCSC Genome Browser provides access online resource that researchers can access and use to study every human chromosome 23 in detail in where a complete base to a Polymorphism in a Single Nucleotide.[4] This tool can be thought of as a multi-powered microscope. For viewing and analyzing the many kinds of data that may be attached to genomic sites, Genome Browser is invaluable.[4] These interpreters include variance, transcription, and the many kinds of nematode genomes particularly useful for multiple-genome of comparative analysis the following primary tools have been created and maintained by the browser team: 1) The Table Browser and Genome 2) BLAT 3) and Hundreds of thousands of series of aligned annotation "tracks" are used in In-SiIico PCR experiments [1]. It contains hundreds of thousands of genes in order of an array of synchronization of the annotation "tracks or path" visible in each additionally each track the section in this UCSC browser. The Genome Browser from UCSC at Table shows that these intergenic regions, or areas between genes, have an impact on SNPs throughout the whole genome.[1] Some of the less-used tools that contain regulatory data, including methylation, transcription factor binding, and disease connections, are Gene Sorter, Genome Graphs, and Data Integrator.[2] 'The University of California illnesses by SNPs relationships' may additionally can be used to track the passing along of illness genes within families and is crucial in determining which SNPs have been identified[3]. Several SNPs might be capable to forecast that personality of someone specific illicit substances and sensitivity comes from the outside influences analogous to pathogens. In the vast majority of instances of specific exons and introns forming. But the SNP is a polymorphism where inside a high linkage disequilibrium (the non-random connection of alleles at two or more loci) area of the genome.[4] where the likelihood of a person's vulnerability to certain diseases and how they react to medications in where the likelihood that a person may get sick from different diseases and response to drugs in SNPs, a useful in whole-genome at SNP association. A well-liked software program that, "The Santa-Cruz (UCSC) Genome Browser," an online application, allows users to quickly see a desired genome segment in any level of detail.[1] It also contains ordered annotations produced by the UCSC Genome Bioinformatics Group and additional collaborators, such as gene and such as UCSC's strong engineering department, particularly in the fields of EE, CS, Biosciences, Physics/Astronomy, and EE.[4] A consistent graphical user interface for genomic sequences, features, and annotations is offered by UCSC across a broad range of the wide spectrum of organisms. Covering seven different nematode genome types, including Caenorhabditis elegans, From Yeast to Humans is very helpful for multiple-genome comparison study. The offered resources have been utilized, and the sequences involving identification and evaluation may be made easier by the visual assistance. When using the UCSC Genome Browser, provided materials have been used, and the visual aids may facilitate the identification and assessment procedures. The UCSC Genome Browser allows users to browse high-quality, researcher-focused whole-genome annotation data as much as is practical and permit scientists to do research at UC Santa Cruz while using their expertise to examine their own from analysis. The provided tools have been used and the visual aids can be facilitating the sequences that feature detection and examination. This Genome Browser used from the perspective like a researcher where Its UCSC Genome Browser's fundamental concept presents high-quality whole-genome annotation data. Provide as much as possible and enable researchers to take advantage of their knowledge to independently evaluate data browse at UCSC.[1]

## 3. MATERIAL AND METHODS:

UCSC Genome Browser

1) The Human Feb. 2009 (GRCh37/hg19) Assembly.
2) The Human Dec. 2013 (GRCh38/hg38) Assembly.

### 3.1. **Inclusion Criteria**:

A human reference genome GRCh19 was released on 27 February 2009, A computerized nucleic acid sequence database which scientists have put together to serve as a hypothetical illustration of the gene set of a single organism of a species also there are human a citations genome GRCh38 and the genome data was made accessible on December 17, 2013, through the Genome Research Collaboration.

### 3.2. Exclusion Criteria:

The Demonstration of Finding All SNP in Using the UCSC Genome Browser to Look for a Gene of PTEN, TP53, MDM2, MYC, BRAF, CYP1 A1, CYP19A1, SMAD1, SMAD2. PTEN Gene (75% mutation) important Gene for Endometrial Cancer.

### 3.3. Examination of the browser used by UCSC:

This is, UCSC Genome and Table Browser may be used to find every single SNP in a gene, as demonstrated in the course. Finding single nucleotide polymorphisms in a gene is another application for it. We chose the "pack," and then we hit the [refresh] button. Using various resolutions, the Genome Browser has been displayed the measurement ranges (for human chr10) according to the various levels of detail. Those resizable from a single base pair to constantly swift and intuitive are made possible by this 250-million-fold difference in knowledge. The interface was designed to make surfing easier for users. Numerous functions may be found in a browser. To discover each and every SNP in a gene applies in the instruction of the demonstrations that to discover UCSC. After choosing the "pack," we pressed the [refresh] button. The variety of scales (human) chr10, in distinguished stages the particulars processes have been shown at in the different scales, which vary depending on the Genomics Browser's resolution. Always those convertibles from just one pair of bases to due to that 250-million-fold had experience fast and easy to understand. The screen has been developed to facilitate user browsing. Browser has an extensive list of distinctive characteristics.


Fig: 1 The PTEN gene is located here; choose it from the list, and clicking the [go] button will take us there. This gene is meant to sporadically illustrate that. With respect to a certain gene, the canonical gene interacts. Several data sets' default data follow the default gene.

### 4.. PROCEDURE: [Genome and Table Browser]

1. At first, we have to do, The Genome Cyber-surfer database tool where Initially, we must perform, The Genome Cyber-surfer database tool is located at genome.ucsc.edu and includes both supervisory and nonsupervisory sections at genome.ucsc.edu. There are 2 links in the upper left side of the internet address there are the fact that go to the homepage where users must select which organism to utilize. Clicking the "click here to reset" button takes us back to the home page, whereas clicking the [submit] button opens the graphical viewer in the browser. To search the "HUMAN" default assembly, hg19 and hg38, we must do so. Next, we look at the UCSC Genes using the default data, where several kinds of data tracks might simplify the view by concealing all the data tracks.

2. The next action we have to take is: The "PTEN gene [E.g.: Endometrial Cancer]" is the default gene in this case. We choose it Clicking the "click here to reset" button takes us back to the home page, whereas clicking the [submit] button opens the graphical viewer in the browser is the rationale for this. they are medically significant. They are

identified as having originated from a locus-specific database.

3. The next action we need to do is to activate three of the single nucleotide's polymorphism tracks, which are located at the bottom of the screen. After that we have to do, the Common SNPs from dbsnp version 138 and All SNPs track are the same version and all the Flagged SNPs track all are clinically relevant or potentially clinically relevant by polymorphisms. After each track has had time to "pack," the "fresh" button is activated, allowing us to view a sizable SNPs together with their "rs" values on the screen. Amino acid modifications are shown by the colours red, blue, and black, respectively, splice sites or untranslated regions, and intronic regions. When the mouse button on the right is pressed, the track will collapse into a single row where the "dense" instruction is located. From there, we can see that they are evenly spaced across the gene. The most clinically significant SNPs are tracked by the Common SNPs track, whereas the All-SNPs track is quite denser.

4. A few green ones that fall within the code area. SNPs with unaltered amino acid composition. It is going to them as well as down to "dense." They will also collapse down to a state of "denseness." It's important to read the track description to determine the precise nature of this SNP track as the gene may or may not be clinically significant That's underneath the button on the left side where we will learn that they are clinically relevant if they have a particular flag from NCBI. They're designated as having been contributed by a locus-specific not disease-causing database. The various types of SNP tracks are there and the All SNPs are the generally overarching track by the other two is subsets of the all-SNPs track.

5. We will go to the Table cyber surfer and use one of these tracks to find the SNPs in the gene. We will simply get the SNPs in this particular gene here and the genome browser remembers where the previous screen, so that the coordinates are preloaded into this area but there are necessary work has to be click in the [position] button. To find the SNP tracks we will elect the variation group to match their position in the track controls beneath the Genome Browser graphical display. The dbsnp edition 138 provides the common SNPs, which all of which can be tracked utilizing the same version, and all of that are flagged as theoretically or clinically significant due to polymorphisms.

6. That's underneath the button on the left side where we will learn that they are clinically relevant if they have a particular flag from NCBI. underneath the button on the left, where it says if they're carrying a particular NCBI flag, which signals whether or not which does not trigger illness. There are several distinct types of SNP tracks, with the All-SNPs track serving as the main one while the other two are subsets of that. To locate the SNPs in the gene, we'll visit the table created by the World Wide World surfer and utilize one of these tracks. We'll.

7. In this case, we have to use the common SNP track to generate a variety of SNP lists. Under the columns, we want to save for the browser's graphical SNP display area. By clicking the [Get Output] button, we are able to pick which features to download together with the file containing the chosen polymorphism portion. UCSC Browser page where, should we be interested, we may select the genetic equivalents. We must choose the various names and [get output] button numbers at that precise moment. Certain SNP-specific tracks are comparable to the "function" and "allele frequency" found in this list.


Fig: 2 View by After table has turning the UCSC Genes track to "pack" and pressing the [refresh] button, we will observe the screen with all data tracks hidden. IN The Common SNPs from dbsnp version 138 are available here. To access the SNPs for each gene segment, use the little left arrow on the side of the window

8. It must indicate that we have a list that is required for us to fully understand the sequence variants (SNPs) that are being utilized by the Genome Browser or Cyber the surfing community. All we have to do is choose the "rs" numbers from the list, leaving the header data remain. The list's contents must then be copied back into the Table Browser or Internet browser and used to create a custom track. Use the drop-down list by tapping the [Back] the button symbol to go return to the Table Browser's main screen. We will copy the URL and use this [paste list] button to load our SNPs into the Table Browser on the website. By selecting "the output format" and pushing the [submit] button, the track sequence of SNPs can then be imported and pushing the [submit] button, the track sequence of SNPs can then be imported button gives us the ability to select the "custom track" for our output.

9. After selecting the [get output] option, a screen appeared on which we had to mark up our track record. For the shortened label, we have to use "common SNPS in PTEN" and for the extended label, "common SNPs in PTEN gene region." By choosing [get output], we were able to view the customized track. They will be shown to us on the genome's browser. To have an improved understanding of the region in question, we may focus in and examine the frequently occurring SNPs that make up the track by bringing them up into placement such that they are next to the newly manufactured track and this exon's select. The corresponding tracks' rs numbers and tick marks, which can collapse them into "dense".

10. It might be demonstrated that there are similarities between our track and the original Common SNPs track. Now that each SNP in 1 has a factor zoom out of 100, we must look at the query's limiting range. The PTEN gene is situated above the novel gene reflection tracks that we have identified. This is the Genome Browser, where DNA sequences are BLAT-aligned with reference genomic assemblies to display genes, gene structures, and gene annotation tracks. Furthermore, Custom Tracks is displaying the data in addition to already stored browser data. The Table Browser may also download intersections and joins across data sets and do bulk data manipulation.

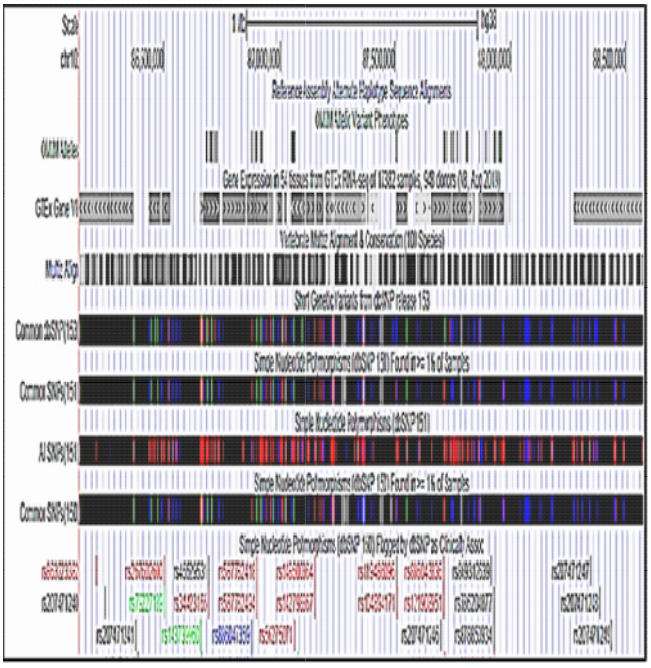


Fig: 3 The Genome Browser at http://genome.ucsc.edu. has two links in the upper left corner. that1] navigate to the home page, where we can choose the organism, we want to use. 2] The human genome default assembly hg38's "click here to reset" option takes us to the default the "submit" button launches the graphical viewer in the browser.

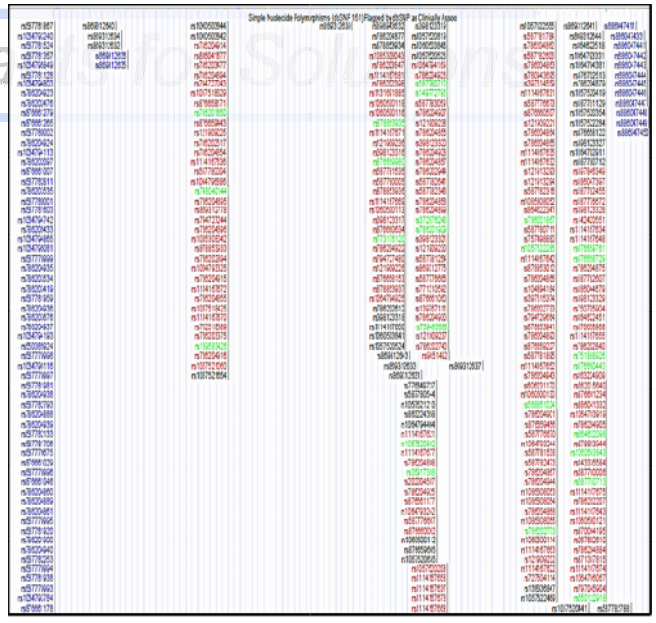


Fig:4 Upon clicking the "fresh" button, every single track starts to "pack," resulting in an enormous number of SNPs along with their rs numbers showing up on the screen. The blue dots in results 1–2 are splice sites or untranslated parts; black dots in outcomes 3–6 is

intronic regions; and red dots in results 4–5 are differences in amino acids. Using the right mouse button, the track may be compacted to a single row where it is "dense". They are uniformly placed across the gene, as you can see. There are a small number of green rs SNPs, or coding region SNPs, whose amino acid composition is unchanged.

## 4. RESULTS:

**Red:** ones are amino acid changes: Coding - Non-Synonymous:

1. The variation in the amino acid sequence when a codon has at least one altered nucleotide, causing the stop codon to halt prematurely and the polypeptide to be truncated.

2. Missense variant: It is a sequence variation that results in a different amino acid sequence with one or more base modifications while maintaining the same length.

3. Stop lost: An extended transcriptome is produced by a sequence variation in which the terminator codon (stop) includes at least one altered nucleotide.

4. Frame shift variant: One variation of the procedure involves selecting a destination by touching the [go] button and selecting through the list that appears. The translational reading frame is disrupted by a gene's "is demonstrated that periodically" as well as the original ones because of the total number of nucleotides that have been inserted or deleted is not a multiple of three.

5. Inframe indel: a coding sequence variant when the transcript's frame remains identical but the CDS's length fluctuates.

6. Splice site:

1. Splice acceptor variant: A "splice acceptor variant" is a splice variation that modifications the two-base regions at an intron's 3' end.

2. Splice donor variant: Splice contributor polymorphisms are two-base pair variants which alter an intron's 5-' end.

**Black:** Intronic areas are the black ones: Locus: A sequence variation that is found at a gene's 3' end.

**Blue:** Splice sites or untranslated regions are represented by the blue ones: Non-coding RNAs (ncRNA): Version of the transcript without coding: Variant of an RNA gene that does not code for anything.

**Untranslated:**

1. 3 prime UTR variant: A 3' untranslated region (variation in UTR format).

2. 5 prime UTR variant: A 5' untranslated region (variation in UTR format).

**Green:** The coding area SNPs with green backgrounds are those that do not alter the amino acid sequence.


*https://genome.ucsc.edu/cgibin/hgTables?hgsid=1374056407_HH8SoT6lKCMM0IaN1C1lhYihC8A8&hgta_ctName=Common+snp+in+pten&hgta_ctDesc=C*

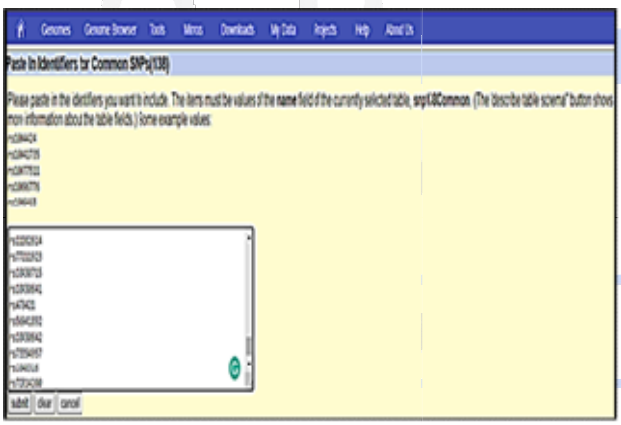
*Fig 5(a): The "Custom Track" page, from which we obtain the "rs" value using the "UCSC GENOME BIOINFORMATICS"*

### 6. (a) The Interpretation of UCSC Browsers:

**• UCSC Genome Browser on Human Dec.2013 (GRCh38/hg38) Assembly: -**

In accordance with dbSNP Releases 153, and the short genetic variant rs12573787 denotes a common (1000 Genomes Phase 3 MAF >= 1%) variant. On December 17, 2013, the Genome Reference Consortium published GRCh38, the whole reference genome for humans. Single-nucleotide variants (SNVs), small deletions, insertions, and complicated deletion/insertions (indels)

are a few instances of short genetic differences (up to about 50 base pairs) from dbSNP version 153 in this track compared to the reference assembled genome. Certain variants in dbSNP are known to be harmful, even if the bulk of variations are rare and not true polymorphisms. Exact matches between distinct variants (RefSNP clusters with rs# ids) and chromosomal sites exceed 702 million. On December 17, 2013, the Genome Reference Consortium released the human reference genome GRCh38. This first version had over 150,000 gaps includes substitute haplotypes and repair patch sequences. connected to over 702 million hg38 (GRCh38) genomic locations, of which more than 667 million are for hg38 (GRCh38). The subsequent screen build had about 250 gaps, while the initial version had almost 150,000 gaps. By using ultra-long reads, 12 gaps in the GRCh38 reference assembly have been closed by nanopore sequencing. Over 100 gaps are reduced in the GRCh38 assembly. There are four subtracts of variations on the track:




https://genome.ucsc.edu/cgibin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&p

Fig 5(b): For the custom tracking table browser output, common SNP values must be copied and pasted.

• All dbSNP (153): The full {track} collection (702 million for hg38) means the entire set of the track is in the 702 million for hg38)

• Common dbSNP (153): There are about 15 million variations in the 1000 Genomes Phase 3 dataset with a Minor Allele Frequency (MAF) of at least 1% (0.01); these variants are deleted in addition to those in the various subsets. There are approximately 15 million variants with a Minor The term "allele mean.

• ClinVar dbSNP (153): ClinVar lists 455,000 variations, including pathogenic and benign (as well as unclear) variants.

• Variants in the multiple dbSNP (153): multiple chromosomes variants have been mapped for example: chr10 and chr17, of the variants if this is a variation, is it only a distinction between two sequences that are repeated? It raises the possibility that an uncommon variation, which is not part of the set mapped to the sequence in a primary chromosomal that alternates the haplotype or fixes the patch sequence allocated to that each chromosome might arise in both the X and Y pseudo-autosomal regions (PARs). Coordinate ranges to which dbSNP mapped as a variation but with incompatible genomic coordinates are highlighted by the fifth subtract. As an illustration, note that several coordinate ranges were given for characterizing distinct alleles; the sixth subtract highlights this. For example, multiple location ranges have been given for different alleles.

• Map Err (153): Over 149,000 gene have been mappings of 86,000 in between unique rs IDs for hg38 were found to have been useful. Because SNPs can differentiate between different sequencing gatherings, snips can be used to mapping disease loci and identify genes that predispose individuals to certain diseases. SNPs can be utilized as biomarkers for complicated characteristics connected to genetic diseases attributable to the result of academics' development of high-density SNP arrays and SNP maps together.



**UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly:**

The expression "reference genome" means to a digital sequence of nucleic acids database that's which is created by sequencing the DNA of multiple individual donors. Note that while reference genomes have been assembled by scientists as an illustrative representation of the human genome's gene set, they do not accurately represent the collection of genes of one particular organism. A haploid mosaic chromosome with distinct DNA sequences from each donor is instead provided by a reference. It is possible to shift the newly generated track up to be next to the Common SNPs track. reference genome (also known as a reference assembly) is a digital nucleic acid sequence database, assembled from the sequencing of DNA from a number of individual donors, reference genomes do not correctly depict the collection of genes from any one creature, even if scientists have gathered these genes in order so they can serve as an illustration for the set of genes in humans. Instead, a reference provides a

haploid mosaic chromosome of different and we'll select this exon over here also zoom into the region and get a closer look. A match between newly formed track and the original Common SNPs track disease loci may be mapped and genes associated with vulnerability to illness can be identified among individuals by comparing our newly developed track to the original Common SNPs track from which it originated and by using SNP-based genetic linkage analysis. The use of SNPs as markers has been made possible by a combination of a high level of SNP arrays and SNP maps for complex trait genetic diseases. SNPs that may be used to predict an individual's response to certain medications, vulnerability to environmental variables like pollutants, and likelihood of contracting specific diseases have been discovered by researchers. In addition to being helpful for tracking the inheritance of disease genes among families, SNPs are essential for determining an individual's susceptibility to different illnesses and how they will react to medications.

**6. (b) Interpretation of Single Nucleotide Polymorphism in UCSC Browser of Endometrial Cancer Gene:**

SNP detection for PTEN gene: PTEN gene detection of Endometrial Cancer in this Cyber-surfer is included with Human Genome Project and associated with UCSC genes have the potential implication in medicinal area.

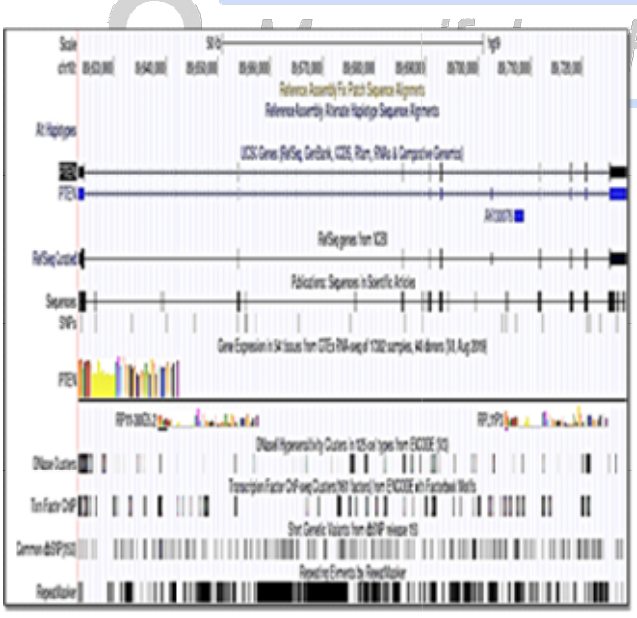


Fig: 5:(c) In UCSC Genes: There are RefSeq, GenBank, CCDS, Rfam, tRNA &comparative genomics.

**Single-nucleotide polymorphisms (SNPs) of dbsnp:**

DNA phosphatase 1 SNPs (single nucleotide polymorphisms) (TEP1). Single nucleotide changes in genes or related genetic components that impact the expression of SNPs in genes are known as tumor suppressor genes. The National Center for Biotechnology Information (NCBI) and the National Human Genome Research Institute (NHGRI) collaborated to build and manage the Single Nucleotide Polymorphism Database (dbsnp), a free genetic database repository for genetic diversity within and across species. Microsatellites, small-scale insertions and deletions, publishing, population frequency, molecular consequence, and genomic and RefSeq mapping data for both common variations and clinical mutations are all available for human single nucleotide variants. The most common type genetic variation can be observed in the DENV database strand. This tumor suppressor gene, which includes single nucleotide alterations, is also known as epithelial cell enriched PTEN, which is TGF-β regulated, and mutated in multiple advanced malignancies 1 (MMAC1).

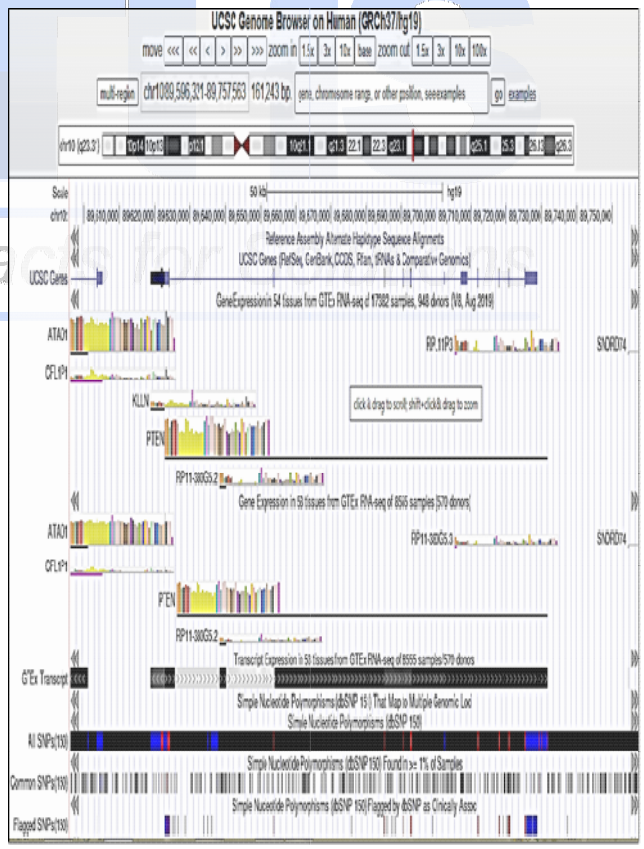



Fig:5: (d) Assembly of the UCSC Genome Browser Human (GRCh37/hg19), February 2009[1] The identification included DNase clusters, repetitive elements, transcription factors identified by ChIP-seq,

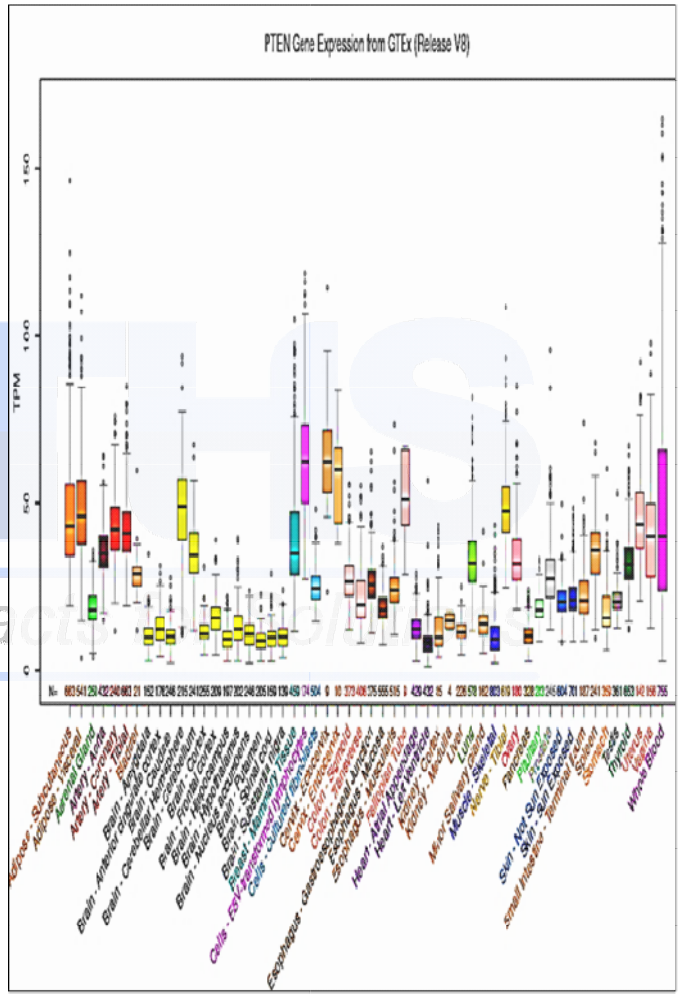
sequences SNPs of gene expression in 54 tissues using GTEXRNA-seq data, PTEN gene SNP, and short genetic variations from dbsnp.

When compared to its cytoplasmic non-ubiquitinated form in gene expression, which controls the appropriate neuron positioning, dendritic development, and synapse formation, the AKT-mTOR signaling pathway's cytoplasmic non-ubiquitinated form induces less tumor suppressive ability and has a higher potential for apoptosis. Cancer Through their interaction with different environmental factors and genetic characteristics of the host that increase cancer risk, "Single Nucleotide Polymorphisms (SNPs)" have a significant contribution to cancer susceptibility. This contribution could aid in the development of therapeutic applications as well as the overall management of this deadly disease. Complicated interplay between genetic and environmental variables lead to polygenic cancer. SNP mutations thus set off all of those mechanisms


Fig:5:(e) Zooming in on Chr10[PTEN], the scale shrank from 50 kb to 1 kb. These DBPs, or simple nucleotide polymorphisms mapping to multiple genomic loci, were detected in 1% of samples as flagged SNPs, all SNPs, and common SNPs.

 Gene expression study has been analysed by some of different sequence variation. On this 3 some are: -

a) Human Gene of PTEN gene (uc001kfb.3) (NM_000314) RefSeq NCBI: -
b) Human Gene of PTEN gene AK130076 (uc001kfc.1) (NM_000314) RefSeq NCBI: -
c) Human Gene PTEN gene (uc 021pvw.1) - RefSeq NCBI (NM_000314):




https://genome.ucsc.edu/cgibin/hgc?db=hg19&c=chr10&l=89623193&r=89728532&o=89623194&t=89731687&g=gtexGeneV8&i=PTEN

Fig 6: The Reference SNP Cluster ID A unique designation known as the "rs" number has been employed by researchers and databases to designate a particular SNP (single nucleotide polymorphism) (rs = "rs" + number). It stands for Reference SNP cluster ID

and is the naming mechanism most SNPs utilize. To view Single Nucleotide Polymorphisms (SNPs) from Clone Overlaps and Select "Single Nucleotide Polymorphisms (SNPs) from Clone Overlaps" or "Single Nucleotide Polymorphisms (SNPs) from Random Repeats" from the Browser information page. This will show each SNP's unique rs number that is present in the sequence that was supplied.

**a) Human Gene of PTEN gene (uc001kfb.3) (NM_000314) RefSeq NCBI: -**

PTEN, the homologue of Homo sapiens phosphatase and tensin, is located at position [chr10: 89,706,925-89,708,694], and its RNA level is ascertained by mRNA and a cDNA clone Hg19. Measurements: 1,770. [In such scenario, total exon count is 1Strand: +]. It has been shown that those genes function as tumor suppressors and are frequently altered in a considerable proportion of cancers. This gene encodes a phosphatedylinositol-3, 4, 5-trisphosphate 3-phosphatase protein. Its catalytic domain, which operates similarly to the dual specificity protein of tyrosine phosphatases, preferentially dephosphorylates phosphoinositide substrate. This domain is similar to Tensins. The protein possesses both a catalytic region and a tensin-like domain. that, by inhibiting the intracellular concentrations of phosphatidyl inositol-3,4,5-trisphosphate in cells and managing the AKT/PKB signalling pathways, inhibits the formation of tumors. The non-canonical (CUG) upstream initiation site produces a larger isoform that is expected to be preferentially associated with the inner membrane of the mitochondria. Starts translation with a leucine and may have an impact on the mitochondria's energy metabolism.

**b) Human Gene of PTEN gene AK130076 (uc001kfc.1) (NM_000314) RefSeq NCBI: -**

The pseudogene for this gene is found on chromosome 9. Several transcripts' variants encoding distinct isoforms are generated via splicing and various translations of start codons. The gene location hg19, chr10:89,623,195-89,728,532, has a transcript (including UTRs) with a size of 105,338 and a total exon count of 9. Strand: + Coding the area at [chr10:89,624,227-89,725,229] position on hg19 Dimensions: 101,003; Coding the exon count gene: 9. Acting as a dual-specificity protein phosphatase, this tumor suppressor dephosphorylates proteins that are phosphorylated on tyrosine, serine, and threonine. This non-phosphorylated variant works with AIP1 to inhibit the activation of AKT1. Dephosphorylates tyrosine-phosphorylated focal adhesion kinase, which in turn suppresses the development of focal adhesions that function as important modulators, integrin-mediated cell spreading, and cell migration. In the region of chromosome 9 and 10. A faux gene that encodes this specific gene was recently identified. Splicing and different translations of start codons result in multiple transcript variations that express different isoforms. The gene located at hg19, chr10:89,623,195-89,728,532 has a transcript (including UTRs) with an average size of 105,338 and a total exon count of 9. strand + Coding the hg19 region located at [Chromosome 10:89,624,227–89,725,229] Dimension Size: 101,003; gene coding exon count: 9. This tumor suppressor functions as a dual-specificity protein phosphatase, dephosphorylating proteins that are phosphorylated on tyrosine, serine, and threonine. This non-phosphorylated form inhibits AKT1
activation by interacting with AIP1. Dephosphorylates focal adhesion kinase that has been phosphorylated by tyrosine, which inhibits the growth of focal adhesions—which are crucial modulators—integrin-mediated cell spreading, and cell migration.

**c) Human Gene PTEN gene (uc 021pvw.1) RefSeq NCBI (NM_000314): -**

Lipid phosphatase eliminates the amount that is needed phosphate from the inositol ring at location D3. With substrate order in vitro, that the following are the phospholipids: 1) inositol 1, 3, 4, 5-tetrakinophosphate; 2) phospholipid 3,4-diphosphate; and 3) phospholipid 3-phosphate. The PtdIns (3,4,5) P3 > PtdIns (3,4) P2 > PtdIns3P > Ins (1,3,4,5) P4. Activity for an enzyme called lipid phosphatase. Lipid phosphatase enzyme activity plays a crucial function because of its potential to inhibit tumors. Lipid phosphatase, which in vitro has the following substrate order, breaks down the phosphate in the D3 position of the inositol ring: The points PtdIns (3,4,5) P3 > PtdIns (3,4) P2 > PtdIns3P > Ins (1,3,4,5) P4. Phosphatidylinositol 3, 4, 5-trisphosphate, phosphatidylinositol 3, 4-Di phosphate, and inositol 1, 3, 4, 5-tetrakinophosphate all depend on this type of enzyme to perform their tumor suppressor functions. Mapped genes for PI3K-AKT/PKB through the dephosphorylation of phosphoinositide, the signalling pathway is hostile to the mechanisms that control the progression of the in-vivo and in silico phases of the cell phase and the resilience of the cell. RP11380G5.2 (V8, Aug 2019) includes the Gene Expression research in PTEN Using "GTEx RNA-seq" in 54 tissues including 17382 samples and 948 donors. ENSG00000171862. 10 is the Ensembl project gene ID; gene class: coding; GENCODE biotype: protein coding; Cells containing the greatest median expression was observed in EBV-transformed tissue

cells, having 62.44 TPM; score: 498; total median expression: 1401.15 TPM; hg19 chr10:89623195-89731687; position in the genome.

**The NIH Genotype-Tissue Expression (GTEx): [The GTEx (Genotype- Tissue Expression):**

(eQTLs) mapping can clarify the genetic factor causing changed gene expression.] The purpose of this program is to build a sample and data repository with the objective of encouraging investigations into the link between variation in genetic makeup and the regulation of genes that express in different human tissues. These RNA-seq data (V8, August 2019) are included in the final data releases for GTEx. The median levels of gene expression in fifty-two tissues and two cell lineages are shown in this track. Gene-Tissue Expression, or GTEx, initiative linked hereditary variations to that affect the genes in the human body's engaged and quiescent tissues and organs. This suggests that genetic variations were discovered by the experiment. DNA polymorphisms that genes may display and transmit are known as gene expression qualitative and quantitative particularity loci, or eQTLs. By mapping these loci, the genetic component underlying variations in gene expression may be identified.

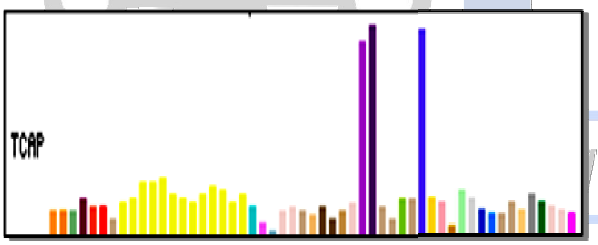



Fig: 6. The amount of expression for each gene can be seen in the two different Full and Pack display modes as a colored bar graph. The height of every bar indicates the tissue's median expression level across all samples, while the bar's colour indicates the tissue. Tissue colours have been allocated in accordance with the GTEx Consortium publishing guidelines.

• H3K27Ac Mark (ENCODE Regulation): - A number of tracks provide information which is important to the ENCODE investigation's transcription regulation. Transcript levels from a variety of distinctive cell types, determined by polyadenylated RNA sequencing, are presented on the transcription track. The H3K27Ac tracks display changes to histone proteins, which might be a sign of enhancer regulatory activity.

• DNase Clusters and Txn Factor ChIP: - Areas that have been investigated or tested in an assortment of cell types are shown on the DNase Clusters track. In several distinct kinds of cells, chromatin is especially susceptible to being broken by the DNase enzymes. Regulatory areas are typically shown to have DNase sensitivities, with promoters being more vulnerable to the enzyme. Measurements using the method of chromatin immunoprecipitation with transcription factor-specific antibodies and sequencing the precipitated DNA (ChIP-seq) can show that the Txn Factor ChIP is bound. ChIP-seq identifies the DNA positions where transcription factors—proteins that control gene transcription—have been found. These tracks can complement each other and together offer a plethora of information about regulatory DNA. While the histone marks provide a wealth of knowledge, the resolution of them is restricted to approximately 200 bases, and they do not provide a lot of details on "Reference SNP Cluster ID" ("rs number") values to aid in the use of SNP by The DNase hypersensitive test provides greater accuracy at the DNA level and is only one experiment that might be run on a variety of cells. The information offered is restricted to a small set of transcription variables across insufficient cell lineages. While each test has benefits and limitations of its own, the relative independence of the previously listed assays increases confidence when monitoring operations across several tracks. Every gene transcription factor has to be tested independently. additional transcription information and histone marks found in ENCODE recordings. Histone mark displays, displaying data from a transparent overlay technique many single tracks—and transcriptional information are used in this integrated super-track. Every cell line in this track has a distinct color that is designed to be rather bright and saturated in order to go well with the transparent overlay. Histone markers and transcription information in ENCODE recordings.
Transcriptional data a combination of histone mark displays, displaying data from several cell lineages on a single track using a transparent overlay technique, are used in this integrated Super-track. Each cell line in this track has a distinct colour that is designed to be rather bright and saturated in order to go well with the transparent overlay.

• The ENCODE Regulation Tracks: - Unlike previous versions of the ENCODE Regulation, the outer tracks do not use colors to store different genetic information tracks derived from DNA sequences. Transcription Factor ChIP tracks include information on many cell lines for which a color standard is insufficient. Instead, these tracks show gray rectangles whose darkness is connected with the greatest value

seen in all adjacent cell lines. Clicking on the item brings up a full page with the values for each cell line that has been evaluated and the raw data for ENCODE Regulation tracks. About the mRNA transcripts' experimental behavior as revealed by Serial Analysis of Gene Expression (SAGE). A "SAGE/UniGene track," which shows information about One of the tracks that combines the SAGE data is the degree of transcription of various UniGene sequences of the alignment groups. The SAGE Map project at NCI. (http://www.ncbi.nlm.nih.gov/UniGene/) from the NCBI's SAGE Map project.

• SAGE Map project: Since a component of the "Serial Analysis of Gene Expression" (SAGE) project, mRNA has been extracted from a cDNA via a method called restriction enzyme digestion, which separating short genetic segments expressed by the ligation or additional labels. Bacterial genes organize were subjected to sequencing in order to offer measurable and comprehensive genetic expression the profiles.


https://genome.ucsc.edu/cgibin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&p

Fig7: ChIP-seq Histone Modifications from ENCODE: These tracks show chromatin state maps of the various ENCODE cell types. The nucleosome is made up of many similar proteins called histones.

Histone families H1, H2A, H2B, H3, H4, and H5 are among the six families. Every nucleosome core has been produced by two H2A-H2B dimers and an H3-H4 tetramer, and H1 and H5 bind both DNA and the nucleosome. There are some specific amino acids, some histone proteins undergo post-transcriptional modifications (methylation, acetylation). The primary purpose of the histone modifications, also known as marks, is to regulate how the DNA winds around the nucleosome. This in turn modifies chromatin accessibility, which in turn affects gene expression. The consequences of histone marks are different. Certain histone marks are linked to enhancers, transcribed areas, silent regions, and promoters. Regions can be identified attributable to the ChIP-seq method's interaction of DNA with certain histone marks. This technique uses formaldehyde inside of cells to produce cross-linking histones and other proteins linked to DNA to the genetic sequence. Following extraction, the cross-linked chromatin is mechanically sheared and immunoprecipitated using certain antibodies. The impact has been chosen by the antibody from the capturing DNA segments that are attached to the protein.

• 100 Vertebrates Conservation by PhastCons and Phylop: - Base-wise conservation of 100 vertebrates using the PhastCons software, which measures evolutionary conservation through numerous evolutionary genetic alignments, and Phylop (phyloP100way Overall), which evaluates evolutionary conservation utilizing individual P-values. The PTEN gene is located where chr10 is, and the numerical range is 89,623,194-89,728,532. The total bases that are accessible are 105, 339, Data on: 99.9734% coverage of 105,311 bases in the SNP. The data values cover a high of 9.431 and a low of -5.777 on a 25-bin histogram based on 105311 values. Using two algorithms Two components of the PHAST package, (1) PhastCons and (2) phyloP, and the track may display the evolutionary conservation for all 100 vertebrate species and measurements, as well as for all species of multiple alignments. Multiple alignments were produced utilizing the "MULTIZ" and other tools in the comparative genomics alignment pipeline established by UC Santa Cruz and Penn State Bioinformatics. Nets ("syntenic") and chains ("align able") are the fundamental building components of the "PHAST/Multiz". For a description of the entire alignment procedure, the written description of the Chain/Net tracks of the • This track also displays preserved components that "PhastCons" discovered to be conserved elements.

• **PhastCons & PhyloP:**

Applying a hidden Markov model, PhastCons methodology utilizing multiple alignments to determine the probability that each nucleotide is a component of a conserved element. When evaluating

conservation at individual column contrast, flanking columns take account of each alignment column's phyloP separately, ignoring neighboring tracks. The phyloP diagrams plot appear not as fluid as the PhastCons strategies due to the fact that is more "texture" at every single site. While the two SNP track approaches have various advantages and disadvantages, PhastCons is more effective at detecting conserved elements since it is dependent on "runs" of conserved sites. Nonetheless, the PhyloP is more outfitted for assessing indications of selection at expected nucleotides or groups of nucleotides (such as the initial positions of miRNA target regions or the third of the translational positions). exponential acceleration (faster-than-expected evolution) There's an essential difference between slow conservation (slower than displayed evolution) and neutral drift. The regions in the phyloP plots that are expected to be maintained receive favorable scores and appear in blue; sites that are projected to change quickly are given negative values and are displayed in red. The -log p-values assuming an unaffected hypothesis with a null test are represented by the absolute values of the scores. While the selection spectrum will be between 0 and 1, the PhastCons scores, conversely, indicate the probability of negative selection. The same circumstances were used for both phyloP and PhastCons, which consider misaligned nucleotides and alignments gap constitute has been missing data. The masked and aligned score of every genome assembly has successfully replicated by UCSC. Hub genome browsers example: -Missing sequence Whenever zooming through, modifications of yellow on the track displays represent significant putting it together and displayed by at base level. The minuscule insertions and deletions (indels) of the genome which constitute single nucleotide polymorphisms, for example (dbSNP 151) that are flagged by dbSNP as Clinically Associated dbSNP build 151 rs587781867 (dbSNP: rs587781867) Position: chr10:89623193-89623197 Band: 10q23.31; Genomic Size: 5 Strand: +, Observed allele: -/GCCCT; Reference allele: GCCCT, Function– upstream gene variant; Weight: 1. Individually Simple Nucleotide Polymorphisms from build 151 of the dbSNP has been used and that are depend This subset contains only SNPs which have been recognized by dbSNP as clinically connected, customized to a single location throughout the whole assembly of the reference genome, and not been reported containing an uncommon allele frequency of at least 1%. Only a medical geneticist with training should read this track in order to determine the relevance of any given variant based on all available information. The frequency subset of all SNPs likely contains several SNPs with actual minor frequencies of alleles in is 1% or higher. For instance, because some variations are included in a locus-specific database (LSDB), they are included in this track of distinct SNP.A proficient specialist has thoroughly reviewed all of the pertinent content. The remaining similar tracks are not believed to be carcinogenic. Therefore, a variant's placement in this track does not always imply danger, nor does its mention in OMIM. [available from ftp.ncbi.nlm.nih.gov/SNP.] The following tracks:

• Common SNPs (151) - Singleton allele frequency (MAF) SNPs that correspond to the collection of reference more than once. Single-copy reference assembly (RCA) SNPs with the frequency of a minor allele (MAF) higher than 1% were found.

• Flagged SNPs (151) - The research of Singlet mapping to reference assembly SNPs having a frequency of minor alleles (MAF) of less than 1% (or unknown) has been designated in dbSNP as "clinically associated"—not certainly a risk allele. SNPs classified in dbSNP as "clinically associated"—not necessarily a risk allele—are those which map to the reference assembly just once and have a minor allele frequency (MAF) of less than 1% (or unknown).

• Multiple SNPs (151) - SNPs which has been correspond from of a particular spot to another mapping on reference assembly has been shown in the track. SNPs that map from one location to another on a reference assembly have been displayed in the track.

• All SNPs (151) - every single SNP from the reference assembly to the dbSNP mapping has been shown in the track. Every SNP from the reference assembly to the dbSNP mapping has been displayed in the track.

**6.(c) Interpretation of SNP that Mutated Effect in PTEN Human Gene with the UCSC Browser:**

Mutations in some advanced cancers affect phosphoinositol 3, 4, 5-trisphosphate 3-phosphatase, the homolog of phosphatase and tensin, and a dual-specificity protein phosphatase PTEN. 1. act as another dual-specificity protein phosphatase that can suppresses the tumors by dephosphorylating all phosphorylated proteins on tyrosine, serine, and threonine. In the phosphatidylinositol 3, 4, 5-trisphosphate, phosphatidylinositol 3, 4-diphosphate, phosphatidylinositol 3-phosphate, and oligosaccharide 1, 3, 4, 5-tetrakisphosphate, lipid kinase eliminates the phosphate with substrates order at the D3 position of the inositol ring. PtdIns (3,4,5) inclination P3 prefers PtdIns (3,4) P2 preference > PtdIns3P > Ins (1,3,4,5)

P4. The capacity of the PI3K-AKT/PKB signalling pathway to dephosphorylate phosphor-inositides, regulate cell cycle progression and survival, and limit of lipid phosphatase activity very much essential to its tumor suppressor role. That leads to localized tyrosine phosphorylation adhesion kinase to become dephosphorylated; the unphosphorylated form of this protein combines with AIP1 to obstruct the formation of focal adhesions, integrin-mediated cell spreading, and AKT1 activation. One important regulator of appropriate neuronal positioning, including dendritic development and synapse formation, is AKT-mTOR signalling. During adult neurogenesis, it controls the pH level at which newly generated neurons are integrated. On the other hand, TGFB1-downregulated cytoplasmic monoubiquitinated form results in less tumor suppressive potential, the monoubiquitinated nuclear form possesses a higher capacity for apoptosis. The C2 domain's capacity to bind phospholipid membranes in vitro without the need for Ca (2+) is crucial for its tumor suppressor activity, since it promotes protein stability through the PTEN ubiquitination interaction with MAGI2. In adipose tissue, this binding is a negative regulator of glucose metabolism and insulin signalling. All the adult tissues, including the heart, brain, placenta, lung, liver, muscle, kidney, and pancreas in human has been, exhibited a relatively high level of TISSUE SPECIFICITY expression. USP7 uses nuclear exclusion to carry out PTEN compartmentalization and de-ubiquitination, and it only need one amino acid, either Lys-289 or Lys-13. Retinoic acid accelerates the process of Monoubiquitination.

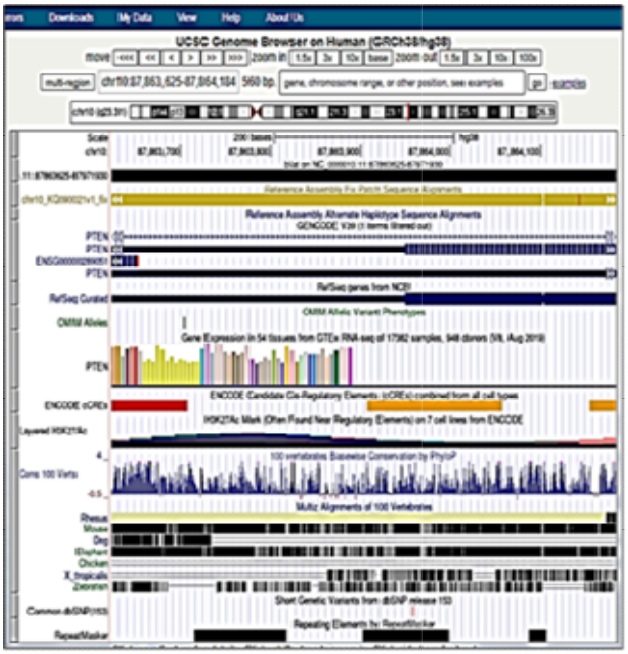
- **DISEASES**

DISEASE (1): **Cowden Disease (CD),** commonly referred to as Cowden syndrome (CS), is a condition characterized by defects in the PTEN protein. The autosomal dominant cancer susceptibility condition known as CD has been linked to an increased likelihood of developing malignancies in the skin, uterus, endometrium, breast, and thyroid. numerous Hamartoma syndrome is the primary expression of CD and affects numerous organ systems, including the skin, CNS (40%) gastrointestinal tract, thyroid (40–60%), breast (70% of CD patients), and skin. The afflicted body part has a higher chance of developing thyroid and breast malignancies independently. Additionally, CD can manifest as symptoms such as mucocutaneous papillomatosis (99%) and trichilemmomas, which are benign tumors of the follicles that produce hair infundibulum are hallmarks of CD can be the symptoms too.

DISEASE (2): **Lhermitte-Duclos Disease (LDD),** often referred to as Cerebelloparenchymal Disorder VI, is caused by defects in PTEN. The kind of cerebellar malignancy known as dysplastic gangliocytomas frequently causes the cerebellar symptoms and seizures that are characteristic of LDD. LDD and CD are regarded as Hamartoma-Neoplasia syndromes and appear to be one and the same.

DISEASE (3): **Endometrial cancer (ENDMC)** vulnerabilities result from defects in the PTEN gene. One factor contributing to endometrial cancer susceptibility is abnormalities in the PTEN gene (ENDMC). A portion of people with Proteus syndrome, a genetically diverse disorder, have PTEN mutations. Patients with Proteus syndrome is classified as belonging to the PTEN Hamartoma syndrome spectrum based on the molecular diagnosis of instances that have a PTEN mutation. Patients who survive Proteus syndrome in its early stages are probably more prone to acquire cancers in the future. Patients who survive the syndrome of Proteus in its early stages are probably more prone to acquire cancers in the future.

DISEASE (4): The cause of Macrocephaly/Autism syndrome (MCEPHAS) is defects in PTEN. People whose average circumference of the head equals +4.0 SD and those whose head circumferences vary between from +2.5 and +8 SD depending to Age and Gender along with age have macrocephaly and autism spectrum disorders. Thus, we may anticipate that chromosome 10q23 deletion syndrome is caused by PTEN. A chromosome 10q23 microdeletion involving a distinct polypeptide sequence is included in the overlapped characteristics, and it has one C2 tensin-type similarity domain. Variations in DNA sequences lead to the development of illnesses which react differently from humans to diseases and chemicals Patients have autism spectrum disorders and macrocephaly, So, we can predict that because of Chromosome 10q23 deletion syndrome is triggered by PTEN. This micro deletion involving different polypeptide sequence and, the overlapping features contain 1 C2 tensin-type similarity domain. It causes DNA sequences variations which respond to pathogens and chemicals then humans develop the diseases.

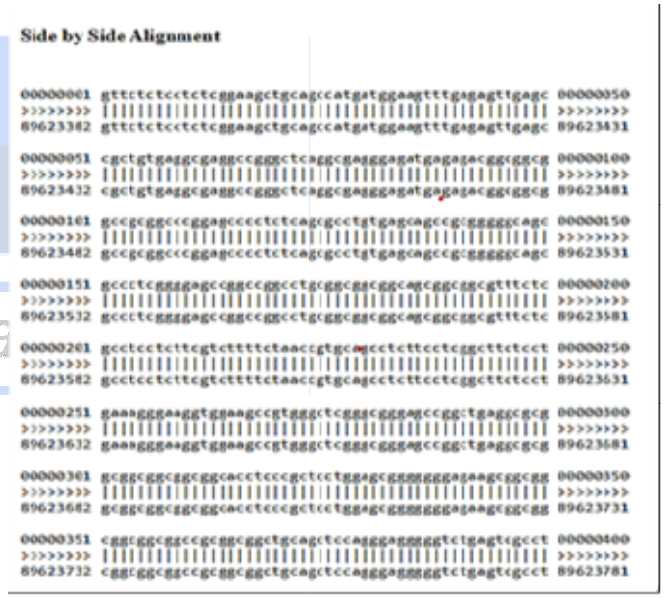



**Genomic chr10 :**

```
gcggcggagg gggcgggcag gccggcgggc ggtgatgtgg cgggactctt  87863574
tatgcgctgc ggcaggatac gcgctcggcg ctgggacgcg actgcgctca  87863624
GTTCTCTCCT CTCGGAAGCT GCAGCCATGA TGGAAGTTTG AGAGTTGAGC  87863674
CGCTGTGAGG CGAGGCCGGG CTCAGGCGAG GGAGATGAGA GACGGCGGCG  87863724
GCCGCGGCCC GGAGCCCCTC TCAGCGCCTG TGAGCAGCCG CGGGGGCAGC  87863774
GCCCTCGGGG AGCCGGCCGG CCTGCGGCGG CGGCAGCGGC GGCGTTTCTC  87863824
GCCTCCTCTT CGTCTTTTCT AACCGTGCAG CCTCTTCCTC GGCTTCTCCT  87863874
GAAAGGGAAG GTGGAAGCCG TGGGCTCGGG CGGGAGCCGG CTGAGGCGCG  87863924
GCGGCGGCGG CGGCACCTCC CGCTCCTGGA GCGGGGGGGA GAAGCGGCGG  87863974
CGGCGGCGGC CGCGGCGGCT GCAGCTCCAG GGAGGGGGTC TGAGTCGCCT  87864024
GTCACCATTT CCAGGGCTGG GAACGCCGGA GAGTTGGTCT CTCCCCTTCT  87864074
ACTGCCTCCA ACACGGCGGC GGCGGCGGCT GGCACATCCA GGGACCCGGG  87864124
CCGGTTTTAA ACCTCCCGTG CGCCGCCGCC GCACCCCCCG TGGCCCGGGC  87864174
TCCGGAGGCC gccggcggag gcagccgttc ggaggattat tcgtcttctc  87864224
cccattccgc tgccgccgct gccaggcctc tggctgctga ggagaagcag  87864274
gcccagtcgc
```

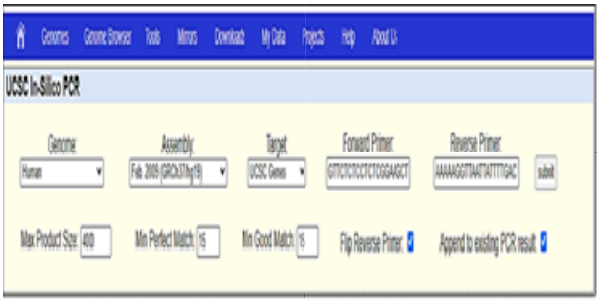
**Another Procedure of the interpretation is BLAT:**

BLAT (BLAST-Like Alignment Tool): BLAT on sequences of proteins with 80% or higher similarity and DNA sequences with 95% or higher similarity, encompassing 20 amino acids or longer crafted and measured at least forty bases in length BLAT is frequently employed to find a gene's exon structure, present a coding section inside a full-length gene, extract an EST of special interest as its own track, search for gene family members finding the human counterparts of a gene, and determine the genomic coordinates of mRNA or protein within a certain gene assembly question posed by a different species. The red indicates that the nucleotides of the query sequence and the genome vary at this location. Orange indicates that there is an insertion in the query sequence or a deletion or alignment gap in the genome at this location. When a sequence is purple, it indicates that it continues past the alignment's conclusion. Green indicates that the polyA tail of the query sequence does not seem to be aligned with the opposite side or the genome. It differs from BLAST, which is used to search sequences for gaps, and BLAT for the genome and protein database. Differences between Genomic and User Sequence Bases that match in both the Sequence vs. Genomic and cDNA sequences are capitalized and colored blue. Heavy Blue is the colour of genomic sequences, Light Blue indicates gaps, and gaps are denoted by capitalized bases those bases indicate the limits of spliced or sequence gaps. Matching bases in either cDNA sequence or splice sites often.

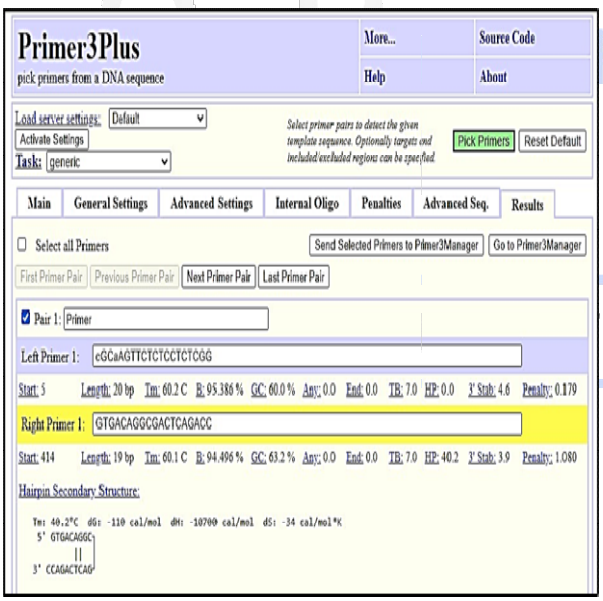



**One another Procedure of the interpretation is:**

**UCSC In-Silico PCR:**

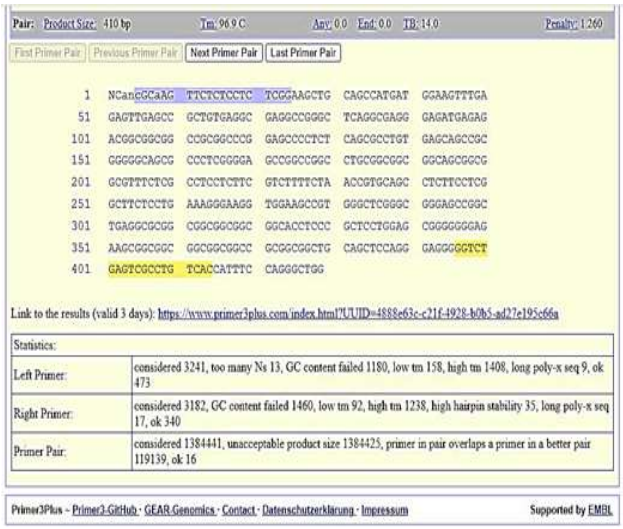


The UCSC In-Silico PCR Using two PCR primers created by Jim Kent, the UCSC In-Silico PCR is a free online server tool for searching sequence databases. A successful sequence file in FASTA format taken from NCBI/PTEN gene containing every sequence in the database is the search's output. The database sequence sections where the FASTA sequence track body has been capitalized are matched by the primer pair of the FASTA header region. Here is a human PTEN gene example: The following rules govern the UCSC in-silico PCR configuration: Look up sequences in the genome assembly database. Quary sequence is the target. Both the Forward and Reverse primers have a minimum length of fifteen bases. Maximum Product Size: increased area. Min Perfect Match: The number and size of bases on the 3' end of the primers match precisely. Min Good Match: On the 3' end of primers, two of the three bases match. Turn over Reverse primer: invert primer complements of the sequence in reverse. 8) Add to the current PCR result PCR result added to the list.




**Primer3:** The widely utilized application Primer3 is used to generate PCR primers (PCR signifies "Polymerase Chain Reaction"). In genetics and molecular biology, PCR is a vital and common technique. Primer3 is also designed for developing sequenced primer and hybridized probe. There are actually multiple applications utilizing PCR.

The Forward: 74.1 cgcaagttctctcctctcgg and Reverse: 49.6 C gtgacaggcgactcagacc calculate that melting temp activated Primer3:https://www.primer3plus.com/index.html?UUID=4888e63c-c21f-4928-b0b5-ad27e195c66a




## 7. Discussions:

Analyzing SNPs— Research investigates the causal connection among modifications to an individual's susceptibility to malignancy and variations in the spectrum of human genetics by analyzing SNPs, which which have been defined as stable single-base mutations with an average frequency of greater than 1% in at least one group. Genetic association studies have taken into consideration numerous investigations on comparable polymorphisms, and the SNP distribution of the control groups was compared with the abundance of specific polymorphisms in the genome. Genetic polymorphisms contain critical genes that are associated with a greater probability to develop cancers that are particular, such as endometrial cancer. In cancer research, scientists depend on genes and SNPs that change the way genes express themselves or function to try to comprehend documented interactions with dangerous pathways. It has been proposed that the activity of the microRNA binding site can be affected by the PTEN single nucleotide polymorphism, rs701848,. SNP rs701848 is found in the 3′ near gene area that microRNAs target in an effort to modify the miRNA binding site's function. PTEN polymorphism for prognosticating cancer patients or employing PTEN variants as markers of therapeutic response were promptly rejected. studies that concentrate on the PTEN protein or mRNA gene expression levels. The connection between two PTEN SNPs, rs701848 and rs2735343, and an increased likelihood for developing cancer has been related both of which have been connected to an increased likelihood of endometrial Several possible functional SNPs, specifically those identified in the 3′ and 5′ UTRs of the prospective

candidate genes (5′ UTR). AKT1 5′UTR region7,15,23–27; rs2494750 and rs2494752 during the mTOR gene promoter region15–23; rs2536 in the 3′UTR of mTOR7,15–18,22,23,28–30; and rs701848 in the region 7 with PTEN 3′UTR , where PTEN is thought to be a key positive regulator, are examples of miRNA bindings that may have an impact on carcinogenesis. Several miRNA bindings may have an key positive regulator. Examples of these include rs2295080, rs2494750, and rs2494752 within mTOR gene region of the promoter region number 15–23, where the Scientist and Researches focusing on the gene expression levels of PTEN proteins or mRNA. The association between rs701848 and rs2735343, two PTEN SNPs, and the probability of malignancies and cancers; both the SNPs have been linked to an elevated probability of endometrial carcinoma of squamous cells. The PTEN/AKT/mTOR pathway's core gene Single Nucleotide Polymorphisms (SNPs) may affect protein transcription and expression, modifying the pathway's ability to operate and potentially contributing significantly to the development of cancer. The relationship of meta-analysis between the susceptibility to cancer and the PTEN SNP, more than 1000 coding-region SNPs (cSNPs), and the SNP rs2735343 (G/C) increases endometrial cancer risk. During the exact moment of mutation in the gene, candidate SNPs may be implicated in steroid receptor activation pathways, cell-cycle regulation, carcinogen metabolism, DNA damage repair, and apoptosis. As a tumor suppressor gene, PTEN (phosphatase and tensin homologue) is important in controlling cell division, growth, and death. Intron 4 is one of the prevalent SNPs in the PTEN gene, and several SNPs have been found in the PTEN gene. IVS4 (rs3830the PTEN IVS4 I/D polymorphism and human malignancy in various populations675) occurs with an ACTAA insertion at 109 bp downstream of exon 4. a meta-analysis to assess the overall relationship between the IVS4 I/D polymorphism with resistance or risk of acquiring cancer. One important factor with the development of cancer could involve the expression and transcription of the proteins, which change the pathway's ability and function. The SNP rs2735343 (G/C) and the meta-analysis-associated risk of endometrial cancer: a relationship between susceptibility to cancer and PTEN gene of SNP more than 1000 coding-region of SNPs Some possible functional SNPs, known as (cSNPs), are particularly found in the 3′UTR and 5′-untranslated regions (5′UTR) of candidate genes. Certain miRNA bindings, such as rs2295080 in the mTOR gene promoter region15–23 and rs2494750 and rs2494752 in the protein AKT1 5′UTR, might communicate with miRNA binding and modify the transcriptional activity of the target genes, eventually triggering carcinogenesis. region7,15,23–27, rs2536 in the 3′UTR of mTOR7 has been ,15–18,22,23,28–30 and rs701848 the PTEN 3′UTR region7 where PTEN considered as a key positive regulator. It has been suggested that the PTEN rs701848 the Single nucleotide polymorphism influences the microRNA binding site's activity. SNP rs701848 is situated in the 3′ near gene region that microRNAs target in order to influence the activity associated with the miRNA binding site. Modifying alteration that long-term stability of the mRNA alters the expression of PTEN, and this in turn changes the susceptibility to cancer. It has been researched how SNP rs701848 is thought to operate. The rs701848 polymorphism and cancer, where the C allele, the CC/TC genotype, and the copy number CC or TC the genotype were associated with a significantly higher overall cancer risk and in the endometrial cancer subgroup. The results obtained highlight the urgent need for further research into the connection between these SNPs and cancer susceptibility as well as the essential function that genetic variations perform in the PTEN/AKT/mTOR pathway. A synopsis of some of the tracks displayed in the UCSC Genome Browser can be obtained from the discussion that accompanies. updates in the yearly Nucleic Acid Research database, subject to problem.

## 8. Bibliography:

One of the most prominent varieties of cancer impacting the female reproductive system, endometrial carcinoma begins in the uterus's endometrial interior lining. According to numerous research performed to date, cancer of the endometrium may have been caused by genetic abnormalities, resulting in an approximated hereditary contribution ranging from 2–10% overall. PCR in silico using BLAT locating alternatives for the query from a different figuring out a gene's exon DNA structure, showcasing a coding region inside an extended gene, and determining the messenger RNA or protein genomic coordinates within an assembly are just a few of the many uses for the BLAT tool. The results of a Table Browser search results may be exported into a spreadsheet or text editor for additional processing, or they can be stored in a particular track format for immediate browser upload. customization. Certain genes considered important in relation to endometrial cancer have been investigated in those studies for the conservation of protein sequences and nucleotide composition. Links to different analysis-facilitating tools are provided by the genomic and annotated data that provide the UCSC Genome Bioinformatics' visual depiction in the web browser home page. In the major portion of the unit, BLAT and the Table Browser have been made unavailable. In-

silico PCR using BLAT is the web server for the interactive indexing technique. The BLAT tool has numerous applications, which includes locating homologs of the query from a different species, identifying an exon The determining the genomic coordinates of messenger RNA or proteins in an assembly, showing a coding region inside a full-length gene, and analysing the DNA structure of a gene. outcomes of a Table Browser search can be for immediate upload into the browser. The phenotypic and disease connections, expression, regulation, comparative genomics, EST and mRNA data, the variations between repetitions, and also annotations unique to ENCODE Pilot areas have all been mapped and sequenced by the SNP tracks of UCSC Browser. Using mRNA and the SNP tracks of UCSC Browser have been helpful in the different mapping and sequencing, phenotypic and disease connections, and gene predictions. Tracks fall into the following categories based on their functionality: fundamental in different categories in the Genome Browser annotation; more classifications will surely be contributed when technological and analytical scientific advancements reach new fields. An extensive range of annotations for the more well researched genomes are available in each of the aforementioned categories on the browsers for humans and mice, along with a detailed explanation of each particular track. activated by the UCSC Browser's SNP tracks. No doubt, as scientific advances in technology and analytical efforts expand into various fields; more classifications will be added. Tracks in the Genome Browser annotation are divided into the following fundamental groups based on their functioning. An extensive range of annotation for the more extensively studied genomes can be found in each of these categories on the browsers for humans and mice, along with a detailed explanation of each particular track in there. Here all Tracks that are selected using full-length transcripts as a basis, like RefSeq, typically have lesser genomic coverage but more precise results. Protein sequence alignment, EST, and mRNA data were used to create the UCSC Genes tracks. UCSC works hard to give the research community with Browser mirror sites, click the Mirrors link on the UCSC Genome Bioinformatics home page, and BLAT service. If there are sporadic power outages or device malfunctions, the UCSC Genome Browser environment replicate the multiple mirror sites that to see a roster of currently being kept. The website at http://genome.ucsc.edu where an Employees and users who use the Genome Browser collaborate on research papers and how-to demonstrations employing the Genome Browser alongside associated technologies. Although it is not required for registration in order to look up and browse the contents, users are encouraged to do so in order to upload and modify content and make advantages of the Sessions utility's UCSC storing function. Technical articles and examples provided from us, the users. Searching and viewing the contents do not require registration, however users are urged to do so in order to modify and upload information, utilize the UCSC storage, and to take advantage of the user-friendly graphical interface. The underlying database is accessible through a variety of channels and utilizes a robust, but very basic, architecture. Visitors can add their own unique annotations to the browser's integrated annotation context, which can be seen in either a public or private format. We want to keep updating this website and modify it to include additional SNP genomes utilizing the UCSC Browser in the upcoming years. Because the Gene prediction tracks depend on various types of experimental evidence, it can occasionally be difficult to tell if something unusual is mistaken or a transcript. Here Tracks that have been meticulously carefully selected depending on certain full-length transcripts, such as RefSeq, typically have lesser genomic coverage but more precise results. Protein sequence alignment, EST, and mRNA data were used to create Genes at UC Santa Cruz. Click in the UCSC Genome Bioinformatics Mirrors link. homepage for further details and the BLAT service. UCSC undertakes significant procedures to provide the researchers with a wide range of Browser mirror sites. The genome of UCSC Browser environment replicates numerous mirror sites that to browse a list of frequently updated websites where a helpful instrument for investigating the human genome website can provide rapid sequence and text-based searching capabilities in case of a periodic function interruptions in power or equipment malfunction. The graphical representation is comparatively the coordinates of the genome that distinguish mRNA or protein and determine a gene's exon structure by smoothly displaying a coding region inside a whole gene homology. The Genome Browser employees as well as customers can access the databases using a variety of methods, and they are quite straightforward. The annotations integrated into the browser may be seen publicly or privately, and users have the option to add their own unique annotations. Our goal for the upcoming years is to keep expanding this website and modifying it to include additional SNP genomes using the UCSC Browser.



## 9. **Acknowledgments**:

The Chancellor of Techno India University in West Bengal is acknowledged by the authors for offering the required infrastructure.

## References


[1] International Human Genome Sequencing Consortium (2001)Initial
sequencing and analysis of the human genome. Nature, 409, 860–921.

[2] Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H.,Zahler,A.M. and Haussler,D. (2002) The Human Genome Browser at UCSC. Genome Res.,12, 996–1006.

[3] Kent, W.J. and Brumbaugh,H. (2002) autoSql and autoXml: Code Generators from the Genome Project. Linux J., 99, 68–77.

[4] Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. Genome Res., 12, 656–664.

[5] Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T., Durbin,R., Eyras,E., Gilbert,J., Hammond,M., Huminiecki,L., Kasprzyk,A., Lehvaslaiho,H., Lijnzaad,P., Melsopp,C., Mongin,E., Pettett,R., Pocock,M., Potter,S., Rust,A., Schmidt,E., Searle,S.,Slater,G., Smith,J., Spooner,W., Stabenau,A., Stalker,J., Stupka,E., Ureta-Vidal,A., Vastrik,I. and Clamp,M. (2002) The Ensembl genome database project. Nucleic Acids Res., 30, 38, 41

[6] Salamov, A.A. and Solovyev, V.V. (2000) Ab initio gene finding in Drosophila genomic DNA. Genome Res., 10, 516–522.

[7] Kulp,D., Haussler,D., Reese,M.G. and Eeckman,F.H. (1996) A generalized hidden Markov model for the recognition of human genes in DNA. Ismb,4, 134–142.

[8] Burge,C. and Karlin,S. (1997) Prediction of complete gene structures in human genomic DNA. J. Mol. Biol., 268, 78–94.

[9] Smith, A.F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr.Opin.Genet. Dev., 9, 657–663.

[10] Benson, G. (1999) Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res., 27, 12–17.

[11] [11] Cheung,V.G., Nowak,N., Jang,W., Kirsch,I.R., Zhao,S., Chen,X.N., Furey,T.S., Kim,U.J., Kuo,W.L. and Livier,M. (2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature, 409, 953–958.

[12] [12] Dowell,R.D., Jokerst,R.M., Day,A., Eddy,S.R. and Stein,L. (2001) TheDistributed Annotation System. BMC Bioinformatics, 2, 7.54 Nucleic Acids Research, 2003, Vol. 31, No. 1.edition.

[13] R.M. Kuhn, et al., The UCSC Genome Browser Database: update 2007, Nucleic Acids Res. 35 (2007) D668–D673 [Database issue].

[14] A.S. Hinrichs, et al., The UCSC Genome Browser Database: update 2006, Nucleic Acids Res. 34 (2006) D590–D598 [Database issue].

[15] W.J. Kent, BLAT—the BLAST-like alignment tool, Genome Res.12 (4) (2002) 656–664.

[16] Pruitt, K.D. and Maglott, D.R. 2001. RefSeq and LocusLink: NCBI gene-centered resources. Nucleic Acids Res. 29: 137–140.

[17] Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L.,Fischer, C., Fizames, C., Wincker, P., Brottier, P. Quetier, F., etal. 2000. Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence.Nat. Genet. 25: 235–238.

[18] Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein,L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey,W.L., et al. 2001." A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature409: 928–933.

[19] Schuler, G.D. 1998. Electronic PCR: Bridging the gap between genome mapping and genome sequencing. Trends Biotechnol. 16: 456–459.Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D.,

[20] 000 Genomes Project Consortium. A map of human genome variation from population-scalesequencing. Nature. 2010;467(7319):1061–73. [PMC free article] [PubMed] [Google Scholar]

[21] [21]. Amberger JS, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM) Nucleic Acids Res. 2009;37:D793–D796. [PMC free article] [PubMed] [Google Scholar]

[22] Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. Nat Genet. 2004;36:431–432. [PubMed] [Google Scholar]

[23] Bejerano G, Siepel AC, Kent WJ, Haussler D. Computational screening of conserved genomic DNA in search of functional noncoding elements. Nat Methods. 2005;2:535545. [PubMed] [Google Scholar]

[24] Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 2011;39(Database issue):D32–7. [PMC free article] [PubMed] [Google Scholar]

[25] Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 1999;27:12–17. [PMC free article] [PubMed] [Google Scholar]

[26] Smit AF. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Gen Dev. 1999;9:657–663. [PubMed] [Google Scholar]

[27] Strausberg RL, Greenhut SF, Grouse LH, Schaefer CF, Buetow KH. In silico analysis of cancer through the Cancer Genome Anatomy Project. Trends Cell Biol. 2001;11:S66–S71. [PubMed] [Google Scholar]

[28] Thomas DJ, Rosenbloom KR, Clawson H, Hinrichs AS, Trumbower H, Raney BJ, Karolchik D, Barber GP, Harte RA, Hillman-Jackson J, Kuhn RM, Rhead BL, Smith KE, Thakkapallayil A, Zweig AS, Kent WJ The ENCODE Project Consortium Haussler D. The ENCODE project at UC Santa Cruz. Nucleic AcidsRes. 2007;35:D66667. [PMC free article] [PubMed] [Google Scholar]

[29] The UniProt Consortium. The Universal Protein Resource (UniProt) 2009. Nucleic Acids Res. 2009;37:D169–D174. [PMCfree article] [PubMed] [Google Scholar]

[30] Verdu P, Barreiro LB, Patin E, Gessain A, Cassar O, Kidd JR, Kidd KK, Behar DM, Froment A, Heyer E, Sica L, Casanova JL, Abel L, Quintana-Murci L (2006) Evolutionary insights into the highworldwide prevalence of MBL2 deficiency alleles. Hum Mol Genet 15:2650–2658. . [PMC free article] [PubMed] [Google Scholar]

[31] The UCSC genome browser and associated tools: Robert M. Kuhn, David Haussler, W. James Kent; Briefings in Bioinformatics, Volume 14, Issue 2, March 2013, Pages 144-161, Published: 18 August 2012; https://doi.org/10.1093/bib/bbs038

[32] The UCSC Genome Browser Database: update 2006 : A. S. Hinrichs, D. Karolchik, R. Baertsch, G. P. Barber, G. Bejerano, H. Clawson, M. Diekhans, T. S. Furey, R. A. Harte, F. Hsu ... Show more Nucleic Acids Research, Volume 34, Issue suppl_1, 1 January 2006, Pages D590–D598, Published: 01 January 2006. https://doi.org/10.1093/nar/gkj144

[33] The UCSC Genome Browser: Donna Karolchik,Angie S. Hinrichs,W. James Kent; First published: 01 December 2012; https://doi.org/10.1002/0471250953.bi0104s40.

[34] The UCSC Genome Browser Database: update 2009; R. M. Kuhn, D. Karolchik, A. S. Zweig, T. Wang, K. E. Smith, K. R. Rosenbloom, B. Rhead, B. J. Raney, A.

Pohl, M. Pheasant ... Show more Nucleic Acids Research, Volume 37, Issue suppl_1, 1 January 2009, Pages D755–D761, https://doi.org/10.1093/nar/gkn875 Published: 07 November 2008.

[35] The UCSC Genome Browser database: update 2010 ; Brooke Rhead, Donna Karolchik, Robert M. Kuhn, Angie S. Hinrichs, Ann S. Zweig, Pauline A. Fujita, Mark Diekhans, Kayla E. Smith, Kate R. Rosenbloom, Brian J. Raney, Andy Pohl, Michael Pheasant, Laurence R. Meyer, Katrina Learned, Fan Hsu, Jennifer Hillman-Jackson, Rachel A. Harte, Belinda Giardine, Timothy R. Dreszer, Hiram Clawson, Galt P. Barber, David Haussler, W. James Kent Nucleic Acids Research, Volume 38, Issue suppl_1, 1 January 2010, Pages D613-D619; Published: 10 November 2009 https://doi.org/10.1093/nar/gkp939.

[36] [34] The UCSC Genome Browser: What Every Molecular Biologist Should Know: Mary E. Mangan,Jennifer M. Williams,Robert M. Kuhn,Warren C. Lathe III; First published: 15 October 2009;https://doi.org/10.1002/0471142727.mb1909s88.

[37] UCSC Genome Browser: Deep support for molecular biomedical research; Mary E.Mangan1Jennifer M.Williams1Scott M.Lathe1DonnaKarolchik2Warren C.LatheIII1; Open Helix, LLC, 12600 SE 38th Street, Suite 230, Bellevue, WA 98006, USA; UCSC Genome Bioinformatics Group: Center for Biomolecular Science & Engineering, CBSE/ITI, 501D Engineering II Building, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA.

[38] The UCSC Genome Browser database: 2019 update: Maximilian Haeussler, Ann S Zweig, Cath Tyner, Matthew L Speir, Kate R Rosenbloom, Brian J Raney, Christopher M Lee, Brian T Lee, Angie S Hinrichs, Jairo Navarro Gonzalez, David Gibson, Mark Diekhans, Hiram Clawson, Jonathan Casper, Galt P Barber, David Haussler, Robert M Kuhn, W James Kent; Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D853-https://doi.org/10.1093/nar/gky1095

## Websites

[1] Ensembl genome browser (http://www.ensembl.org)
[2] VISTA (http://genome.lbl.gov)
[3] NCBIMap-Viewer (http://www.ncbi.nlm.nih.gov/mapview/)
[4] ECR Browser (http://ecrbrowser.dcode.org)
[5] Combo (http://www.broad.mit.edu/annotation/argo/)
[6] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5707100/
[7] https://www.nature.com/articles/s41598-017-17250-z
[8] https://www.frontiersin.org/articles/506887
[9] https://www.researchgate.net/publication/289795366_Analysis_of_association_between_PTEN_gene_single_nucleotide_polymorphism_and_stomach_neoplasms
[10] https://www.sciencedirect.com/science/article/pii/S0888754320319947
[11] https://www.pnas.org/content/106/24/9761
[12] https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-11-208
[13] https://academic.oup.com/jmcb/article/11/7/578/5509869
[14] https://commonfund.nih.gov/gtex
[15] https://www.nature.com/articles/ng.2653
[16] https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0155270
[17] https://cancerres.aacrjournals.org/content/73/2/725
[18] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5557368/nihms123730.
[19] https://aacrjournals.org/cancerres/article/73/2/725/586396/SMAD2-SMAD3-and-SMAD4-Mutations
[20] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7650668/
[21] http://shgcwww.stanford.edu/Mappin
i. g/TNGMAPS/; Radiation hybrid maps at Stanford University.
[22] http://genome www.stanford.edu/Saccharomyces/ The Saccharomyces Genome Database (SGD) at Stanford University.
[23] http://www.biodas.org; Distributed Annotation System web site.
[24] http://www.cse.ucsc.edu/~Kent/intronerator; C. elegans genome browser with an emphasis on alternative splicing.
[25] http://www.sanger.ac.uk/Software/
[26] Formats/GFF; Description of the
[27] Gene Finder Format (GFF).
[28] www.mysql.com; the main web site for the MySQL database.
[29] UCSC In-Silico PCR
[30] https://genome.ucsc.edu/cgi-bin/hgPcr
[31] Primer3:https://www.primer3plus.com/index.html?UUID=4888e63c-c21f-4928-b0b5-ad27e195c66a