IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 09, Issue 10, October. 2019, Series -III, PP 55-66

Analysis of Human Genome Sequences to Identify Disease Genes

Md. Jashim Uddin[1], Md. Islamul Haque[2], Dr. Paresh Chandra Barman[3], Khandaker Takdir Ahmed[4], Kazi Mowdud Ahmed[5], Md. Ibrahim Khalil[6], MM Asaduzzaman Sabbir[7]

[1]Assistant Professor Dept. of ICT, Islamic University, Kushtia-7003, Bangladesh,
[2]Student Dept. of ICT, Islamic University, Kushtia-7003, Bangladesh,
[3]Professor Dept. of ICT, Islamic University, Kushtia-7003, Bangladesh,
[4,5]Lecturer Dept. of ICE, Islamic University, Kushtia-7003, Bangladesh,
[6,7]Student Dept. of CSE, First Capital University of Bangladesh, Chuadanga-7200, Bangladesh.
Corresponding Author: Md. Islamul Haque
Received 16 October 2019; Accepted 31 October 2019

Abstract: A genome is the collection of DNAs that comprises an organism Each individual organism's genome contains the genes and other DNA elements that ultimately define its identity. Genomes range in size from the smallest viruses, which encode fewer than 10 genes, to eukaryotes such as humans that have billions of base pairs of dans encoding tens of thousands of genes. The recent sequencing of genomes from all branches of life including viruses, bacteria, archaea, fungi, nematodes, plants and humans presents us with an extraordinary moment in the history of biology. By analogy, this situation resembles the completion of the periodic table of the elements in the nineteenth century. As it became clear that the periodic table could be arranged in rows and columns, it became possible to predict the properties of individual elements. A logic emerged to explain the properties of the elements, but it still took another century to grasp the significance of the elements and to realize the potential of the organization inherent in the periodic table. Today we have sequenced the DNA from thousands of genomes without putting on a lab cost. This process will take decades. A variety of tools must be applied, including bioinformatics approaches, genetics and cell biology. In this paper, we will analyze the Genome Sequence to identify disease gens.

Keywords: DNA, EST ,Genome, Gene, Genome Sequence, Hemochromatosis.

I. INTRODUCTION

A. Problem Overview

The genetic material of all multi-cellular organisms that is the famous double helix of Deoxyribonucleic-Acid (DNA) which contains all of our genes. In term, DNA is formed of four chemical bases, pairs of which make the "rungs" of the twisted, ladder-shaped DNA molecules [1]. All genes are created stretches of these four bases, arranged in several ways and in different lengths. HGP (Human Genome Project) researchers have decoded the human genome in three major ways:

1. Determining the sequence of all the bases in our genome's DNA.
2. Creating maps that show the locations of genes for major sections of all our chromosomes.
3. Manufacturing what are referred to as linkage maps.

The Human Genome Project (HGP) was the international co-operative research program whose goal was the entire mapping and understanding of all the genes of human beings. All our genes together are called genome. The HGP has stated that there are probably about 20,500 human genes [1][2]. Now the completed human sequence can determine their locations. This final result of the HGP has given the world a resource of detailed information regarding the structure, organization and function of the entire set of human genes. The International Human Genome Sequencing Consortium published the primary draft of the human genome in the journal Nature in February 2001 with the sequence of the entire genome's 3 billion base pairs about 90 percent completed. A tremendus finding of this primary draft was that the amount of human genes appeared to be considerably fewer than previous estimates, which ranged from 50,000 genes to as several as 140,000 [3]. The final sequence was completed and published in April 2003 [3][4].

B. Contributions

The tools created through the HGP also continue to inform efforts to characterize the complete genomes of many alternative organisms used extensively in biological research project such as mice, fruit flies and flatworms. These efforts support one another, because most of the organisms have many similar, or

"homologous," genes with similar functions. Therefore, the determination of the sequence or function of a gene in a model organism, for instances, the roundworm C. elegans, has the potential to demonstrate a homologous gene in human beings or in one of the other model organisms. These imaginative goals required and will continue to demand a variety of latest technologies that have made it possible to relatively rapidly construct a primary draft of the human genome and to continue to refine that draft. These techniques include:

- DNA Sequencing
- The Employment of Restriction Fragment-Length Polymorphisms (RFLP)
- Yeast Artificial Chromosomes (YAC)
- Bacterial Artificial Chromosomes (BAC)
- The Polymerase Chain Reaction (PCR)
- Electrophoresis

Therefore, advanced methods for widely disseminating the data generated by the HGP to scientists, physicians and others, is important in order to make sure the most rapid application of analysis results for the advantage of humanity. Biomedical technology and research are particular advantages of the HGP.

In our paper, we will introduce the ways to identify disease genes from the expressed sequence data such as that may have been obtained from patients. Firstly, provides an introduction to the human genome assembly and the resources such as:

- Basic Local Alignment Search Tool (BLAST)
- Genome Data Viewer (GDV)
- Single Nucleotide Polymorphism database (dbSNP)
- Online Mendelian Inheritance in Man (OMIM)

Then demonstrate of these resources to the identification of genes related to disease Hemochromatosis.

C. Proposed Solutions

- To analysis the human genome sequences
- To identify the hidden message in DNA
- To identify the disease genes

II. HEMOCHROMATOSIS

Hemochromatosis (HE-mo-kro-ma-TO-sis) is a disease in which an excessive amount of iron builds up in your body (iron overload). Iron is a mineral which found in many foods. An Excessive amount of iron is toxic to your body. It will poison your organs and cause organ failure [5]. In hemochromatosis disease, iron can build up in most of your body's organs, but particularly in the liver, heart, and pancreas. An excessive amount iron in the liver can cause an enlarged liver, liver failure, cancer of liver, or cirrhosis (sirRO-sis). Cirrhosis is scarring of the liver that causes the organ to not work well. An excessive amount of iron in the heart can cause irregular heartbeats referred to as arrhythmias (ah-RITH-me-ahs) and heart failure [6][7]. An excessive amount of iron in the pancreas can lead to diabetes. If hemochromatosis is not treated, it may even cause death.

A. Signs, Symptoms, and Complications

Hemochromatosis can affect several parts of the body and cause various signs and symptoms. Many of the signs and symptoms are just like those of different diseases. Signs and symptoms of hemochromatosis typically do not occur until middle age. Firstly, women are more likely to have possessed general symptoms such as fatigue (tiredness). In men, complications such as diabetes or cirrhosis (scarring of the liver) often are the primary signs of the disease [6]. Signs and symptoms conjointly vary based on the severity of the disease. Common signs and symptoms of hemochromatosis embrace joint pain, fatigue, general weakness, weight loss, and stomach pain. Not everyone who has hemochromatosis has the same signs or symptoms of the disease. Estimates of what number of people develop signs and symptoms vary greatly. Some estimates recommend that as many as half of all people who have the disease do not have signs or symptoms [7].

B. Risk factors

There are some known risk factors for hemochromatosis:

- Genetic factors: Having two copies of a mutated "high iron" or, HFE gene, is the greatest risk factor for hereditary hemochromatosis. The person inherits one copy of the mutated HFE gene from each parent. H refers to high, and FE means iron.
- Family history: A person with a parent, child, brother, or sister with hemochromatosis is more likely to have it.

- Ethnicity: People of British, Scandinavian, Dutch, German, Irish, and French ancestry have a higher risk of having the HFE gene mutation and of developing hemochromatosis.
- Gender: Men are significantly more likely to develop hemochromatosis than women, and they tend to experience signs and symptoms between the ages of 40 and 60 years, while women are more likely to develop it after menopause.

C. Primary hemochromatosis: A genetic mutation

Every living organism has genes. Genes are a set of collection of instructions that decide what the organism is like, how it survives, and how it behaves in its environment. A mutation in one gene can change the way the body works. HFE is the gene that controls the amount of iron we have tendency to absorb. The two common mutations in the HFE gene are C282Y and H63D [7]. In the U.S., most people with inherited hemochromatosis have inherited two copies of C282Y, one from the mother and the other from the father. Around 31 percent of people with two copies of C282Y develop symptoms by their early fifties. A person who inherits only one gene with the C282Y mutation is not demonstrated to develop iron overload syndrome, although they will probably absorb a lot of iron than normal, and they will be a carrier. If both parents are carriers, there is a 1 in 4 chance of inheriting two mutated genes, one from each parent. However, some people with two copies of the C2H2Y mutation never experience symptoms. Some individuals may inherit one C282Y and one H63D mutation. A small proportion of these people will develop hemochromatosis symptoms. Inheriting two copies of H63D is rare. Having two copies of the H63D mutation may increase the risk of developing hemochromatosis, but this is not confirmed. Men with HFE defects can develop symptoms from the age of 40, but in women, symptoms normally appear after the menopause [8].

D. Hemochromatosis Complications

If hemochromatosis is not found and treated early, iron builds up in your body and may lead to:

- Liver disease, including an enlarged liver, liver failure, cancer of liver, or cirrhosis (scarring of the liver)
- Heart problems, such as arrhythmias (irregular heartbeats) and heart failure
- Diabetes, particularly in people who have a family history of diabetes
- Joint damage and pain, including arthritis
- Reproductive organ failure, such as erectile dysfunction (impotence), shrinkage of the testicles, and loss of sex drive in men, and absence of the menstrual cycle and early menopause in women
- Changes in skin color that build the skin look gray or bronze
- Underactive pituitary and thyroid glands
- Damage to the adrenal glands

III. GENERAL PROTOCOL AND REQUIRED RESOURCES

A. Outline of Steps

1. Compare the sequences of ESTs from a patient to the sequences of the human genome (using Basic Local Alignment Search Tool [BLAST]).
2. Identify the genes aligning to the ESTs and download their sequences (using Map Viewer).
3. Identify whether the EST sequences contain any known SNPs (using dbSNP).
4. Determine whether a gene variant is known to cause a phenotype (using Online Mendelian Inheritance in Man [OMIM]).

Thus, starting from the transcribed sequences derived from patients, we will obtain information about expressed genes and determine whether these genes contain known variations that lead to the disease phenotype [8].

B. Descriptions of Resources

Used NCBI assembles component sequences from the human genome sequencing project into longer sequences called **contigs** whose accession numbers begin with prefix "NT_". NCBI also performs a number of annotations on the assembly to identify genes, transcripts, clones, repeats, markers, and SNPs. NCBI releases the updated human genome assembly or the new "Build" periodically. For more information about the human genome assembly and annotation, reference no [7]. and the help document (http://www.ncbi.nlm.nih.gov/mapview/static/humansearch.html).This project use of NCBI resources such as BLAST, Map Viewer, dbSNP, and OMIM as tools to identify disease genes [7][8].

C. BLAST

BLAST provides a method for rapid searching of nucleotide and protein databases for similarities with a query nucleotide or protein sequence. The human genome BLAST page at (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606) provides centralized access to

the NCBI human genome assembly and annotated transcript and protein sequences. The BLAST output links directly to the **Human Genome Data Viewer**, where database hits can be analyzed in their genomic context to see the relationship with other annotated features.

D. Genome Data Viewer

The Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/) allows us to view and search an organism's complete genome [9]. It shows integrated views of a collection of genetic, physical, and sequence maps for annotated genes, expressed sequences, SNPs, and other features, and, thus, is a valuable tool for the identification and localization of genes that contribute to human disease[10][11].

E. dbSNP

NCBI's SNP database (http://www.ncbi.nlm.nih.gov/SNP/) contains both single nucleotide substitutions, and short deletion and insertions [12]. The data in dbSNP are integrated with other NCBI genomic data. SNPs are aligned to the human genome and the locations of SNPs with respect to the annotated genes and mRNAs are identified.

F. OMIM

OMIM (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM) is the database of human genes and genetic disorders developed and edited by Dr. Victor A. McKusick and his colleagues at Johns Hopkins and elsewhere, and adapted for the Internet by NCBI (Note 1 about Online Mendelian Inheritance in Animals) [13].

IV. Methodology and Result Analysis

We will identify the hemochromatosis disease genes, which is characterized by an iron overload. Consider that we are working on the hemochromatosis disease and needs to obtain information about the gene(s) causing the phenotype. The following steps will describe the analysis of EST sequences that might have been obtained from a hemochromatosis patient. Sample procedures are given below:

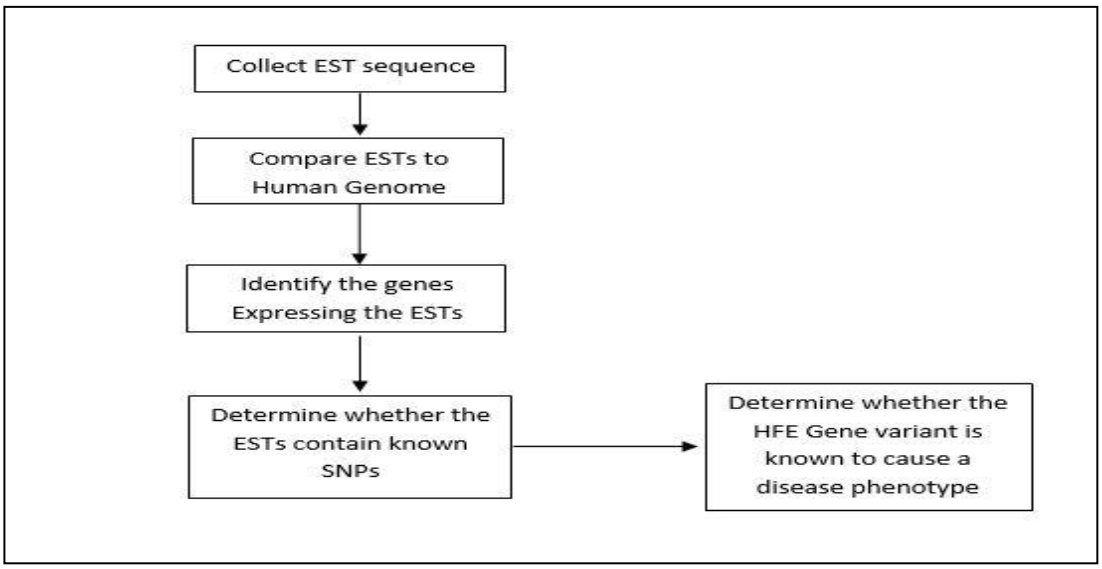


Fig. 1. Flow diagram of the methodology

Step 1: The EST sequences are given below:

TGCCTCCTTTGGTGAAGGTGACACATCATGTGACCTCTTCAGTGACCACTCTACGGTGTCG
GGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCCAATGGATG
CCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAGGGCTGGATAACC
TTGGCTGTACCCCCTGGGGAAGAGCAGAGATATACGTACCAGGTGGAGCACCCAGGCCTGGATCA
GCCCCTCATTGTGATCTGGG

Step 2: Compare ESTs to The Human Genome One way to identify the genes expressing the ESTs is to compare their sequences using BLAST with the human genome assembly and the genes annotated on it. The specialized BLAST page for searching against the annotated human genome assembly is at (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606). We can concatenate a number of EST sequences to run the search as a batch. However, we will use only one EST sequence as a query for this

analysis. Paste the patient's EST sequence in the query box of the BLAST page and select the "Refseq genome" database from the pull-down menu and use the default program Mega Blast [9][11].

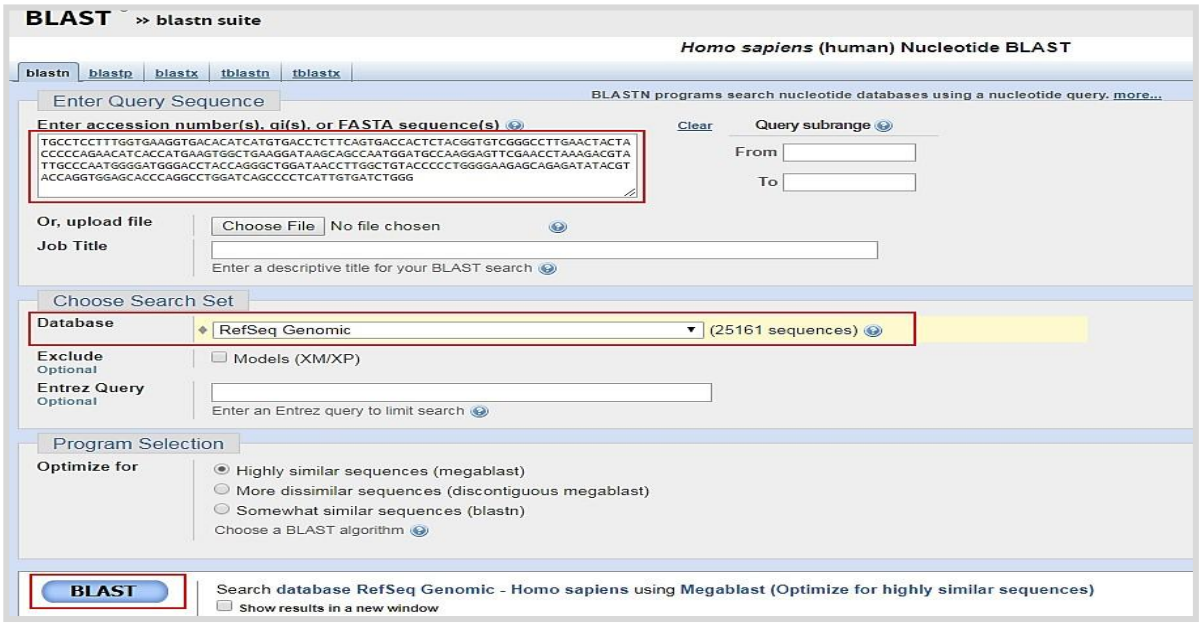


Fig. 2. Step 2 Compare ESTs to the Human Genome

Start the search by clicking on the "Blast". The BLAST results page shows only one match to the contig sequence NT_007592.16 on chromosome 6 in the human genome Build GRCh38.p12 Primary Assembly. In certain cases, there may be multiple matches to the human genome assembly.

Fig. 3. The alignment of the query EST sequence (indicated by "query") and the matched sequence from chromosome 6 (indicated by "sbjct") shows that the EST sequence is only 99 % identical to the genomic sequence.

Note the location of the nucleotide that is different between the two sequences (a G to A variation at the nucleotide 26,032,913 of the contig NT_007592.16 and nucleotide 26,092,913 of the contig NC_000006.12). Click MSA viewer menu

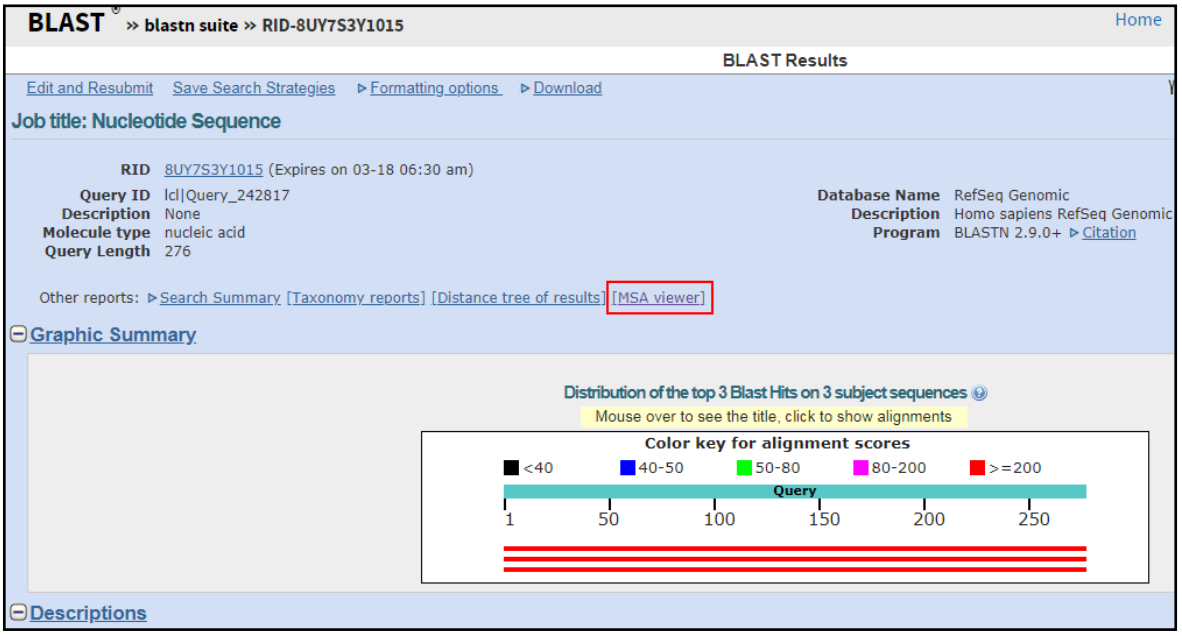


Fig. 4. BLAST result

Then we see the mismatch sequence and it's contig and nucleotide positions. The difference may be due to a sequencing error in the low-quality EST sequence or it may represent a real SNP in the human genome.

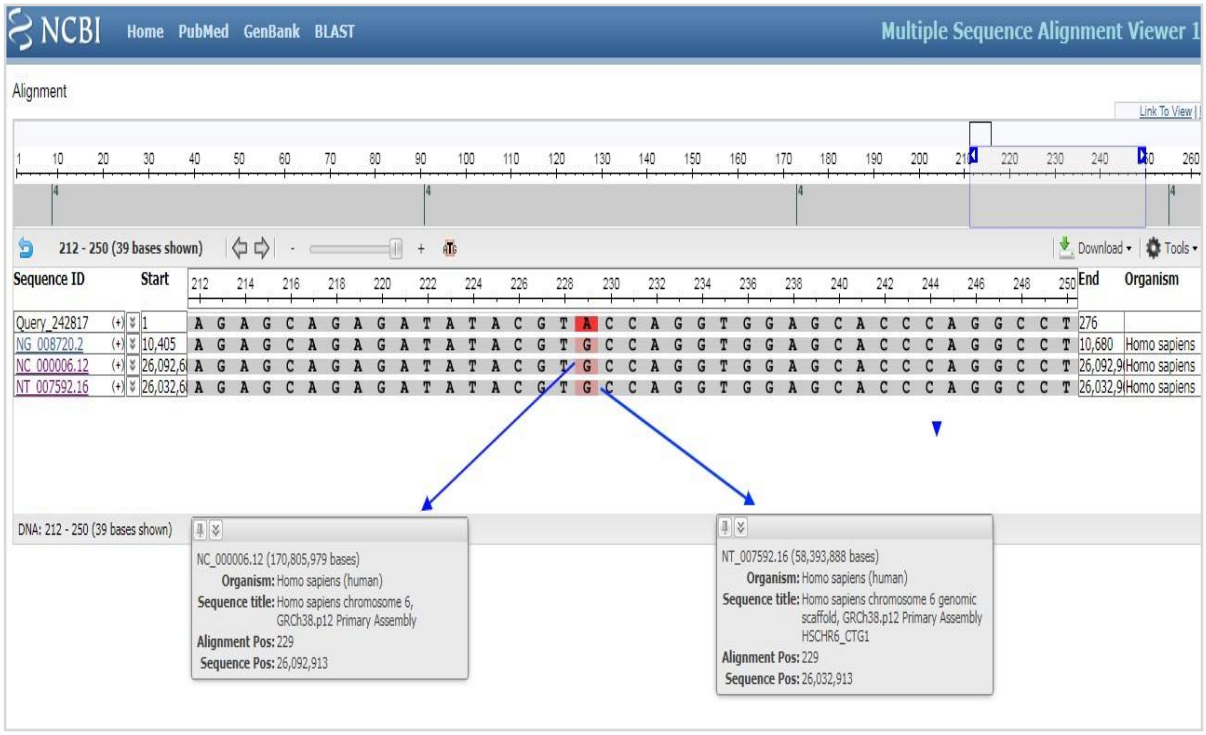


Fig. 5: Multiple Sequence Alignment Viewer Result

Step 3: Identify the Genes Expressing the ESTs and Download Their Sequences

Fig.6. Genome Data Viewer display of the Basic Local Alignment Search Tool. The four maps displayed in this view, Model, RNA, Gene-seq, Contig, are highlighted by rectangles.

We will now take advantage of the NCBI annotation of the human genome assembly to identify the gene corresponding to the EST by using the Genome Data Viewer. To visualize the BLAST hit on the genome using Genome Data Viewer, click the "Genome Data Viewer" button of the BLAST results page, then on the Map element "NC_000006.12." Currently, it should be displayed (Model, RNA, Genes seq and Contig).

The Genes seq map shows the "known" genes annotated by alignment of EST and/or mRNA sequences to the assembly. The Contig map shows the assembled genome contig sequence in the region, the Model map shows the Ab initio model genes predicted by the NCBI's program Gnomon and the RNA map shows the alignments of the known alternatively spliced transcripts. The BLAST hit, indicated by the red bar, is within the region of one of the exons of the HFE gene annotated on the human genome sequence. The thick bars in the Genes seq map indicate the exons and the thin lines joining them indicate introns of the gene. Zoom out several times until the user sees the entire HFE gene structure by clicking on the gray line and selecting option "Zoom out 2 times" from the menu that appears. The query EST represents a known gene, HFE. The orientation of the arrow next to the gene link indicates the orientation of the gene on the forward or the reverse strand. A gene annotated on the forward strand is indicated by an arrow pointing downward whereas a gene annotated on the reverse strand is indicated by an arrow pointing upward. The HFE gene is annotated strand of chromosome 6.

The current map, Genes seq map, has links to resources that provide more information about the HFE gene such as OMIM, sv(Sequence Viewer), pr (Reference Proteins), dl (Download Sequence), ev (Evidence Viewer), mm (Model Maker), and hm (Homologene). Display the entire HFE gene sequence by clicking on the download "dl" link and then on "Display" on the next page (Notes 8 and 9). Note the accession number of the longest transcript, NM_000410. We will use this information in the next step.

Step 4: Determine Whether the ESTs Contain Known SNPs Go back to the Genome Data Viewer report:

In this case SNPs. Zoom in on the blast hit area by clicking on the map line next to it and choosing the appropriate zoom level. There are one SNPs in the area; rs1800562.

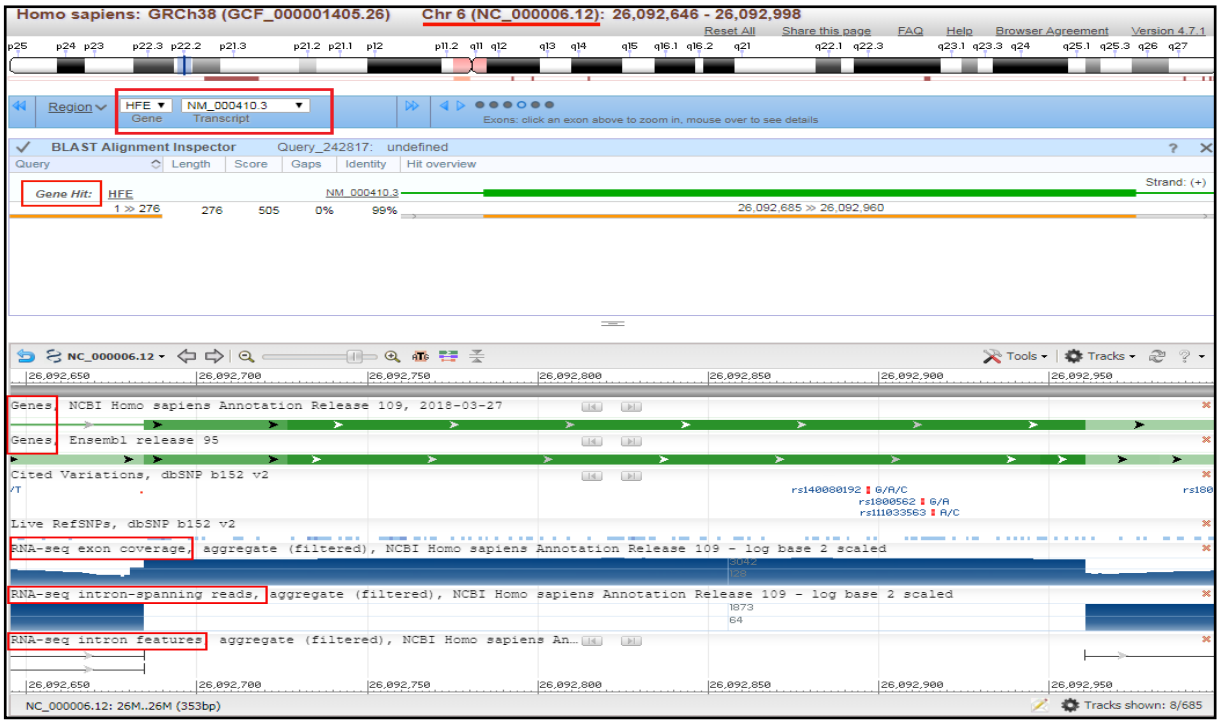


Fig. 7. Finding The HFE gene is annotated strand of chromosome 6

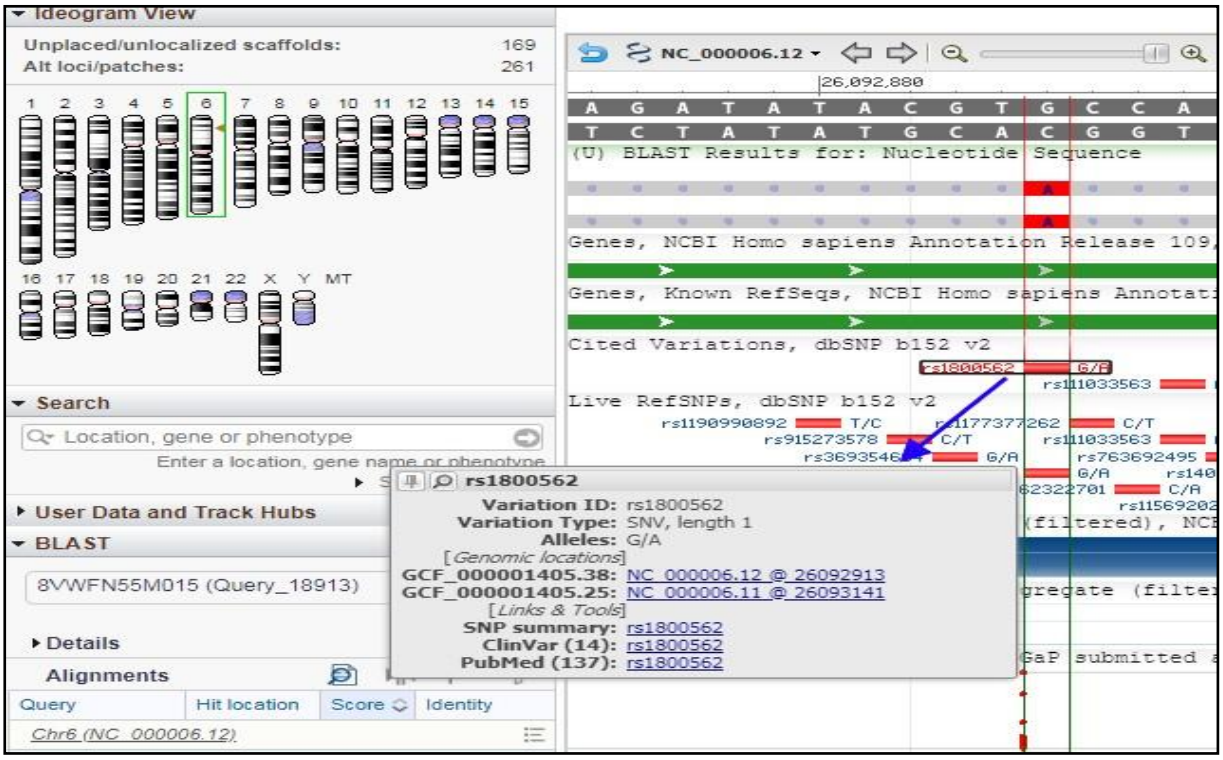


Fig. 8. Genome Data Viewer display, containing the variation and Genes_seq maps, zoomed in the region of the Basic Local Alignment Search Tool (BLAST) hit in Subheading 5. There are two SNPs, rs1800562.

Then we check,

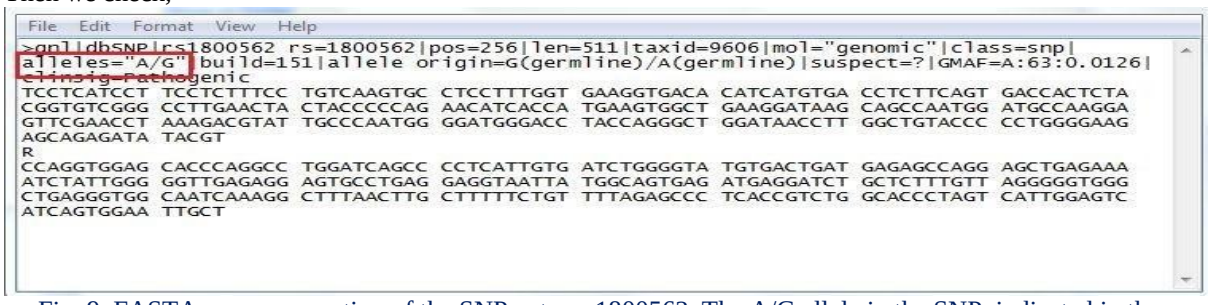


Fig. 9. FASTA sequence section of the SNP entry rs1800562. The A/G allele in the SNP, indicated in the definition line on the record, is highlighted by a rectangle.

Fig. 10: Integrated maps section of the SNP entry rs1800562. The location of the SNP, nucleotide position 26092913 on the contig NC_000006.12 of the reference assembly, is highlighted by a rectangle.

Gene Model (mRNA alignment) information from genome sequence

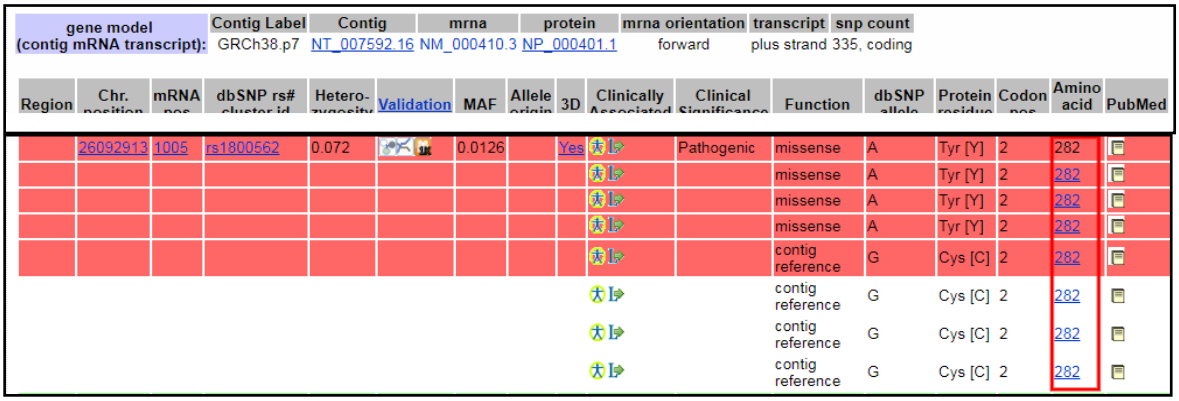
Total gene model (contig mRNA transcript):				11		
mrna	transcript	protein	mrna orientation	Contig	Contig Label	List SNP
NM_000410.3	plus strand	NP_000401.1	forward	NT_007592.16	GRCh38.p7	<- currently shown
XM_011514543.2	plus strand	XP_011512845.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139011.2	plus strand	NP_620580.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139010.2	plus strand	NP_620579.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139009.2	plus strand	NP_620578.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139008.2	plus strand	NP_620577.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139007.2	plus strand	NP_620576.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139006.2	plus strand	NP_620575.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139004.2	plus strand	NP_620573.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_139003.2	plus strand	NP_620572.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel
NM_001300749.1	plus strand	NP_001287678.1	forward	NT_007592.16	GRCh38.p7	View snp on GeneModel

Fig. 11. Gene Model (mRNA Alignment) information from genome sequence

This is the same nucleotide variation on the contig NT_007592.14 found in the BLAST result in Subheading 5.1 (26092913 G to A). To identify whether this change represents a change in an encoded amino acid, we will refer to the GeneView panel. This view shows the location of the SNP in the alternatively spliced products annotated on all the assemblies. It also provides information at the protein level; the amino acid number and the change in the sequence, if any. Refer to the panel for the longest transcript, transcript variant 1 NM_000410.3, on the reference assembly contig NT_007592.16. The SNP would result in the change of 282nd amino acid in the protein NP_000401.1, encoded by the mRNA NM_000410.2, from cysteine to tyrosine.

GeneView section of the SNP entry rs1800562 for the mRNA NM_000410 alignment on the reference assembly contig NT_007592. The resulting amino acid change, 282nd amino acid in the protein NP_000401.1, from cysteine to tyrosine, is highlighted by a rectangle.

Thus, the query EST sequence contains a known SNP in the HFE gene that results in a cysteine to tyrosine change in the 282nd amino acid (Cys282Tyr) of the protein expressed by the longest transcript variant, variant 1 (Note 13). The next obvious step is to find out whether the SNP in the HFE gene is known to be associated with a disease phenotype.



gene model (contig mRNA transcript):	Contig Label	Contig	mrna	protein	mrna orientation	transcript	snp count
	GRCh38.p7	NT_007592.16	NM_000410.3	NP_000401.1	forward	plus strand	335, coding

Region	Chr. position	mRNA pos	dbSNP rs# cluster id	Hetero-zygosity	Validation	MAF	Allele origin	3D	Clinically Associated	Clinical Significance	Function	dbSNP allele	Protein residue	Codon pos	Amino acid	PubMed
	26092913	1005	rs1800562	0.072		0.0126		Yes		Pathogenic	missense	A	Tyr [Y]	2	282	
											missense	A	Tyr [Y]	2	282	
											missense	A	Tyr [Y]	2	282	
											missense	A	Tyr [Y]	2	282	
											contig reference	G	Cys [C]	2	282	
											contig reference	G	Cys [C]	2	282	
											contig reference	G	Cys [C]	2	282	
											contig reference	G	Cys [C]	2	282	

Fig. 12. Gene Model (Cotig mRNA Transcript)

Step 5: Determine Whether the HFE Gene Variant is Known to Cause a Disease Phenotype

To determine whether the Cys282Tyr amino acid change is linked to a phenotype, we will access the OMIM database. Make the Genes_seq map a master map by clicking the arrow at the top of the Genes_seq map. Click on the OMIM link next to the HFE gene.

Allele description

NM_000410.3(HFE):c.845G>A (p.Cys282Tyr)

Gene:	HFE:homeostatic iron regulator [Gene - OMIM - HGNC]
Variant type:	single nucleotide variant
Cytogenetic location:	6p22.2
Genomic location:	Chr6: 26092913 (on Assembly GRCh38) Chr6: 26093141 (on Assembly GRCh37)
Preferred name:	NM_000410.3(HFE):c.845G>A (p.Cys282Tyr)
HGVS:	NC_000006.12:g.26092913G>A NG_008720.2:g.10633G>A NM_000410.3:c.845G>A ...more
Protein change:	C282Y; Cys282Tyr
Links:	UniProtKB: Q30201#VAR_004398 OMIM: 613609.0001; dbSNP: rs1800562
GMAF:	0.0126(A), 1800562
NCBI 1000 Genomes Browser:	rs1800562

Fig. 13. Determine the HFE Gene Variant is Known to Cause a Disease Phenotype

This takes us to the OMIM report for the HFE gene. It describes the relationship between the mutations in the HFE gene and the hemochromatosis phenotype. Click the Allelic Variants "View list" in the side blue bar to get information about the mutant proteins from patient.

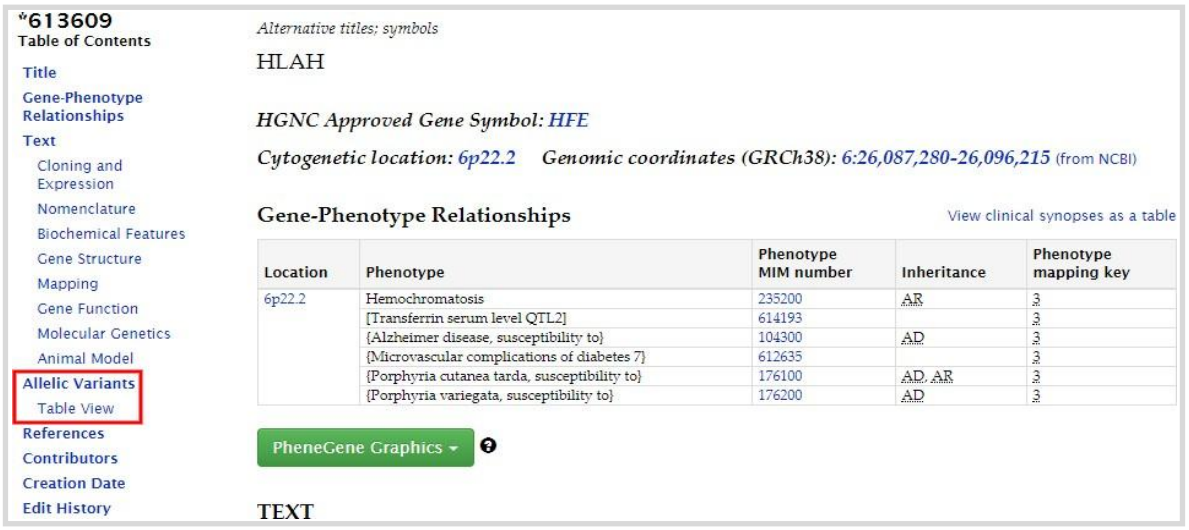


Fig. 14. Gene Phenotype Relationships

One variant, Cys282Tyr, is reported to cause the hemochromatosis phenotype. The query EST contains a known variation that would lead to the expression of the Cys282Tyr variant protein associated with the hemochromatosis phenotype.

Fig. 15. The Cys282Tyr change in the HFE protein is associated with hemochromatosis

Allelic variants list section from the Online Mendelian Inheritance in Man report for the HFE gene. The Cys282Tyr variant, highlighted by a rectangle, is reported to be associated with hemochromatosis.

V. CONCLUSION

This project describes the steps needed to identify the gene producing an EST obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene, and find SNP-associated phenotypes. The query EST sequence was found to align to contig NT_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).The query EST was found to align to the HFE gene. The query EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT_007592.14) that results in the Cys282Tyr change in the hemochromatosis protein expressed by the longest HFE mRNA variant. The Cys282Tyr change in the HFE protein is associated with hemochromatosis.

ACKNOWLEDGEMENTS

This work was supported by Dept. of Information and Communication Technology, Islamic University, Kushtia-7003, Bangladesh.

REFERENCES

[1]. Kitts P, McEntyre J, Ostell J, "Genome assembly and annotation process.", The NCBI Handbook. National Library of Medicine (US), NCBI; Bethesda, MD: 2002–2005.
[2]. Wheeler DL, Barrett T, Benson DA, et al, "Database resources of the National Center for Biotechnology Information. Nucleic Acids", Res. 2006; 34:D173–D180. [PMC free article] [PubMed].
[3]. Altschul SF, Madden TL, Schaffer AA, et al, "Gapped BLAST and PSI-BLAST: a new generation of protein database search program. Nucleic Acids", Res. 1997; 25:3389–3402. [PMC free article] [PubMed].
[4]. Madden T. ,"The BLAST sequence analysis tool.", In: McEntyre J, Ostell J, editors, "The NCBI Handbook. National Library of Medicine (US)", NCBI; Bethesda, MD: 2002–2005.
[5]. Dombrowski SM, Maglott M, "Using the Map Viewer to Explore Genomes", In: McEntyre J, Ostell J, editors, "The NCBI Handbook. National Library of Medicine (US)", NCBI; Bethesda, MD: 2002–2005.
[6]. Kitts A, Sherry S, "The single nucleotide polymorphism database (dbSNP) of nucleotide sequence variation", In: McEntyre J, Ostell J, editors, "The NCBI Handbook. National Library of Medicine (US)", NCBI; Bethesda, MD: 2002–2005.
[7]. Maglott D, Amberger JS, Hamosh A, "Online Mendelian Inheritance in Man (OMIM): a directory of human genes and genetic disorders.", In: McEntyre J, Ostell J, editors, "The NCBI Handbook. National Library of Medicine (US)", NCBI; Bethesda, MD: 2002–2005.
[8]. Zhang Z, Schwartz S, Wagner L, Miller W, "A greedy algorithm for aligning DNA sequences", J Comput Biol. 2000; 7:203–214.
[9]. Jonathan Pevsner, "Bioinformatics and Functional Genomics."
[10]. M. Lesk, "Introduction to Bioinformatics."
[11]. Cymbia Gibas & Jambeck, "Bioinformatics Computer skill."
[12]. Devid W. Mo, "Bioinformatics: Sequence and Genome Analysis."
[13]. Philip Compeau, Pavel Pevsner, "Bioinformatics Algorithms: An Active Learning Approaches."

Md. Islamul Haque." Analysis of Human Genome Sequences to Identify Disease Genes." IOSR Journal of Engineering (IOSRJEN), vol. 09, no. 10, 2019, pp. 55-66