Introduction to Bioinformatics - The Molecular Diagnosis and Treatment of Cancer

Objective

The information that is critical to directing the function of a cell is encoded in the DNA sequence of its genes. DNA sequence is a succession of letters (G, T, C and A) that correspond to the order of bases within the DNA molecule. This sequence of bases dictates the amino acid sequence of the protein encoded by a gene. The amino acid sequence of a protein uniquely determines it's structure and function. The analysis of DNA sequence is a key skill in the field of Bioinformatics.

Cancer results when the control of cell growth is dysregulated. Dysregulation of cell growth is typically caused by the alteration of the DNA sequence that encodes a protein that controls cell growth. Such DNA sequence changes (called mutations) are typically

caused by carcinogens (factors that make cancer). For example, tobacco smoke contains a chemical carcinogen that can change a G base into a T base and thereby alter the sequence of DNA. The bad news is that carcinogens cause mutations and cancer, the good news is that knowledge of the nature of the mutation can direct selective molecular cancer therapies.

In this exercise, you will play the role of a Cancer Bioinformatician. Your job is to analyze a cancer at the DNA level and to suggest molecular therapies. In the year 2000, it cost 2 Billion dollars to determine the complete DNA sequence (a sequence of 3 billion bases !) of a normal human individual. In May 2014, the cost of sequencing the entire DNA of an individual decreased to about $1,000. So, it is now economical to sequence the DNA of human cancer cells to find out what has gone wrong and to predict molecular therapies

In this introductory exercise, a preliminary analysis has already been performed to identify sequences in a lung tumor that are different from the normal human sequence. By answering the following questions you will determine whether an example of this altered sequence could have caused lung cancer.

After you complete this introductory exercise, you will be given real patient case studies and will be challenged to suggest molecular therapies.

Here is one atypical DNA sequence that was obtained from the complete sequencing of tumor from a lung cancer patient :

GTAGTTGGAGCTGTTGGCGTAGGCAAGAGT

Question - What human gene does this sequence correspond to ?

(Hint:

Compare this sequence with all the known human DNA sequences that are stored in the federal data base. To do this use the BLAST program (see appendix) which searches for sequences which best align to your input sequence)

Question: Has this gene been mutated ?

(Hint :

Look closely at the alignment , does it match exactly with the normal sequence (the normal human sequence is the sequence in the data base)

If there is a mutation, what carcinogen might have caused this mutation?

(Hint - Is there a known carcinogen that would create this particular base change ? Googling is useful here.

Was the patient exposed to this Carcinogen ?

(Hint :

What question might you ask the patient to determine this ?)

If the gene is mutated , what effect would this mutation have upon the structure of the protein encoded by the gene ?

(Hint - Does the mutation in DNA sequence change the amino acid that it encodes ? Use the codon wheel in the appendix)

Does this mutation change the function of the protein , deregulate cell growth and thereby cause cancer ?

(Hint - Is it known whether this change in the structure of the encoded protein can deregulate cell division and growth? This can be investigated by Googling the question.)

How would you develop therapies that selectively target this cancer ? In this case, selectively means therapies that would destroy tumor cells but not harm normal cells.

(Hint -The unique shape of a mutant protein should be a target for selective molecular therapeutics)

APPENDIX

1) To identify a gene that corresponds to a given DNA sequence.

Google NCBI

The federal NCBI site will be top of the list, click on it!

Next click on the "Analyze section"

Next look at the tools section and click on Basic Local Alignment Tool (BLAST), next, under basic blast, click on "nucleotide blast".

Then input the sequence that you want to analyze into the query box.

First type the "greater than" symbol followed by the name of your sequence, for example, "My test sequence ". Press Return and in the following line, type in your DNA sequence. It should look like this:

> My test sequence
NNNNNNNNNNNNNNNNNNNNN (where N can be A G C or T)

Set the appropriate DNA sequence database. For example, if you know the sequence is human use the human genomic plus transcript. If you want to check the sequence against all known sequences from all species use nr.

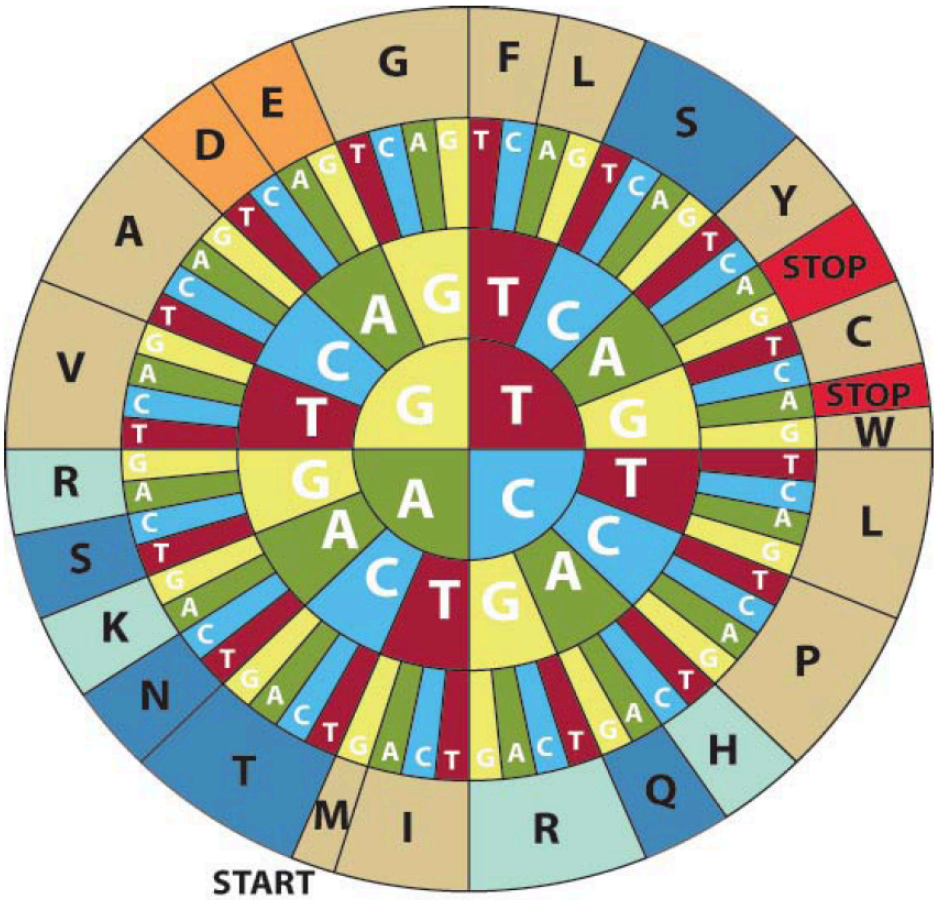
Then... click on the BLAST button and wait for results.

The gene that most matches the input sequence will appear first in a ranked list. Further information on the gene can be obtained by clicking on the relevant icons.

2) How to determine the amino acid encoded by a codon

Use the codon wheel to translate DNA codons into amino acids.

To decode a codon find the first letter of your sequence in the inner circle and work outwards to see the coressponding amino acid. For example: CAT codes for H (Hisitidine)



Amino acid code

A - Alanine	G - Glycine	M - Methionine	S - Serine
C - Cysteine	H - Hisitidine	N - Asparagine	T - Threonine
D - Aspartic Acid	I - Isoleucine	P - Proline	V - Valine
E - Glutamic acid	K - Lysine	Q - Glutamine	W - Tryptophan
F - Phenylalanine	L - Leucine	R - Arginine	Y - Tyrosine