Lab 1

BST 280

Oct 31, 2017

Note: Most info in summary can be found in lecture ppt. But I posted page number in the summary refers to textbook "Introduction to Genomics" in case if you need further details.

Summary of lectures:

Lecture 1:

1. Basics of Molecular Biology
    a. Mol bio in 7 words
    b. Central dogma
2. Sanger sequencing (p.96)

Sanger sequencing is a method of DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication

3. Human Genome Project
    a. Whole genome sequencing strategies
        i. Sequence tagged connectors
        ii. Shotgun sequencing
    b. Why Eukaryotic genomes are larger
    c. HGP
    d. Other sequencing projects
        i. (HapMAP, ENCODE, Cancer genome, p.151)

Lecture 2:

1. Genomics and Evolution
    a. Comparative Genomics
    b. different types of selection
    c. Advantages of WGS in population genetics
2. Metagenomics
    a. HMP
    b. strategy for sequencing-based microbial community identification
        i. 16sRNA used for bacteria
        ii. OTU Calling

Lecture 3:

Summary for resources:

1. PubMed : It was primarily designed to provide the access to references and abstracts from biomedical and life sciences journals. However, PubMed provides links that allow to access the full-text journal articles at Web sites of participating publishers

2. Bookshelf : The NCBI Bookshelf is a collection of freely accessible, downloadable, on-line versions of selected biomedical books
3. Nucleotide : The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.
4. Gene : Gene has been implemented at NCBI to characterize and organize the information about genes. It serves as a major node in the nexus of genomic map, expression, sequence, protein function, structure and homology data. A unique GeneID is assigned to each gene record that can be followed through revision cycles
5. Protein : It maintains the text record for individual protein sequences, derived from many different resources such as NCBI Reference Sequence (RefSeq) project, GenbBank, PDB and UniProtKB/SWISS-Prot. Protein records are present in different formats including FASTA and XML and are linked to other NCBI resources. Protein provides the relevant data to the users such as genes, DNA/RNA sequences, biological pathways, expression and variation data and literature.
6. Genome : This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.
7. SNP : free public archive for genetic variation within and across different species developed and hosted by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI). Although the name of the database implies a collection of one class of polymorphisms only (i.e., single nucleotide polymorphisms (SNPs)), it in fact contains a range of molecular variation: (1) SNPs,Bio (2) short deletion and insertion polymorphisms (indels/DIPs), (3) microsatellite markers or short tandem repeats(STRs), (4) multinucleotide polymorphisms (MNPs), (5) heterozygous sequences, and (6) named variants.
8. BLAST : In bioinformatics, BLAST for Basic Local Alignment Search Tool is an algorithm for comparing primarybiological sequence information, such as the amino-acid sequences of different proteins or the nucleotides ofDNA sequences. A BLAST search enables a researcher to compare a query sequence with a library ordatabase of sequences, and identify library sequences that resemble the query sequence above a certain threshold
9. OMIM : continuously-updated catalog of human genes andgenetic disorders and traits, with a particular focus on the gene-phenotype relationship. OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.

Finding out more about the BRCA2 gene

NCBI

1. Search “brca2” in the NCBI homepage
2. Under the “Genes” heading, click on “Gene”
3. Click on “Search Gene for brca2 as a symbol”

4. Click on BRCA2 with description “BRCA2, DNA repair associated [Homo sapiens (human)]”

Questions:
Where is the BRCA2 gene located? *(format with ‘q’ in the middle)*

How many different transcripts of BRCA2 are shown in the sequence viewer? What are their RefSeq accession numbers?

What are the RefSeq accession number for protein?

What chromosome is the BRCA2 homolog on Norway Rat(Rattus norvegicus) located?

5. Go to the SNP link on the right panel “Related Information”.
6. Click on the first result (in this case rs199250601)

What type of mutation is it? What does that mean?

Use NCBI to learn about SNP

In this lab activity, we will like to learn more about rs2273535 and we will use various tools to obtain information.

1. Find SNPs in the neighborhood rs2273535 by dbSNP

Using BLAST to get Gene from Sequence

MC4R
NG_016441.1

1. Please give the Gene Name and a GenBank Accession of the following gene.

```
AAAGCAACGCTCAGGCTGGAAACAGAAGCTTCCGAGAGGCAGCCGATGTGAGCATGTGCGCACAGATTCG
TCTCCCAATGGCATGGCAGCTTCAAGGAAAATTATTTTGAACAGACTTGAATGCATAAGATTAAAGTTAA
AGCAGAAGTGAGAACAAGAAAGCAAAGAGCAGACTCTTTCAACTGAGAATGAATATTTTGAAGCCCAAGA
TTTTAAAGTGATGATGATTAGAGTCGTACCTAAAAGAGACTAAAAACTCCATGTCAAGCTCTGGACTTGT
GACATTTACTCACAGCAGGCATGGCAATTTTAGCCTCACAACTTTCAGACAGATAAAGACTTGGAGGAAA
TAACTGAGACGACTCCCTGACCCAGGAGGTTAAATCAATTCAGGGGGACACTGGAATTCTCCTGCCAGCA
TGGTGAACTCCACCCACCGTGGGATGCACACTTCTCTGCACCTCTGGAACCGCAGCAGTTACAGACTGCA
CAGCAATGCCAGTGAGTCCCTTGGAAAAGGCTACTCTGATGGAGGGTGCTACGAGCAACTTTTTGTCTCT
CCTGAGGTGTTTGTGACTCTGGGTGTCATCAGCTTGTTGGAGAATATCTTAGTGATTGTGGCAATAGCCA
AGAACAAGAATCTGCATTCACCCATGTACTTTTTCATCTGCAGCTTGGCTGTGGCTGATATGCTGGTGAG
CGTTTCAAATGGATCAGAAACCATTGTCATCACCCTATTAAACAGTACAGATACGGATGCACAGAGTTTC
ACAGTGAATATTGATAATGTCATTGACTCGGTGATCTGTAGCTCCTTGCTTGCATCCATTTGCAGCCTGC
TTTCAATTGCAGTGGACAGGTACTTTACTATCTTCTATGCTCTCCAGTACCATAACATTATGACAGTTAA
GCGGGTTGGGATCATCATAAGTTGTATCTGGGCAGCTTGCACGGTTTCAGGCATTTTGTTCATCATTTAC
TCAGATAGTAGTGCTGTCATCATCTGCCTCATCACCATGTTCTTCACCATGCTGGCTCTCATGGCTTCTC
TCTATGTCCACATGTTCCTGATGGCCAGGCTTCACATTAAGAGGATTGCTGTCCTCCCCGGCACTGGTGC
CATCCGCCAAGGTGCCAATATGAAGGGAGCGATTACCTTGACCATCCTGATTGGCGTCTTTGTTGTCTGC
TGGGCCCCATTCTTCCTCCACTTAATATTCTACATCTCTTGTCCTCAGAATCCATATTGTGTGTGCTTCA
TGTCTCACTTTAACTTGTATCTCATACTGATCATGTGTAATTCAATCATCGATCCTCTGATTTATGCACT
```

CCGGAGTCAAGAACTGAGGAAAACCTTCAAAGAGATCATCTGTTGCTATCCCCTGGGAGGCCTTTGTGAC
TTGTCTAGCAGATATTAAATGGGGACAGAGCACGCAAT