6/3/24

PAGE:
DATE: / /

Notes

Q] Unit 1 Elective

1] Genes, Introns & Exons

→ i] Genes are functional unit of heredity as they are made up of DNA, the chromosome is made up of DNA containing many genes.

2] Every gene comprises of the particular set of instructions for a particular function or protein coding 3] There are about 30,000 gene in each cell of human body. Gene comprise of alternating pattern of introns & exons, promoter, open reading frame & splice sites. which collectively contributes for protein building

a] Introns:-

i) An Introns are is a region that resides within a gene but does not remain in the final mature mRNA molecule following transcription of that gene & thus not code for any aminoacid that make up the protein.

ii) The protein coding sequences contains both exons & introns wherein introns are non coding sequences whereas exons are coding sequences.

iii) Introns are removed during the process called splicing so only exons are included in the mature mRNA.

iv) Introns are much longer than exons.

v) Introns may contain sequences that regulate how genes are expressed or transcribed & how mRNA is processed.

Function:-
Transcription terminator
Genome organization
Transcription initiation

b] Exons:-

i) A part of gene encode for final mature RNA produced by that gene after introns have been removed by splicing

ii] Exons usually include both the 5' & 3' untranslated regions of mRNA, which contain start & stop codons, in addition to any protein coding sequences

iii] There are 8 exons & 7 introns per gene

Function:

(1) ~~Take~~ Transcription proceed ahead

(2)

iv) Exons are coding sequences that code for a protein's amino acid sequence

v) After post transcriptional alteration the exons are translated into mature mRNA

vi) These are highly conserved sequences meaning they don't change much over time

Functions of exons:-

1) Exons are the parts of a gene that code for a protein

2) Exons are mRNA coding regions that code for amino acids

3) Various exons code for different protein domains

4) A single exon or numerous exons spliced together can encode the domain

GeneAlign is a coding exon prediction tool for Predicting protein coding genes by measuring the homologies between a seq of a genome & related seq which have been annotated of other genome

5] When exons on sister chromosomes are switched during recombination, exon shuffling occurs

6] This enables creation of new genes

7] Exons also allow for alternative splicing which allows several proteins to be translated from same gene

8] Introns are deleted from mature mRNA and exons are joined together

9] After introns have been eliminated by RNA splicing, an exon is any component of a gene that will constitute a part of the final mature RNA generated by that gene

Note:- The tool used for exon prediction is GeneAlign

Q] ORF Finder:

(1) In molecular genetics, an open reading frame is the part of a reading frame that has the ability to be translated. An ORF is a continuous stretch of codons that begins with a start codon and ends at a stop codon. An ATG codon within the ORF may indicate where translation starts. In other words we can say that the region of a nucleotide (codon that starts from an initiation codon and ends with a stop codon is called ORF.

(2) The CDS (coding Sequence) is the actual region of DNA that is translated to form proteins while the ORF may contains introns as well. The CDS refers to those nucleotides (concatenated exons) that can be divided into codons which are actually translated into

aminoacids by process of translation.

ORF finder

1) ORF finder is a program or graphical analysis tool available at NCBI website which searches for open reading frames (ORFs) in the DNA sequence you enter.
2) The program or tool returns the range of each ORF, along with its protein translation.
3) Use ORF finder to search newly sequenced DNA for potential protein encoding segments. This tool identifies all open reading frames using the standard or alternative genetic codes.

Importance of ORFs

1. ORFs is a piece of evidence to assist in gene prediction
2. Long ORF's are often used, along with other evidence to initially identify candidate protein coding regions or functional RNA coding regions in a DNA sequence.

Q Gene prediction.

1) Gene prediction by computational methods for finding the location of protein coding regions is one of the essential issues in bioinformatics
2) The gene sequencing of a gene is productive only when it is analysed & predicted correctly.
3) Gene prediction is carried out to identify the structure of genes in order to differentiate protein coding genes from non coding regions, and to identify promoters & other regulatory elements
4) Gene prediction basically means locating genes along genome. Also called genefinding. It refers to the process of identifying the regions of genomic DNA that encode genes.
5) This includes protein coding genes, RNA genes & other functional elements such as regulatory genes.

Importance of Gene Prediction

1) Helps to annotate large, contiguous sequences.
2) It provides information on the evolution of genes, speciation & evolution of species.
3) It gives an understanding of the structure & function of genomes of different organism
4) distinguish between coding & noncoding regions of a genome

Types of approaches

The current gene prediction method can be classified.

1) AB initio based (prediction based on given sequence only)

a) The first feature is the existence of gene signals which include start & stop codons, introns, splice signals, transcription factor binding sites, ribosomal binding site.

b) Second feature - used by ab initio algorithms is gene content which is statistical description of coding regions. It has been observed that nucleotide complementary & statistical pattern of coding region tend to vary significantly from non coding region. Thus unique features can be detected by applying probabilistic models such as Markov rules.

2) Homology based.

Predict based on significant matches of the query sequence with sequences of known genes.

Eg:- If a translated DNA is found to be similar to a known protein family from a database search, this can be strong evidence that the region codes for a protein. Also if possible exons

DATE : / /

of a genomic DNA region match a sequenced cDNA, this also provides experimental evidence of a coding region

Promoters :

1) Promoter sequences are DNA sequences that define where transcription of a gene by RNA polymerase begins.
2) Promoter sequences are typically located directly upstream or at the 5' end of the transcription initiation site.
3) RNA polymerase and the necessary transcription factors bind to the promoter sequence & initiate transcription
4) Promoter sequences define the direction of transcription and indicate which DNA strand will be transcribed, this strand is known as sense strand.
5) Many eukaryotic genes have a conserved promoter sequence called the TATA box, located 25 to 35 base pairs upstream of the transcription start site. Transcription factors bind to the TATA box and initiate the formation of the RNA polymerase transcription complex, which promotes transcription.

Splice sites

A genetic alteration in the DNA sequence that occurs at the boundary of an exon & intron known as splice site.

Most commonly RNA sequence that is removed

begins with the dinucleotide GU nucleotide at its 5' end & ends with AG at its 3 end. The GU-AG rule originally called GT-AG rule interms of DNA seq.

Why splice sites important
Mutation in these sequence may lead to retention of large segments of intronic DNA by the mRNA or the entire exons being spliced out of the mRNA. these changes could result in production of a nonfunctional protein
Prediction of splice site more accurate, localization of splicesites can substantially help to explore the structure of genes

Regulatorsite
1) Regulator gene encodes for a DNA binding protein that act as a repressor.
2) A variety of DNA regulatory elements are involved in the regulation of gene expression & rely on the biochemical interactions involving DNA, the cellular proteins that make up chromatin & transcription factors. Promoters & enhancers are the primary genomic regulatory components of gene expression
Role of regulatory gene

DATE : / /

Q) Microarray technology

1) Microarray technology is a general laboratory approach that involves binding an array of thousands to million of known nucleic acid fragments to a solid surface, referred to as a "chip".

2) The chip is then bathed with DNA or RNA isolated from a study sample (such as cells or tissue) complementary base pairing between the sample and the chip immobilized fragments produce light through fluorescence that can be detected using a specialized machine

3) Microarray technology can be used for a variety of purposes in research and clinical studies, such as measuring gene expression & detecting specific DNA sequences (eg single nucleotide polymorphisms, or SNPs).

4) Microarrays were revolutionary. They really allow genomic analysis without sequencing, which tremendously reduced the cost of doing large studies across a wide area of biology & biomedicine.

Goals:-

1) Gene expression or the amount of gene product, RNA, from any given gene that you found in a cell.

2) Single nucleotide polymorphisms or SNPs, which were useful for genome wide association studies, or GWASS.

Application

1) Microarray is one of the most recent advance being used for cancer research it provides assistance in pharmacological approach to treat various diseases including oral lesions.
2) Microarray helps in analyzing large amount of samples which have either been recorded previously or new samples; it even helps to test the incidence of a particular marker in tumor
3) Microarray provides a basis to genotype thousands of different loci at time, which is useful for association & linkage studies to isolate chromosal region related to a particular disease
4) Genome microarrays have been used for comparative genomic hybridization. In this technique, genomic DNA is fluorescently labeled & used to determine the presence of gene loss or amplification.

23/8/24

Aparna Ma'am

Note Continue

PAGE No.
DATE / /

Genome Reference Sequence

→ (1) A human genome reference sequence is an accepted representation of the human genome sequence that is used by researchers as a standard for comparison to DNA sequences generated in their assembling and updating such reference sequences aim to provide the highest quality, best possible consensus representations of the sequence and structural diversity found in the human genome among populations.

(2) The genome reference sequence provides general framework and is not the DNA sequence of a single person

(3) Reference genome assemblies looked after by the Genome reference consortium

Example

(1) Zebrafish genome reference assembly is a high quality clone based assembly that has gone through decade of continuous improvements. These improvements were based on a huge amount & variety of generated sequencing & mapping data & applied both through automated edits & manual interventions.

Tool used for Genome reference sequence

(1) Refseq - The reference sequence (Refseq) collection provides a comprehensive, integrated, non redundant, well annotated set of sequences, including genomic DNA, transcripts & proteins.

RefSeq sequences form a foundation for medical, functional & diversity studies. They provide a stable reference for genome annotation, gene identification & characterization mutations & polymorphism analysis.

Advantage

① The ability to characterize particular genes, or gene families, that are relevant to species specific conservation.

② The reference genome sequencing used for exploratory analysis of gene families involved in key biological process of threatened species such as immunity, reproduction & behaviour.

Integrated Genomic Maps (Intro).

① Genetic maps provide the outline & physical maps provide the details. It is easy to understand why both types of genome mapping techniques are important.

② Information obtained from each technique is used in combination to study the genome. Genome mapping is being used with different model organisms that are used for research.

③ Genome mapping is still an ongoing process and as more advanced techniques are developed, more advances are expected.

④ Mapping information generated in laboratories all over the world is entered into central databases, such as GenBank at the NCBI.

1) Physical maps.
2) Genetic maps

Gene expression profiling

1) The central dogma of biology describes the method by which information is taken from genes & used to create proteins. DNA transcription produces RNA, then RNA translation makes proteins. This process is known as gene expression and all life forms use it to create the building blocks of life from genetic information.

2) A variety of DNA regulatory elements are involved in the regulation of gene expression and rely on the biochemical interactions involving DNA, the cellular proteins that make up chromatin & transcription factors. Promoters & Enhancers are the primary genomic regulatory components of gene expression.

3) Gene expression profiling measures which genes are being expressed in a cell at any given moment. This method can measure thousands of genes at a time, some experiments can measure the entire genome at once.

4) Gene expression profiling measure mRNA levels, showing the pattern of genes expressed by a cell at the transcription level. This often means measuring relative mRNA amounts in two or more experimental conditions, then assessing which conditions resulted in specific genes being expressed.

(5) Different techniques are used to determine gene expression. These include DNA microarrays & sequencing technologies.
(6) The former measures the activity of specific genes of interest and the latter enables researchers to determine all active genes in a cell.
(7) A gene expression profile tells us how a cell is functioning at a specific time. This is because cell gene expression is influenced by external & internal stimuli including whether the cell is dividing, what factors are present in cell's environment, the signals it is receiving from other cells, & even the time of day.

Application

(1) Gene expression profiling used by a variety of biomedical researchers, from molecular biologists to environmental toxicologists. This technology can provide accurate information on gene expression, towards countless experimental goals.
(2) Gene expression profiling enables you to investigate the effects of different conditions on gene expression by altering the environment to which the cell is exposed & determining which genes are expressed.
(3) Gene expression profiling is often used in hypothesis generation. If very little is known about when & why a gene will be expressed, expression profiling under

different conditions can help to design a hypothesis to test in future experiments.

④ Gene profiling can also investigate the effect of drug like molecules on cellular response. You could identify the gene markers of drug metabolism, or determine whether cells express genes known to be involved in response to toxic environments when exposed to drug.

⑤ Gene profiling can also be used as a diagnostic tool. If cancerous cells express higher levels of certain genes, and these genes code for protein receptor, this receptor may be involved in the cancer, and targeting it with a drug might treat the disease. Gene expression profiling might then be a key diagnostic tool for people with this cancer.

Q Orphan GPCR

① The superfamily of G-protein coupled receptors (GPCRs) included at least 800 seven transmembrane receptors that participate in diverse physiological & pathological functions.

② GPCR represent the largest superfamily & most diverse grp of mammalian transmembrane protein.

③ GPCR play major role in numerous physiological & pathological roles in transducing extracellular signals into intracellular effector pathways through the activation of heterotrimeric G protein by binding to a broad range of ligands

④ Human GPCRs can be divided into five main families on the basis of phylogenetic criteria; Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 & Secretin.

⑤ The first GPCR to be identified was rhodopsin in 1878. It was later proven that rhodopsin consists of the GPCR protein opsin and a reversibly covalently bound cofactor, retinal.

⑥ After completion of the human genome sequence in 2004, the number of human GPCRs increased to about 800 based on the screening approaches, such as low stringency hybridization, PCR derived methods & bioinformatic analyses.

⑦ Besides the olfactory receptor family more than 140 GPCRs have not yet been linked to endogenous ligands. These are the so called orphan GPCRs

⑧ These orphan GPCR represent vast opportunities for discovering new therapy for disease that has been intractable that targeting the well known GPCR and other protein family.

⑨ Approach for identification of oGPCR
a) Screening of putative small molecule & peptide ligands
b) Reverse pharmacology
c) Use of bioinformatics to predict candidate ligands

Q) Reverse Pharmacology

→ (1) Reverse Pharmacology is a science of integrating documented experimental bits, into leads by transdisciplinary exploratory studies & further developing into drug candidates by experimental research.

(2) Reverse Pharmacology enhances the connection, communication & collaboration between modern science & technology with traditional medicine & modern ~~science~~ biomedicine.

(3) Classical drug discovery process is an expensive & time consuming process whereas RP is an economical time sparing & has least bottlenecks. It allows understanding the mechanism of drug action at multiple levels & helps in optimizing the safety, efficacy & acceptability of the leads from natural products

(4) RP utilizes traditional knowledge of medicines to discover drugs and is also called as a path of pharmacology from the bedside to bench experiments.

(5) The aim of forward pharmacology is to enhance the desired physiological effect of a compound.

(6) Forward pharmacology (phenotypic based screening) involves first identifying the functional (phenotype) activity of a compound through cellular or animal models.

(7) Once knowing the physiological affect of the certain compound, only then the compound ligand & its derivatives are

Drug discovery can be divided into two process (a) Classical drug discovery (b) Reverse pharmacology

identified, purified & synthesized repectively & their binding capabilities with a target receptor are determined through biological assays / screenings.

(7) The most potent & selective ligands was identified as the new possible drug & further research is done with ligand.

(8) The aim of reverse pharmacology is to screen the natural products are screened against receptors / targets of known physiological function in order to determine functional activity.

(9) The aim of reverse pharmacology (target based screening) began with the growth of molecular biology and caused a paradigm shift in drug discovery worldwide.

(10) First potential ligands are screened through binding assay where the highly selective ligand that binds with the molecular target is identified. This is known as ligand fishing then this potential ligand (compound) undergoes functional studies (animal models) to significantly show the desired physiological effect.

There are three methods in Reverse Pharmacology

(1) library based approach - used in pharmaceutical & some large labs have

(2) Tissue based - Used in labscale basis (extract)

(3) Information based
- Identifies prospective ligands by database screening & testing them.

PAGE No.
Date / /

Deorphanization:-

(1) GPCRs are the most important & prominent family of pharmacological target in biomedicine

(2) The deorphanization of orphan GPCRs is the important mission in orphan GPCR research

(3) It is a process of identifying identification of ligands that are highly selective for orphan GPCR.

(4) In general the standard assay are radio-ligand binding, calcium flux, GTPx binding & modulation of CAMP levels 92, 93, 94, 95, 96, 97 & 98.

(5) With the development of molecular biology technologies several line of approach have been used for deorphanization. i.e one of them is according to the sequence & functional similarity, ligand of the identified receptor & are used to examine GPCR with identical sequence or domain

Q) Syntony & Gene order

1) In genetic the term syntony refers to two related concepts

a) To classical genetic syntony describes the physical co-localization of genetic loci on the same chromosome within an individual or species

b) In current biology syntony more commonly refers to colinearity. Ie. conservation of blocks of order within two sets of chromosomes that are being compared with each other. These blocks are referred to as syntenic blocks.

Uses:

1) Provides a framework in which conservation of homologous genes & gene order is identified between genome of different species. The availability of human & mouse genomes paired the way of algorithm development in large scale based on syntony mapping

2) Comparing two genomes reveals homologous sequences that reflect their evolutionary origins & subsequent conservation.

Gene order is much less conserved with gene sequence. Therefore syntenic relationships are normally carried out between relatively close lineages

Gene orders are the permutations of genome arrangement. A fair amount of research has been done trying to determine whether gene orders evolve according to a molecular clock hypothesis or in jumps (punctuated equilibrium).
Some research on gene orders in animals mitochondrial genomes reveals that the mutation rate of gene orders is not a constant in some degree.

Q) Edman's degradation.

(1) Edman's degradation is the process of purifying protein by sequentially removing one residue at a time from the amino acid end of peptide.
(2) To solve the problem of damaging the protein by hydrolysing condition.
(3) Pehr Edman created a new way of labelling & cleaving the peptide, where phenyl isothiocyanate was added this compound reacts as phenyl thiocarbonyl derivate with N-terminal.
(3) The N-terminal is then cleaved under less harsh conditions creating a cyclic compound of phenilthiohydantoin, PTH aminoacid this does not damage the rest protein & leave two constituents of the peptide.
(4) Edman sequencing is done but if composition of aa is known

Advantage
(1) the whole sequencing of the protein can be done without damaging the protein.

(2) Allows sequencing of protein in less time.

a) Shotgun proteomics:

1) Shotgun proteomics also known as 'Bottom up proteomics' is a widely used & mature technology for protein identification & characterization of their amino acid sequence along with posttranslational modifications (PTMs).

2) This technique requires the proteolytic digestion of proteins prior to mass spectrometry analysis.

3) Shotgun proteomics has been demonstrated to be a valuable tool for the identification of both large or small proteomes & protein complexes, enabling the discovery of previously unknown protein-protein interactions.

4) The most distinctive feature of shotgun proteomics is that it enables identify a wide range of proteins at the same time with minimal protein separation needed.

5) Shotgun proteomics refers to the use of bottom up proteomics techniques in identifying protein in complex mixtures using a combination of high performance liquid chromatography combined with mass spectrometry.

(6) The name is derived from shotgun sequencing of DNA which is itself named after rapid, expanding, quasi random firing pattern of shotgun.

(3) Targeted proteomics using SRM & data independent acquisition methods are often considered alternative to shotgun proteomics in the field of bottom up proteomics

Advantages

(1) Shotgun proteomics allows global protein identification as well as the ability to systematically profile dynamic proteome.
(2) It also avoids the modest separation efficiency & poor mass spectral sensitivity associated with intact protein analysis.

Disadvantages

(1) The dynamic exclusion filtering that is often used in shotgun proteomics maximizes the number of identified proteins at the expense of random sampling.
(2) This problm may be generated by the undersampling inherent in shotgun proteomics

Applications

(1) Shotgun proteomics can be used for functional classification or comparative analysis of these protein products
(2) It can be used in projects ranging from large scale whole proteome to focusing on a single protein family. It can be done in research labs or commercially

Q] Protein identification with antibody

→ (1) Antibodies are proteins synthesized by a animal in response to the presence of a foreign substance known as antigen

(2) This antibody have specific affinity for a particular region of antigen

(3) This region is termed as Epitope

(4) The antibody epitope interaction can be utilized for highly specific & sensitive detection of a protein that has been immobilized on a membrane, in a process termed as immunodetection

(5) The antibody that binds to the protein of interest is termed as the primary antibody

(6) The primary is applied to the membran & is allowed to bind to the target protein in order to locate the primary antibody & the protein of interest, a secondary antibody is required.

(7) The secondary antibody recognize & binds to all IgG antibodies because IgG antibody react specifically with the introduced protein & can be obtained from animal serum.

(8) It is important that secondary antibody used in a experiment is directed against IgG from the species of origin of the primary antibody.

Antibody based Protein detection techniques

① ELISA -

Enzyme linked Immunosorbent Assay is a method that is analogous to immunodetection of proteins on a membrane & it is used for quantitative assay of proteins in a sample.

② In ELISA, proteins are immobilized onto a solid support known as well plate this step is known as fixation.

③ A wash is given to the plates containing the proteins so that to remove the non specifically bound material.

④ A secondary antibody which will be specific for this protein will be added.

⑤ This secondary antibody ~~which~~ will be ~~by~~ is usually conjugated to an enzyme, that allows its detection by the chromogenic or chemiluminescent methods.

Advantages

① High sensitive & specific can detect antigens at the picrogram level in a very specific manner due to the use of antibody

② High throughput

③ Accessibility to tut various sample types.

Disadvantage

① Temporary readout - Detection is based on enzyme substrate reaction & therefore readout must be obtained in a short span of time,

② Limited antigen information

(2) Western blotting

1) It is a laboratory technique used to detect a specific protein in a blood or tissue sample.
2) The method involved the use of gel electrophoresis to separate the sample protein
3) The separated proteins are transferred out of the gel to the surface of the membrane
4) The membrane is exposed to an antibody specific to the target protein
5) Binding of the antibody is detected using a radioactive or chemical

Principle - The sample is separated by using electrophoresis gel
The proteins are then resolved & transferred onto a membrane of special paper. The membrane is then probed with an antibody specific protein of interest. Because the antibody is labeled with a molecule that we can visualize
We can see whether the protein of interest is expressed in this sample or not &
can also know the abundance of the protein