MEDTEC SHOOL

Course: Molecular and Computational Biology and Medical Genetics

Year (1st-2nd-3rd-4th-5th-6th): 2nd

Period (1st-2nd semester – annual): 2nd semester

Credits: 9

Objectives

The course will merge molecular biology and medical genetics with computational science and bioinformatics to give an up-to-date vision of molecular genetics mechanisms involved in human diseases, with a particular focus on genetic disorders. Besides the theorical part, large space will be given to practical sessions in which the students will learn how to use genomics data present on the web and how to analyze real DNA and RNA sequencing data.

The “Molecular biology” module will be focused on understanding structure, function, and turnover of macromolecules, structure and regulation of genes and genomes, the mutability of the genome and the mechanisms of DNA repair. The impact of recombinant DNA technologies on basic and applied biomedical research will also be illustrated.

The “Medical genetics” module will illustrate the types and extend of genetic variation seen in the human genome, and how this variation affects disease susceptibility. The module will particularly emphasize how molecular and population genetics have impacted on our understanding of the mechanisms of human disease, providing us with novel diagnostic and therapeutic strategies.

The “Computational science” module will focus on sequence alignment to support biomolecular sequence analysis with an attention on algorithms, complexity and available resources to be able to exploit computer-based data processing and analysis in the biomedical context. The module will also introduce the R programming environment and the design and implementation of pipelines of programs to prepare, manipulate and process data through practical examples, to enable students to use tools adopted in these and other courses.

The “Bioinformatics” module will introduce computational strategies and challenges in omics data analysis, focusing on genetic variation and transcription. Aspects of analytics for raw data processing will be illustrated highlighting their relatedness with technology. Fundamentals of data mining techniques applied to gene expression profiling will be presented along with tools for functional annotation.

Notes

- To be allowed to the CMBMD exam students must have passed the exam of “Cell Biology, Embryology and Histology”.
- Questions are always welcome during and after the lectures.
- We strongly encourage you to take notes during lessons, to share materials, and, when possible, to study together, helping each other.
- Slides of the lectures will be uploaded in the LMS platform, where additional material (video, papers, animations) will be posted by lecturers.

Prerequisites

To be allowed to the CMBMD exam students must have passed the exam of “Cell Biology, Embryology and Histology”.

Contents

Module: Molecular Biology (S. Duga & S. Bottaro)

The mutability of the genome

1. Causes of mutation
Topics: spontaneous and induced causes of mutation. Agents that induce mutation. Common mutations due to DNA replication.
Learning objectives:

- Describe examples of the main categories of DNA damaging agents
- Describe mechanisms by which DNA can be damaged
- Discuss the relationship between DNA damage and DNA mutation
- Discuss the balance between mutation-inducing mechanisms and DNA repair and its role in evolution

2. DNA repair mechanisms
Topics: Main mechanisms of DNA repair and related diseases.

- Discuss the cell responses to DNA damage
- Describe the main mechanisms of DNA repair and the specific damages they are apt to
- **Clinical drop**: Inherited diseases due to defects in the DNA repair systems

Recombinant DNA

3. Essentials of DNA recombinant technology
What is rDNA? Tools of rDNA technology, making a recombinant DNA molecule, DNA probes and hybridization, main applications of rDNA.
Learning objectives:

- to understand the principles of rDNA technology
- to understand the concept of genomic and cDNA library

- to be able to describe a cloning experiment
- to know what a molecular probe is and what is a hybridization experiment

4. Polymerase Chain Reaction (PCR) & DNA sequencing
Topics: The discovery of PCR, the principles of a PCR reaction, what can you do with PCR? Sanger sequencing, Next generation sequencing.
Learning objectives:
- Discuss why DNA is amplified during a PCR reaction
- Illustrate the main applications of PCR
- Describe the theory of Sanger sequencing
- compare Sanger sequencing and next generation sequencing (NGS)
- describe the concepts of clonal amplification and parallel sequencing
- list the main technologies for NGS
- discuss the concept of sequencing depth
- **Clinical drop:** how we can diagnose a genetic disease by PCR and DNA sequencing

5. Applications of recombinant DNA technology
Topics: Overview on the main approaches used to produce recombinant proteins of medical relevance by DNA recombinant methods.
Learning objectives:
- to understand how it is possible to produce recombinant therapeutics
- to know the basics of transgenic animal production technologies
- to understand the difference between transgenic, knock-out and knock-in animals
- **Clinical drop:** treating diabetes with recombinant insulin

6. Gene therapy
Topics: Overview on the main approaches used for gene therapy. Traditional approaches with their pros and cons and genome editing (CRISP/Cas9)
- to be able to describe the concept of gene therapy
- to understand the differences between in-vivo and ex-vivo gene therapy

The regulation of gene expression

7. General introduction of the multilayer regulation of gene expression
Topics: How cells modulate gene expression. Levels of gene expression regulation: from chemical modification of DNA to post-transcriptional and post-translational regulation. The principles of DNA-protein interaction.
Learning objectives:
- Discuss the importance of gene expression modulation in driving the processes of cellular differentiation and morphogenesis
- Discuss the principles of transcriptional regulation in prokaryotes and eukaryotes
- Describe the main protein modules interacting with DNA
- Explain the genomic organization and function of homeotic genes

8 The transcriptional regulation

Topics: The transcriptional regulation of gene expression in eukaryotes. Enhancers, silencers and insulators. The epigenetic regulation of gene expression.
Learning objectives:

- Understand the concept of in-cis and in-trans regulation
- Discuss the interaction between modulation by transcription factor and chromatin status
- Explain how epigenetic information influences gene expression

9. Post-transcriptional regulation
Attenuation. Alternative splicing, differential polyadenylation, RNA editing. Control of mRNA localization and stability. RNA interference.
Learning objectives:

- Explain the mechanism of attenuation in Bacteria
- Understand the importance of alternative splicing regulation in eukaryotic cells
- Describe examples of post-transcriptional regulation
- Understand the role on Nonsense-mediated mRNA decay in regulating gene expression and in modulating disease phenotypes
- Describe mechanisms regulating mRNA stability
- **Clinical drop:** Exploiting splicing modulation to treat inherited diseases

10. Noncoding RNAs
Topics: The non-coding RNA revolution. Overview on the main classes of non-coding RNAs. Post-transcriptional regulation by small RNAs in prokaryotic and eukaryotic organisms. Long non-coding RNAs: structural features and mechanism of action. Competing endogenous RNAs.
Learning objectives:

- Explain the importance of RNA-mediated gene expression regulation
- Describe the main classes of noncoding RNAs (small and long)
- Illustrate the mechanism of RNA interference
- Describe microRNA biogenesis and mechanism of action
- **Clinical drop:** miRNA therapeutics
- Illustrate the features of long noncoding RNAs
- Recognize the complexity of mechanisms of action of long noncoding RNAs
- **Clinical drop:** Prostate cancer diagnostic test based on measurement of a specific noncoding RNA

Genes and genomes

11. Human genome organization
Nuclear and mitochondrial genomes. The C-value paradox. Classes of repetitive DNA elements. Gene families. Organization of repetitive elements in the human genome. Segmental duplications.
Learning objectives:

- discuss the differences between the nuclear and mitochondrial genomes
- list the main classes of genomic sequences, their function and their origin
- discuss the peculiar characteristic of the human genome that may be related to the

human brain size
- describe the concept of psudogene and transposon

Module: Medical Genetics (R. Asselta)

12. Genetic variation
Topics: Polymorphisms and mutations: classification and functional consequences.
Learning objectives:
- Describe the different types of point mutations
- Discuss differences between germinal and somatic mutations, and between polymorphisms and mutations
- Describe the functional consequences of the different types of mutation at the RNA and protein levels

13. Mendelian pedigree patterns and their complications
Topics: How to build up a pedigree. Examples of autosomal dominant, autosomal recessive, X-linked, Y-linked and mitochondrial inheritance. Incomplete penetrance, expressivity, male lethality, de-novo mutations, mosaicism, phenocopies, complementation, and mitochondrial inheritance.
Learning objectives:
- Being able to collect a genetic anamnesis and to draw a pedigree
- Describe the main patterns of inheritance
- Discuss the main complications of the classical Mendelian pattern of transmission

14. Genetic mapping of Mendelian traits
Topics: Polymorphisms as a tool for genetic mapping, recombinants and non-recombinants, two-point and multi-point mapping, the concept of LOD score.
Learning objectives:
- Understand how polymorphisms can be used to trace Mendelian traits
- Understand the meaning of LOD score in genetic analysis

15. Prototypic Mendelian diseases: cystic fibrosis and thalassemia
Topics: The quest for the gene causing cystic fibrosis. The organization of the globin loci. Sickle cell anemia. Alpha and beta thalassemia.
Learning objectives:
- Describe a typical approach to identify disease genes
- Describe the gene arrangement of the globin loci
- Illustrate the molecular mechanisms underlying sickle cell anemia and thalassemia
- **Clinical drop**: neonatal screening and genetic diagnosis of cystic fibrosis

16. Dynamic mutations
Topics: The concept of microsatellite instability. Trinucleotide repeat expansion disorders and their classification. Anticipation. Fragile X syndrome, Huntington disease, myotonic dystrophy.
Learning objectives:
- Describe the general features of trinucleotide repeat expansion disorders

- Understand the concept of anticipation
- Illustrate the molecular characteristics of Fragile X syndrome, Huntington disease, and myotonic dystrophy

17. Population genetics
Topics: Allele frequencies in populations. Hardy-Weinberg equilibrium. Genetic drift, population bottlenecks and founder effects. Genetic selection. The concept of the heterozygote advantage.
Learning objectives:
- Understand the applications and limitation of Hardy-Weinberg law
- Distinguish heterozygous advantage from founder effect
- Being able to calculate heterozygote frequency from disease prevalence

18. Genetics of complex (multifactorial) diseases
Topics: The polygenic and multifactorial nature of common diseases. Estimating the contribution of genetic and environmental factors. Linkage disequilibrium. Principle of allelic association.
Learning objectives:
- Understand the difference between dichotomic and continuous traits
- Describe the concept of linkage disequilibrium.
- Understand the difference between linkage analysis and association analysis

19. NGS for the identification of the molecular basis of Mendelian diseases. From genome sequence to the causative variant
From candidate gene analysis to whole genome sequencing. Targeted resequencing vs whole genome sequencing. Exome sequencing: flowchart and examples on how to design a study. The big challenge to fish out the pathogenic variants. How to find a novel disease gene: from theory to practice.
Learning objectives:
- Compare different NGS-based approaches for the molecular diagnosis of inherited disease
- Illustrate what is an exome and what are the critical steps in a standard exome-sequencing experiment
- Understand the hypotheses underlying different study design choices

20. Genome-wide approaches to complex diseases
LD structure of human genome. Genome-wide association studies (with a focus on COVID-19). Missing heritability. Burden of rare variants.
Learning objectives:
- Describe the modalities for conducting a genome-wide association analysis
- List possible genetic approaches to fill the gap of the missing heritability

21. What we have learned from omics approaches
Genomic consortia, Databases. From phenotypes to mutations, to mutations without a phenotype: redefining the classical concept of Mendelian diseases. The problem of incidental findings.
Learning objectives:

- Illustrate the key experiments that lead to the discover of the LD structure of human genome
- List the principal consortia and publicly available databases with genomic data

22. Epigenetics

Introduction to epigenetics. DNA methylation during development & Genomic imprinting. Chromatin structure & spatial organization of chromosomes. Regulatory RNAs: X-inactivation and sex determination. Environment & Epigenome

Learning objectives:

- Describe apparently unexplained phenomena now enclosed in the term Epigenetics
- Describe the Waddington epigenetic landscape
- Learn the difference between epigenetic and genetic modification
- Describe different functions of DNA methylation
- Describe the developmental process from the DNA methylation point of view
- Describe how genomic imprinting works
- Learn the most common genetic diseases associated with genomic imprinting problems
- Understand the different mechanisms leading to UPD
- Learn the most common genetic diseases associated with abberant pattern of histone modifications
- Describe the organization of the nucleus, in terms of pores, lamina, and internal structures/granules
- Describe chromosome territories and TADs
- Learn the most common genetic diseases associated with a dys-organization of the nucleus structure
- Learn the basic modes of sex determination
- Describe the molecular mechanisms leading to X inactivation
- Describe how environmental factors can influence the epigenome

23. Cancer genetics

Epidemiology of cancer & Cancer as genetic disease. Cancer as a hereditary disease. Genetic counselling of cancer.

Learning objectives:

- Describe the basic molecular mechanism leading to cancer
- Describe the genetic landscape of cancers
- Describe the most common inherited cancer syndromes and their molecular mechanisms
- Learn the basic steps that are taken during a genetic counselling for cancer predisposition

24. RNA-based pathogenic mechanisms

MicroRNA biogenesis, mechanism of action, and nomenclature. MicroRNA-mediated pathogenic mechanisms: implications in Mendelian and complex diseases. The ceRNA hypothesis

Learning objectives:

- Describe the biogenesis of microRNA

- Describe the 4 main mechanisms leading to miRNA-based genetic disorders
- Learn how a ceRNA network works, and how they can be related to genetic disorders

Module: Computational science (C. Bolchini)

25. Sequence alignment

Importance of biological sequence comparison, local or global alignment of two biomolecular sequences, sequence similarity search, exact and heuristic approaches

Learning objectives:

- Learn how sequence alignment is performed
- Use of available software tools to perform sequence alignment and result interpretation

26. Introduction to available tools

Introduction to R and alignment with python for processing data, visualizing information and extracting knowledge

Learning objectives:

- Understanding how engineering methods for data access, manipulation and analysis can contribute to increasing biomedical knowledge
- Ability to use software tools prepare and manipulate data

Module: Bioinformatics (L. Pattini)

27. Exploring genetic variation

Applications of Next-generation sequencing (NGS) in biomedicine. NGS of whole genomes and exomes as a powerful tool in biomedical research and clinical diagnostics. Genome informatics for DNA variant characterization.

Learning objectives:

- List the main applications of NGS technologies in biomedical research and medicine
- Compute and visualize the mutation burden and patterns of variation
- Learn basic computational methods for annotation and prioritization of genetic variant candidates

28. Gene expression data analysis

High throughput technologies for gene expression profiling. Statistics for differential analysis. Unsupervised learning: clustering, dimensionality reduction. Functional annotation sources. Enrichment analysis.

Learning objectives:

- Learn the general pipeline for RNA-seq data analysis
- Understand potentialities of RNA-seq based transcriptomics
- Learn basic explorative analysis of gene expression profiling

Practical lessons:

A) Surfing the genome

These practical lessons will allow to acquire confidence with online tools for medical doctors and researchers in the field of molecular genetics. Students will become able to get information on genes, mutations, and associated diseases starting from an anonymous sequence of DNA.

B) Analysis of transcriptomic and genomic data

B1) Analysis of transcriptomic data

Total RNAseq data from cells treated or not with different compounds acting on splicing will be analyzed.
Object: identifying genes and splicing isoforms differentially expressed in treated vs untreated cells.

B2) Analysis of common germline variants

Genotyping data from COVID-19 patients and controls.
We will perform association and epistatic analyses on polymorphisms mapping in pathways relevant for disease pathogenesis. The most suggestive association data will be meta-analyzed with those reported in COVID-19 genetic databases.

B3) Analysis of somatic variants

Whole exome sequencing data obtained from carcinomas and nearby non-tumor tissue will be analyzed. We will define the pattern of somatic mutations, learn how to produce Waterfall plots of mutations, and how to compare the results with literature data. Mutated genes will be investigated by looking into gene structure and protein structure deposited in public databases (UCSC, pdb, etc.).

Teaching Methods

Lessons and practical lessons

Verification of learning

Students'evaluation will be assessed through a written examination with multiple choice questions. The exam will include 30 questions, proportionally distributed among the different modules. Each question will score 1 point; no penalties will be applied for wrong answers. The threshold score for passing the exam will be 18 points.

Texts

- Strachan & Read. Human Molecular Genetics 5th edition, Garland, 2018
- Alberts B. et al. Essential Cell Biology 5th edition, Norton, 2018
- Watson J et al. Molecular biology of the gene. 7th edition, Pearson, 2013
- Thompson & Thompson. Genetics in medicine. 8th edition, Elsevier, 2015
- Alberts B et al. Molecular biology of the cell. 6th edition, Garland Science, 2014