LESSON 1- 140518

Welcome to "APPLICATIONS IN MEDICINE", module of Advanced Molecular Biology.

I am CUTRUPI SANTINA, professor in Molecular biology. My research field is the role of estrogen receptor (ER) in the immune systems. The immune systems play a crucial role to defense organisms against pathogens. However, the immune systems dysregulation is involved in autoimmune disease (AD), such as Multiple Sclerosis. In our lab, we are studying how ER can act as immunomodulatory factors in the AD.

The topic of this module is how the biomolecular approach is used to understand disease. However, the goal of this module is to develop specific expertise in molecular biology.

The skills that the course wants to develop are:

PROBLEM SOLVING in the application of molecular biology IN MEDICINE

EXPERIMENTAL DESIGN to understand molecular mechanisms linked to disease

How can we improve these skills?

- Papers analysis
- Design experiments.

The lesson structure is composed from one part to explain main concepts and the other part with quiz.

We'll use a "TRAINING TASK".

What is the meaning of this approach?

- **Help you to understand the main concept in the deep way**
- **Help you to REMEMBER the main concept**
- **Help you to apply the main concept to solve problem**

No "classic" evaluation using number, but I'll use a comment that help you to understand the mistake and improve learning.

The goal is your learning, so stop me if you don't understand or want to explain one topic in more details. You can send me email, too.

In this lesson, we'll discuss about:

- **What is the main focus of the course**
- **Definition of Functional Genomics**
- **Molecular biology pre-requisite: DNA genomic elements as cell-type specific regulatory regions**
- **SNPs meaning in the disease**
- **How Functional Genomics is the basis to understand diseases**

We'll start with a quiz on the Moodle platform. 30 min.

Activity 1: what is Functional Genomics. Search by google. Compare your definition with the proposal of teacher.

The main focus of this course is **FUNCTIONAL GENOMICS**. This research field has an important impact in the comprehension of the molecular mechanism underpinning disease and open the way to biomarkers identification and drug discovery.

What is the meaning of "functional genomics"?

Functional genomics uses genomic data to study gene expression, regulation and biological functions on a global scale (genome-wide or system-wide), focusing on gene transcription, epigenetic modifications, chromatin remodelling enzymes, transcription factors association involving high-throughput methods.

In the body there are many types of cells, such as neuron, muscle cell, lymphocytes and so on. Each cell has the same genome, but transcribes different sets of genes, leading to differences in the character and function. In different cells, or in the same cell at different times, a sets of genes might be transcribed.

The genomic regulatory regions that are used from the cells control gene expression. The genes pattern expressed from the cells define cell identity and biological functions. the genomic regulatory regions are defined by:

- **EPIGENETIC MARKS**
- **TRANSCRIPTION FACTORS BINDING**
- **CHROMATIN REMODELLING ENZYMES**

- **NUCLEOSOME POSITIONING**

One alteration in the genomic regulatory regions, such as enhancer, may induce transcription activation of the gene that play a role in the disease. So, activation of enhancer respect healthy state can induce pathological conditions.

The alteration in the enhancer may be single nucleotide variants.

Types of SNPs, that can play a role in the disease, can be associated with non coding sequence, linked to genomic regulatory regions, or coding sequence, linked to protein structure and function.

What types of alterations in the molecular mechanism could induce diseases:

- Single nucleotide variations into the genomic regulatory regions change the consensus sequences for transcription factors binding
- Single nucleotide variations into the genomic regulatory regions change long range interactions between two regulatory regions
- Single nucleotide variations into the coding sequence of proteins change:

a) Enzimatic activity

b) Protein-protein interactions

c) Cofactors binding

In this view, there is the description of the possible function of SNPs in the genomic regulatory regions. The SNPs can change the enhancer activation and, for example, induce the activation of gene C expression. During the course, we will describe in more details the role of SNPs in the disease. We focus attention on one genomic regions, however we can extend this view for all genome.

The studies that have indentified the SNPs linked to specific disease are called GWAS (**Genome-wide association studies)**.

How functional genomics is a tool to understand disease

In order to show an overview of research plan for studying the role of SNPs I will present an example.

In this number of Neuroscience there is a description of the meaning of SNPs linked to psychiatric diseases.

A SNP may change the molecular mechanisms that control gene expression, such as TFs association with chromatin.

Which are the steps to understand the SNPs meaning?

The first step is the sequencing of patients genome. The comparison of the genome sequences between healthy and patients show the nucleotide variations. The SNPs associated with disease are collected in GWAS database. The next steps are: functional annotation, experimental validation in vitro, using cellular or disease animal models.

The second step to understand whether SNP has a role in disease outcome is FUNCTIONAL ANNOTATION. For localization of the SNP in the specific genomic regions we can used genome wide data.

FUNCTIONAL ANNOTATION is the method used to search whether SNPs are associated with specific genomic regions and which are the features of these regions, such as transcription factors association, epigenetic modifications and others. In order to describe SNP-associated regions we can used genome-wide sequencing data. Therefore, SNPs may be associated with a genomic regions described as enhancers, promoter, silencer or repressor regions. These information permit us to formulate hypothesis about SNP function, for example SNP in the genomic regulatory region may change the expression of nearest gene.

SNPs in the genomic region may alter a binding site of a specific TFs, such as PU.1 and chromatin states change in the same region

The genome Reference (Ref) is compare with the genome "Altered" (Alt) (SNPs). We observed the analysis of several aspects in the enhancer region, Tatem-VCM: PU.1 binding, histone modifications and PolII binding. In Ref sample, we find a peak that corresponds with the PU.1 binding, while a low signal is associated with Alt sample. In the gene body, we find an high signal of PolII binding and PU.1 in the vcmQTL regions in the Alt sample suggesting that the alteration in the genomic regulatory regions induced PU.1 binding and this SNP is linked with gene transcription.

Why?

Is a variant regulatory?

If we found a change in the signal of peak we can suppose that this variant changes the transcription binding sites. The specific algoritm indicates whether this variant impairs the transcription factor binding.

In enhancer, several TFs form a regulatory complex that control gene expression, but the variant in one TF binding site can impair the formation of regulatory complexes. This alteration can change the cell fate and define the disease trait.

Distinct Modes of Genetic Variation-Mediated Changes in TF-DNA Binding

(A) Only a minority of variable TF-DNA binding events are caused by DNA variants disrupting the cognate TF recognition motif. (B–D) The majority of variably binding events are motif variation independent, indicating that a variant located either proximally (<200 bp,Band C) or distally (D) to the focal motif affects the binding of the respective TF. Proximal variants can affect local cooperative DNA binding (B), which involves physical protein-protein interactions that require overlapping or very closely located (a few bp) motifs, or collaborative DNA binding (C), which reflects TF interdependencies needed, for example, to compete with nucleosomes and thus to access DNA. In contrast, distal variants (D) may alter chromatin state or conformation (e.g., DNA loops), which could affect the stability of interactions with DNA and between TFs.

Using animal or cell lines models we can verify the functions of SNPs. For example, the preclinical disease model could be a transgenic mice with mutation in the genomic regulatory regions that correspond with SNP of interest. Several levels of studies can be applied: DNA, RNA (look at the diapo), Behavior and Cognition tests.

Another example for identification of SNP role is the tumorigenesis.

The pipeline to investigate SNP is the same that is describe for neurological disorder.

1) Sequencing the genome and select the SNPs associated with disease
2) Functional annotation using bioinformatic tool that help us to localize SNP in a specific genomic regions and to know the information about that genomic locus.

Experimental validation:

Mutations in cloned DNA fragments can be generated using **site-directed mutagenesis** or the **CRISPR–Cas system**. **Synthetic oligonucleotides** with a wild-type or mutant sequence can also be chemically synthesized. **b** | Functional output of the non-coding mutations can be determined either using a single or combinatorial approach involving high-throughput sequencing and/or luciferase (LUC) reporter assays. **In high-throughput sequencing,** effects of mutations in *cis*-regulatory elements (promoters and enhancers) can be studied by an approach called *cis*-regulatory element analysis by sequencing (CRE-seq)[118,122]. For **CRE-seq**, synthetic regulatory element constructs with wild-type and mutated sequence are cloned into reporter construct, which is tagged at the 3′ end using a specific nucleotide barcode that identifies the upstream promoter or enhancer element. In **an alternative method** for characterizing enhancer variants, self-transcribing active regulatory region sequencing (**STARR-seq**)[119], enhancer libraries are flanked by synthetic adaptor DNA sequences and cloned downstream of a transcription reporter construct. For both approaches, RNA transcripts from these libraries are used for cDNA synthesis followed by high-throughput sequencing. The expression driven by each element is measured by the ratio of the fraction of reads in the cDNA pool and the genomic DNA pool for each library construct; the particular element driving the expression of each transcript is identified based on the sequence of the transcribed barcode (for CRE-seq) or the transcribed enhancer (for STARR-seq). This enables accurate quantification of the reporter transcript as a direct measure of the regulatory element activity. For the **LUC reporter assays**, DNA fragments cloned into the reporter vectors are transfected in cells followed by

measuring the reporter activity. **c** | Oncogenic properties, such as cell proliferation, migration and invasion, can be tested *in vitro* using cell lines, and tumorigenesis can also be tested *in vivo* using model organisms.

Also in tumorigenesis there are association with clinical profile. SNPs meaning is the basis for application as biomarkers or to develop new drug.

SNPs role.

SNP has a gain of function when creates a consensus sequence for TF binding, therefore transcription is active.

SNP has a loss of function when disrupts a consensus sequence for TF binding, therefore transcription is repressed.

SNP may have a role in the affinity of TF binding, so when hormone concentration is low TF does not bind the regulatory regions and transcription is impaired.

In gene locus, for example IL2RA, there are several data of RNA-Seq, Histone Acetylation and cell-type specific TFs binding. In the upper part, there are disease-associated SNPs. For autoimmune disease, high number of SNPs linked to IL2RA, receptor for IL2, suggests autoimmune susceptibility.

The SNPs function could be tested in patients and we will find the correlation with clinical profile. How can we use these knowledge? SNPs-associated molecular mechanisms could be the basis for drug discovery or biomarkers. Biomarkers are the tools to follow disease outcome, while the molecular mechanisms that are involved in the disease help us for the identification of the target for a new drug.

All materials for the course are present in the web site. You will find the slides presentation, reviews and articles. In addition, you will find lesson audio and lesson notes that help you to mark the main concepts.

Final exam is a test composed of different types of questions: multiple choice, finding associations between phrases and two open questions.