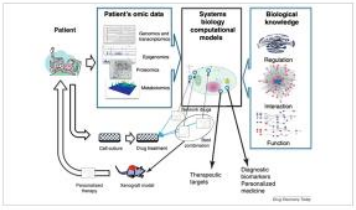
Welcome to everybody.

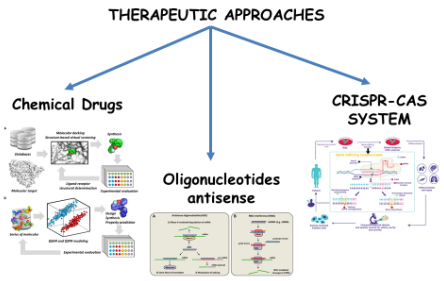
I would like to introduce APPLICATIONS IN MEDICINE, module of "Advanced molecular biology" course.

APPLICATION IN MEDICINE

- GENE FUNCTION AND REGULATION
- THERAPEUTIC APPROACHES



This part is focusing on the strategies to study gene function and uncovering biological mechanisms that regulate gene expression and biological functions. These knowledges are the base for new drug discovery.



The therapeutic approaches that were developed today are three main categories:

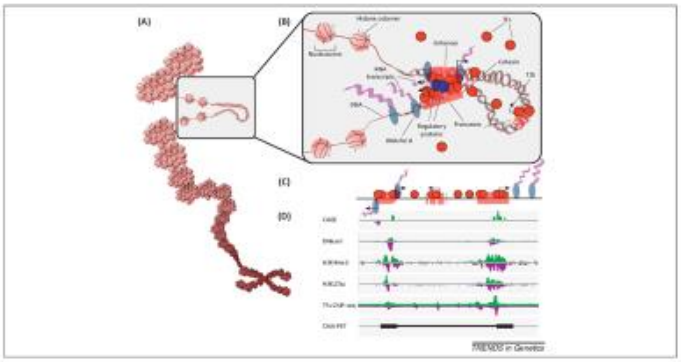
- Chemical drug: you have discuss in the biochemistry course
- Antibodies
- Oligonucleotide antisense
- CRIPR-Cas systems

The knowledges that derive from these studies is fundamental to find a specific drug that acts on specific target. Drug design depends on the target. If the target is a complex and we want to interfe with interaction between two proteins, we can develop a synthetic drug. If the target is a transcript and we want to inhibit or downregulate the transcript or regulate the splicing, we can develop an oligonucleotides antisense. If the target is a specific genomci regions at a gene and we can modify these regions, we can use CRISPR-CAS9 SYSTEM.

Oligonucleotide antisense and CRIPR-Cas systems are discussiong during this module. The tool that help us to discover new target of a specific drug is FUNCTIONAL GENOMCIS.

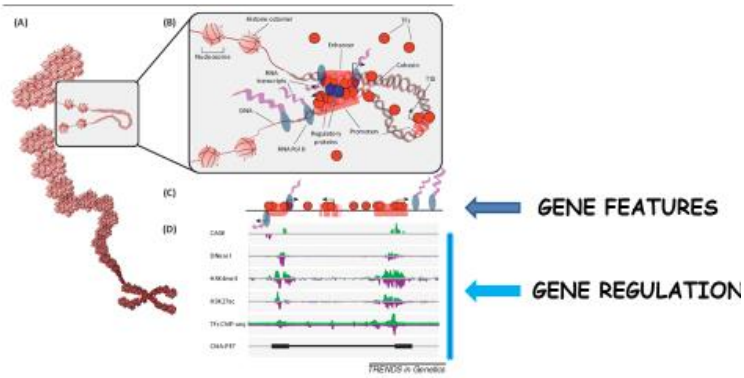
Functional genomics

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.



The funzional genomics is a field that collect genome wide data derived from genomics, transcriptomics and proteomics, and the integration data is a tool to identify a set of genes or a specific target to develop a drug.

GENOMIC REGULATORY REGIONS



Genome-wide data, such as RNA-Seq, ChIP-Seq and ChIPA-PET help us to identify the gene features, as gene body, promoter and enhancer and this information is important to reconstruct a model in which we can recognize the regulatory elements.

GENE REGULATION
How we can understand gene regulation
Using genome-wide sequencing data

- FUNCTIONAL GENOMICS
- INTEGRATION DATA APPROACH

The integration of genome-wide data is a bioinformatic tool to develop a drug against a specific target. First step is the annotation of reads that derive from sequencing to a reference genome, compare the data and interpretate the data.

An example to understand the steps to identify target from genome-wide data to drug discovery is the study to find the role of single nucleotide variants associated with a disease. We can see each steps.

Framework for interpretation of individual disease-associated variants

- Single nucleotide polymorphisms (SNPs) is the nucleotide variations associated with disease

- Genome-wide association studies (GWAS) have successfully identified thousands of common genetic variants associated with complex diseases (http://www.ebi.ac.uk/gwas/)

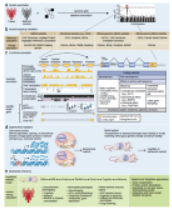
- Functional annotation: to define genomic regulatory regions by genome-wide integration data

- Experimental validation

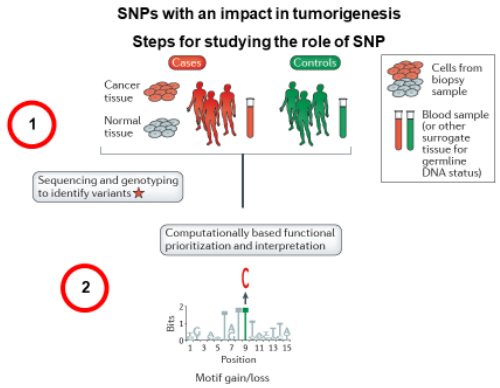
- Disease Animal models

- Correlation between molecular mechanisms and disease symptoms

- Drug Discovery

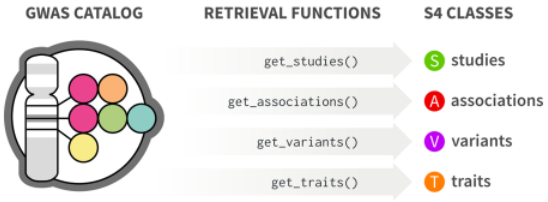


The first step is the identification of single nucleotide variants associated with a disease and integrate the data to identify the role of variants.

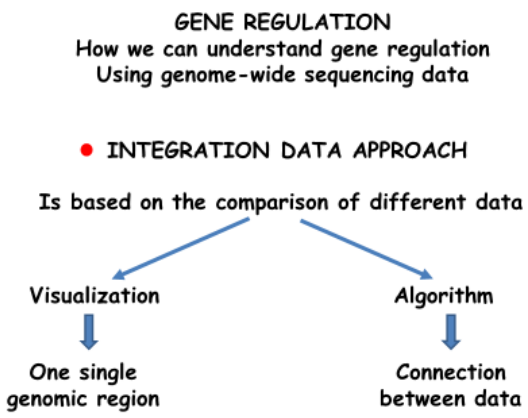


The picture shows how starting from a collection of genome-wide data we can identify the role of SNPs in the trascription factors binding sites.

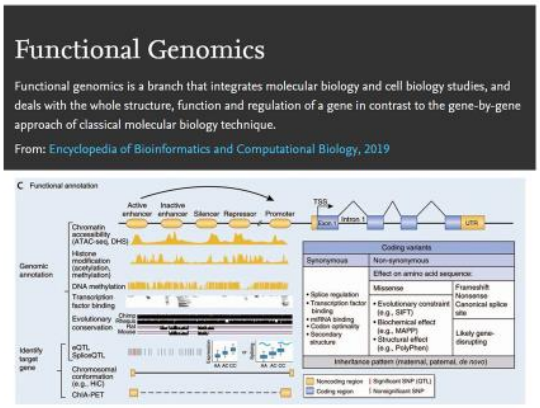
Genome-wide association studies (GWAS) have capitalized on the millions of common single nucleotide polymorphisms (SNPs) to identify those SNPs that are genome-wide significantly associated with a disease or trait.



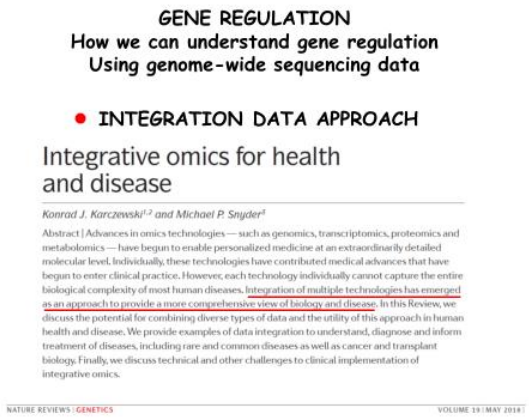
GWAS is a collection of SNPs associated with disease. You can find where is the location of specific SNPs and the finding of the role of SNPs stating with the exploration of the genomic regions in which SNPs is present.



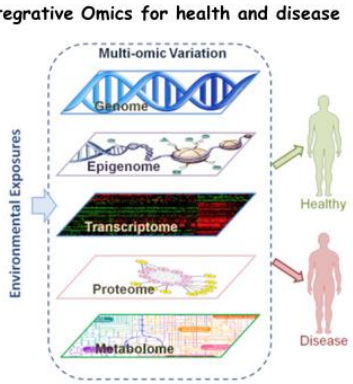
Integration data approaches are based on two main analysis: one is the visualization of a specific gene and another is the use of bioinfomatics tools by algorithm that find the connection between different types of data.



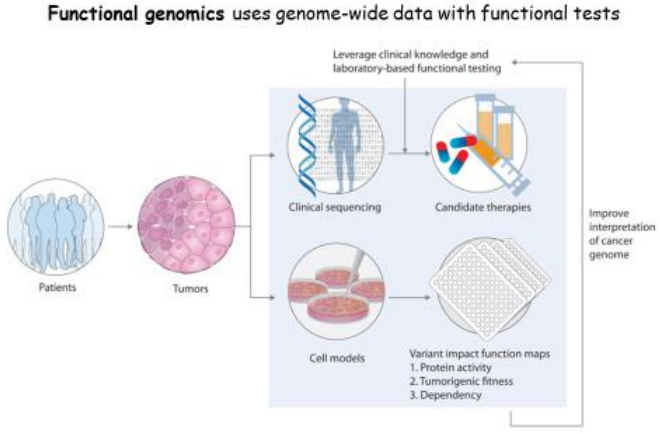
Integration data based on the visualization focuses the attention on the specific genomic region. Genome-wide data give you information about promoter, enhancer and so on.

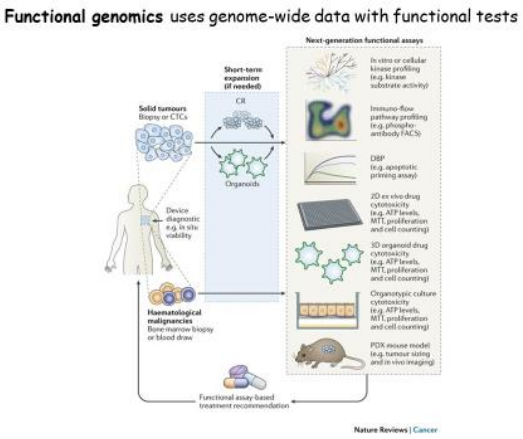


This review gives you an overview of integration data. This is a representation how multiomics give information about patient's profile.

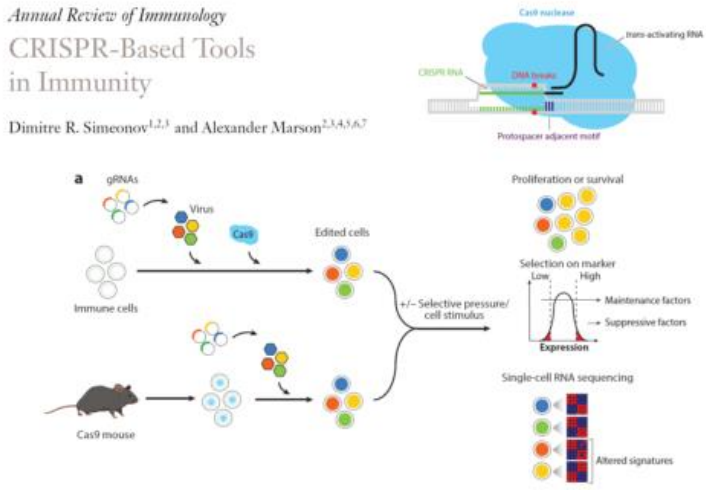


This step of analysis is important to select the genes that are involved in the disease and experimental assay and animal models are some tools to demonstrate the role of gene in the disease development. The candidates that derived from integration data can be tested in biological assay, as migration assay, apoptotic assay and cell-cycle analysis. These results are important to validate the target that could be used in the therapy.





The next picture shows the analysis on stem cells and you can see a lot of assays that can be used to validate the target.

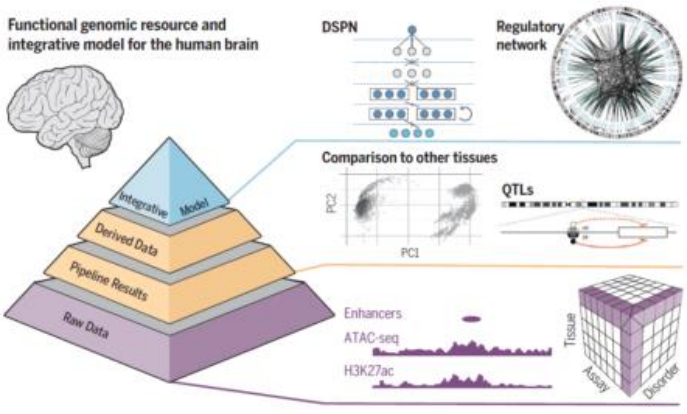


The selection of the targets included the experiment in which CRISPR-Cas9 system is used. A library of RNA guides is created to identify which target plays a role in a specific biological processes, as survivial or cell cycle and single-cell RNA expression profiling shows the impact of a specific RNA guide, therefore the role of a specific target.

We can summarize the steps to identify the role of SNPs:

- Functional annotation
- Experimental validation
- Drug discovery.

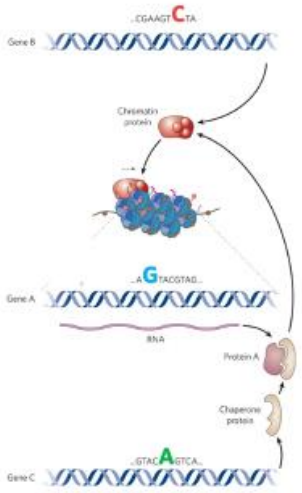
The article that discuss about the role of SNPs in psychiatric disorder is an example of functional genomics and the application of integrative approach to indentify the potential genes involved in this disease (Wang et al., 2018, Science). This is very complex article and I included it in the bibliography and you can see all experiments. This course do not want to describe these data in details, however the goal is understand the steps to know the



impact of SNPs associated with disease. The sequencing of patients' genome gives information about SNPs linked to disease and specific for psychiatric disorder, in this case. The goal of functional genomcis is found the role of SNPs in the development of disease. Thanks to genome wide data, a lot of data can be compared using biofinformatics tools. The base of the pyramidal resourse is the datasets of transcriptome, chromatin immunoprecipitation and long-range interaction (Hi-C) on bulk samples. Single cell-transcriptomic data describes the differences between cells and this is a crucial point in the brain in which we can find a high grade of cell specilization.

INTRODUCTION: Strong genetic associations have been found for a number of psychiatric disorders. However, understanding the underlying molecular mechanisms remains challenging.

RATIONALE: To address this challenge, the PsychENCODE Consortium has developed a comprehensive online resource and integrative models for the functional genomics of the human brain.



RESULTS: The base of the pyramidal resource is the datasets generated by PsychENCODE, including bulk transcriptome, chromatin, genotype, and Hi-C datasets and single-cell transcriptomic data from ~32,000 cells for major brain regions. We have merged these with data from Genotype-Tissue Expression (GTEx), ENCODE, Roadmap Epigenomics, and single-cell analyses. Via uniform processing, we created a harmonized resource, allowing us to survey functional genomics data on the brain over a sample size of 1866 individuals.

1. New genome-wide data

2. Comparison of New genome-wide data with data derived from several databases

From this uniformly processed dataset, we created derived data products. These include lists of brain-expressed genes, coexpression modules, and single-cell expression profiles for many brain cell types; ~79,000 brain-active enhancers with associated Hi-C loops and topologically associating domains; and ~2.5 million expression quantitative-trait loci (QTLs) comprising ~238,000 linkage-disequilibrium–independent single-nucleotide polymorphisms and of other types of QTLs associated with splice isoforms, cell fractions, and chromatin activity. By using these, we found that >88% of the cross-population variation in brain gene expression can be accounted for by cell fraction changes. Furthermore, a number of disorders and aging are associated with changes in cell-type proportions. The derived data also enable comparison between the brain and other tissues. In particular, by using spectral analyses, we found that the brain has distinct expression and epigenetic patterns, including a greater extent of noncoding transcription than other tissues.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat8464

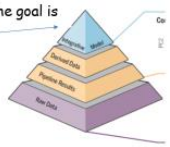
3. Interpretation of data

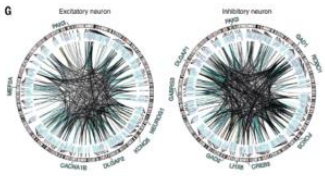
4. Gene expression profile is specific for a group of cells as shown by Single Cell-RNA-Seq

5. Identification of SNPs with specific functions

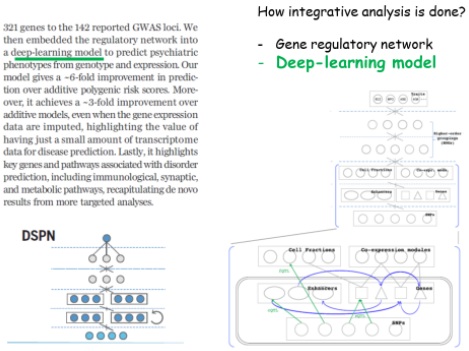
The top level of the resource consists of integrative networks for regulation and machine-learning models for disease prediction. The networks include a full gene regulatory network (GRN) for the brain, linking transcription factors, enhancers, and target genes from merging of the QTLs, generalized element-activity correlations, and Hi-C data. By using this network, we link disease genes to genome-wide association study (GWAS) variants for psychiatric disorders. For schizophrenia, we linked 321 genes to the 142 reported GWAS loci. We

The goal is





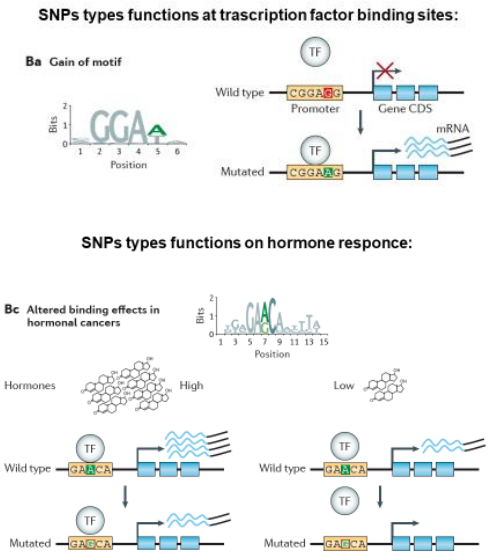
These data revealed the different cell type-specific gene profile and the comparison with the presence of SNPs in a specific genomic regions and the other features, as epigenetic profile and long-range interactions, shows the potential role of SNPs in the biological functions. The tools that are used to identify the role of SNPs are the reconstruction of gene regulatory network and deep-learning.



You can appreciate the diagram the represent the results of deep-learning analysis and I would like that you remember how this is a way to find the role of SNPs.

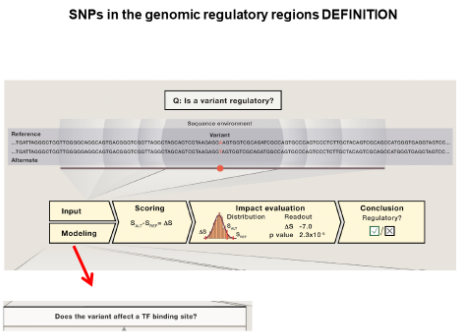
Now we will explore the role of SNPs. ClirVar is a catalog of variants that are linked to Ensembl. The classification of variants is complex. You know the variants in the coding sequence that can change the function of proteins and, in the case of enzyme, a variant in the catalitic site can inhibit enzymatic activity. We can find some variants along the gene, in 5' prime UTR or 3' prime UTR, in splincing site with consequence on the protein synthesis. The variants in the non coding sequence are important because have an impact on the gene regulation.

Now we are pointing our attention on variants in the regulatory regions.

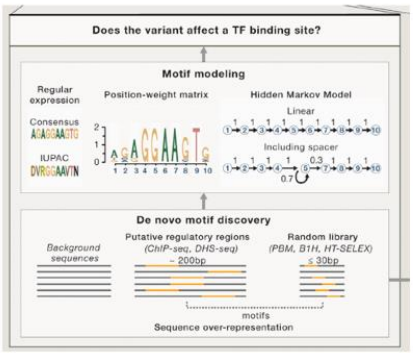


A variant in the trascription factor binding site can induce the binding of transcription factor at the regulatory region inducing trascription activation, called Gain of function.

A variant in the trascription factor binding site can inhibit the binding of transcription factor at the regulatory region imparing trascription activation, called Loss of function.A variant in the trascription factor binding site after hormone treatment can alter the sensitivity to the hormone. At high level of hormone, TF binds at the regulatory regions, while low level of hormone is not able to allow TF binding when there is a variant in the consensus sequence for TF. The alteration of hormone effect occurs also for drug responce, therefore some variants represent a component of suscettibility to disease and can play a role in the efficacy of drug treatment.

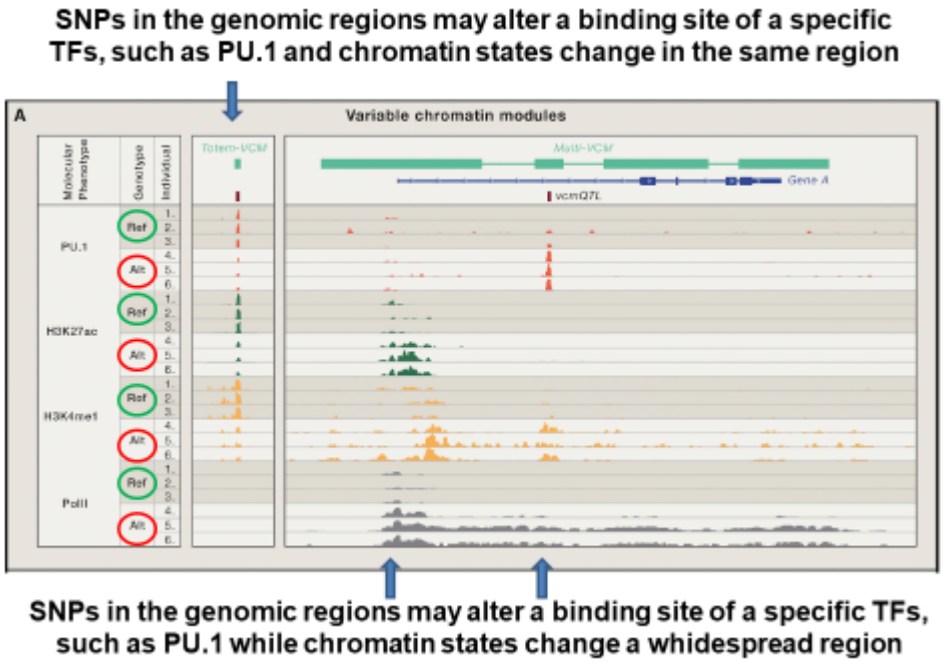


The attribution of SNPs effect at TFBS and the role in the alteration of TF binding derives from calculation of several parameters that consider the consensus sequence.



The choice of the strategy depends on the posed question: does the variant impact on the binding of a TF or the local chromatin landscape.

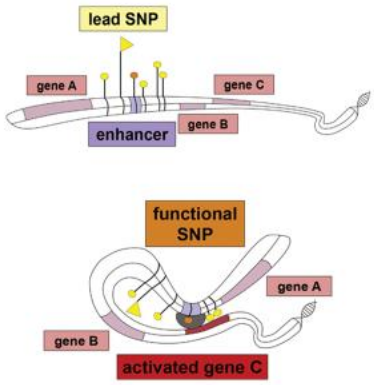
We can see an example. In this picture, Gene A has a variant in the genomic regions, called vcmQTL and TOTEM-VCM. The picture shows examples of regulatory regions that change in two sample, as reference and "altered" or pathological sample. In the "TOTEM region" the variant associated with "altered" sample inhibits the binding of PU.1, a transcription factor. In the same location of PU.1 binding, epigenetic modifications change inducing a chromatin remodelling in this region. Note that in this region, there is no PolII binding, so gene is not expressed. In the "VCM-QTL", the variant gives a gain of motif for which PU.1 binds regulatory regions in "altered" sample. The comparison of reference and altered samples shows the changes in the epigenetic marks in the same location of PU.1, and also in proximal region, in addition there is



PolII binding suggesting that PU.1 binding regulates chromatin organization inducing an active state and gene expression.

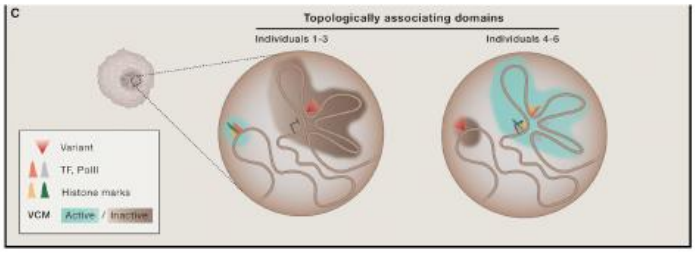
The impact of variant can be associated with the region in which varianti is present, and also with distal region, in this case, you can see an alteration of long range interaction.

Single nucleotide variants in genomic regulatory regions



This picture shows that a variant induces trascription activation, as the previous example, acting on a distal enhancer that allows a long range interaction. We have to consider the chromatin structure that is dynamic and is regulated by epigenetic changes.

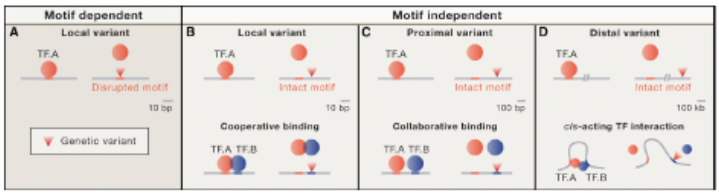
SNPs may change long range interactions



VCMs constitute a nuclar area where higher-order chromatin organization embedded within topologically associating domains (TADs) and provide a molecular rationale as to how TF-DNA binding can be affected by distal genetic variation.

Deplancke et al., 2016

SNPs mechanims for alteration of regulatory transcrption factors complexes



Deplancke et al., 2016

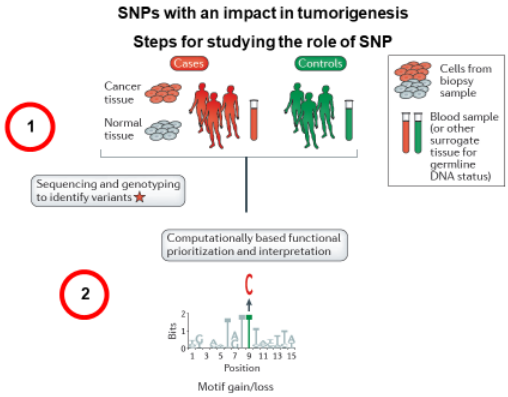
The variant can act directly on TF binding when it is into binding site. There are variant effect independent from motif:

- Variant impairs the binding of complexes between two TFs, called cooperative binding. One TF interact with other TF and the complex do not bind regulatory regions due to variant.
- The binding of one TF depends on the other TF and the variant can impair one TF binding acting on opportunity of other TF binding. This is called "collaborative binding".
- Third model was described before.

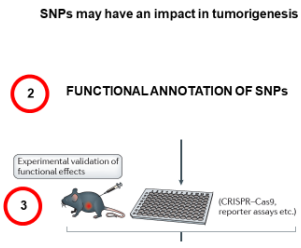
  After the finding of potetial role of SNPs, the demonstration of the effective role depends on experiments. The experiments

In summary, SNPs in the genomic regulatory regions have an impact on:

- Changing the transcription factor binding site
- Disrupting long range interaction
- Inhibition of interaction between transcription factors
- Changing epigenetic modifications

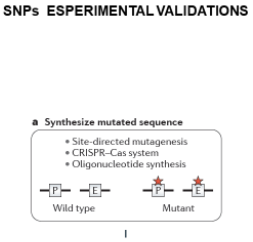


The summary of the steps important to find the role of SNPs is starting with the collection of the SNPs associated with the disease and identifying the location of the SNPs on the genome.

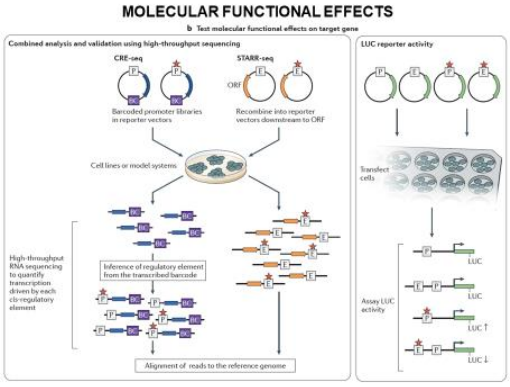


The position of SNPs on the genome allows the comparision with genome-wide data suggesting the potential impact of SNPs that we have described before for genomic

regulatory regions. The experimental assays are necessary to understand how SNPs act because the functional annotiation indicates the potential role.

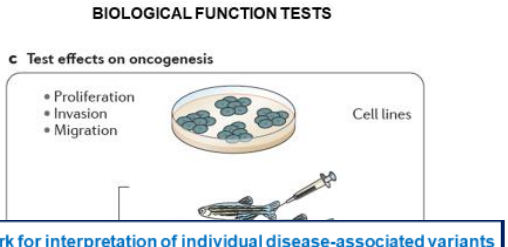


We can study the SNPs using the site-directed mutagenesis, CRISPR-Cas system and oligonucleotide synthesis. For genomic regulatory regions, we can use CRE-Seq and STARR-Seq when you want to analyze several sequences or luciferase assay when you analyze a specific SNP.



In CRE-Seq, the sequences are cloned into the vectors with barcodes to recognize samples and they are transfected in the cells. The measure of promoter activity depends on the number of trascripts.

In STARR-Seq, enhancer are cloned into the vectors creating a library that are trasfected in the cells. After sequencing and annotation of reads, it is possible to identify if SNPs have an impact on enhancer activity. The application of luciferase assay is discussed in the activities associated with lessons.



The biological functional tests conclude the studies to understand the role of SNPs using several models.

Altogether these data are applied to understand molecular mechanisms and finding the correlation between SNPs involved in gene expression and clinical synthoms.

Framework for interpretation of individual disease-associated variants

- Single nucleotide polymorphisms (SNPs) is the nucleotide variations associated with disease

- Genome-wide association studies (GWAS) have successfully identified thousands of common genetic variants associated with complex diseases (http://www.ebi.ac.uk/gwas/)

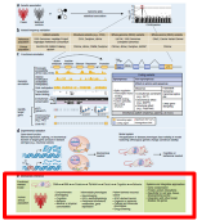
- Functional annotation: to define genomic regulatory regions by genome-wide integration data

- Experimental validation

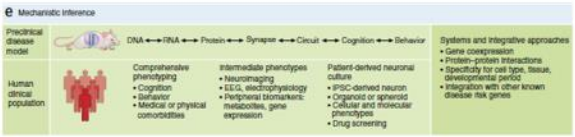
- Disease Animal models

- Correlation between molecular mechanisms and disease symptoms

- Drug Discovery



Correlation of SNP/functions with several clinical analysis

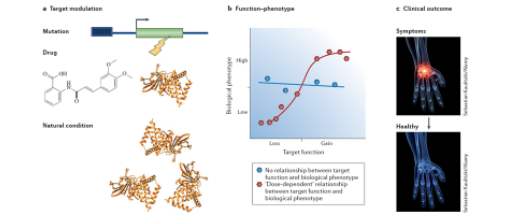


How can we use these knowledge?

The final step is to discover new drug or identify biomarkers to describe disease and outcome.

EXAMPLE

Gene expression alteration in disease May be used as BIOMARKERS (molecules acting as sensor of disease)	Gene expression alteration in disease May be used as DRUG TARGET (drug discovery to stop disease and restore health)



In Summary:

- Functional genomics is a field of molecular biology based on genome-wide sequencing data.
- Genome-wide sequencing data describe genomic regulatory regions that control gene expression
- Gene expression disregulation and SNPs can be linked to the disease
- Understanding molecular mechanisms of disease outcome opens the way to discovery drug and identify biomarkers

In this period, it is very important to develop new strategy to find Covid19 prevention and cure.