In the previous lesson, we have discussed about the definition of SE and their function.

Super-Enhancers are:

large clusters of transcriptional enhancers formed by binding of master transcription factors

are associated with genes that control and define cell identity

are enriched for disease-associated DNA sequence variation

are associated with oncogenes that become active during the process of tumor pathogenesis

Super-enhancers were described as a class of regulatory regions with unusually strong enrichment for the binding of transcriptional coactivators, specifically Mediator (Med1). Super-enhancers in mouse embryonic stem cells (mESCs) were defined by the following method.

(1) Sites bound by all three master regulators, Oct4, Sox2 and Nanog, according to ChIP-seq, were considered enhancers.

(2) Enhancers within 12.5 kbp of each other were concatenated ('stitched') to define a single entity spanning a genomic region.

(3) The stitched enhancer entities and the remaining individual enhancers (those without a neighboring enhancer within 12.5 kb) were then ranked by the total background-normalized level of Med1 signal within the genomic region.

4) Enrichment for H3K27Acetylation

Disease-associated SNPs occur in super-enhancers of disease-relevant cells and that this occurs more frequently for super-enhancers that drive the expression of genes that control and define cell identity, suggesting that altered expression of cell identity genes may often contribute to these diseases. These observations also suggest that how SNPs may affect SE to regulate gene identity and the tole of SNPs in disease outcome.

An another approach for studying enhancer function is the deletion of enhancers and measure the impact on gene expression. The deletion of enhancers consists of genome editing by CRISPR-Cas9 assay.

Description of CRISPR-Cas9 assay.

Genome editing is an approach to add, remove or alter particular locations in the genome.

CRISPR-Cas9 was adapted from a naturally occurring genome editing system in bacteria. The bacteria capture snippets of DNA from invading viruses and use them to create DNA segments known as CRISPR arrays. The CRISPR arrays allow the bacteria to "remember" the viruses (or closely related ones). If the viruses attack again, the bacteria produce RNA segments from the CRISPR arrays to target

the viruses' DNA. The bacteria then use Cas9 or a similar enzyme to cut the DNA apart, which disables the virus.

CRISPR is an acronym for **Clustered Regularly Interspaced Short Palindromic Repeat**. This name refers to the unique organization of short, partially palindromic repeated DNA sequences found in the genomes of **bacteria** and other microorganisms.

While seemingly innocuous, ***CRISPR sequences*** are a crucial ***component of the immune systems*** of these simple life forms. The immune system is responsible to protect an organism's health and well-being. Just like us, bacterial cells can be invaded by viruses, which are small, infectious agents. ***If a viral infection threatens a bacterial cell, the CRISPR immune system can thwart the attack by destroying the genome of the invading virus***. The genome of the virus includes genetic material that is necessary for the virus to continue replicating. Thus, by destroying the viral genome, the CRISPR immune system protects bacteria from ongoing viral infection.

The CRISPR immune system works to protect bacteria from repeated viral attack via three basic steps [5]:

Step 1) **Adaptation – DNA from an invading virus** is processed into **short segments** that are inserted into the CRISPR sequence as new spacers.

Step 2) **Production of CRISPR RNA** – CRISPR repeats and spacers in the bacterial DNA undergo transcription, the process of **copying DNA into RNA** (ribonucleic acid). Unlike the double-chain helix structure of DNA, the resulting RNA is a single-chain molecule. **This RNA chain is cut into short pieces called CRISPR RNAs.**

Step 3) **Targeting – CRISPR RNAs guide bacterial molecular machinery to destroy the viral material**. Because CRISPR RNA sequences are copied from the viral DNA sequences acquired during adaptation, they are exact matches to the viral genome and thus serve as excellent guides.

CRISPR RNA (crRNA) derived from CRISPR repeats protospacers and are associated with bacterial genomic regions.

tracrRNA, transactivating RNA, form a structure for Cas9 binding.

crRNA is complementary with DNA target while tracrRNA associates Cas9.

In a) there is the model in the bacteria, while in b) there is biomolecular technique where crRNA and trascrRNA are linked and is used for manipulation of specific genomic regions.

The modifications into the Cas9 protein allow to several genome changes.

The CRISPR-Cas9 system works similarly in the lab. Researchers create a small piece of RNA with a short"guide" sequence that recognize a specific target sequence of DNA in a genome. The RNA also associates to the Cas9 enzyme. As in bacteria, the modified RNA is used to recognize the DNA sequence, and the Cas9 enzyme cuts the DNA at the targeted location. Once the DNA is cut, researchers use the cell's own

DNA repair machinery to add or delete pieces of genetic material, or to make changes to the DNA by replacing an existing segment with a customized DNA sequence.

In type II **CRISPR** (**clustered regularly interspaced short palindromic repeats**)-Cas (CRISPR-associated) systems, the **endonuclease Cas9** associates with a dual-RNAguide structure consisting of a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA) to cleave double-stranded DNA (dsDNA) using its HNH and RuvC nuclease Domains.

Cas9 contains two nuclease domains: an **HNH nuclease** domain that cleaves the target strand of DNA (complementary to the guide RNA) and a **RuvC-like nuclease** domain that cleaves the non target strand. Mutating one of the two nuclease domains creates a nickase Cas9 (nCas9) that cleaves only one strand of DNA. Two nCas9s can be targeted to adjacent DNA sites to cause a double-strand break (DSB). Pairs of nCas9s have been used to increase the specificity of Cas9-based genomic editing, as only two adjacent nicking events will generate a DSB. Mutating both nuclease domains generates dCas9, which lacks nuclease activity but retains its RNA-guided DNA-binding activity. This allows dCas9 to be fused to other effectors to mediate site-specific genetic and epigenetic regulation without cleaving the target DNA, as well as specific DNA binding for several other applications.

The **Cas9 nuclease** associates the guide RNA and binds the structure with specific interactions in a number of domains. Specifically, the **REC1 and REC2 domains bind the complementary region of the guide RNA**, and eventually the **guide RNA target DNA heteroduplex upon DNA binding**. Mutations to the REC2 domain causes a small decrease in Cas9 activity, while mutations in the REC1 domain eliminate activity completely. The Rec1 domain is likely essential for Cas9 activity because it binds the repeat/anti-repeat duplex. The **Protospacer Adjacent Motif (PAM) Interacting (PI) domain** and **RuvC nuclease domain bind the stem loops on the guide RNA**. Overall, the exact process of guide RNA binding by the nuclease is unknown. However, a dramatic conformational change upon binding has been shown to produce an activated Cas9 nuclease capable of binding and cleaving specific sequences of double-stranded DNA.

RNA-guided DNA recognition and cleavage strictly require the presence of a **protospacer adjacent motif (PAM) in the target DNA**.

Description of the CRISPR-Cas9 vectors.

CRISPR-Cas9 system is based on a vector that contains a sequence for RNA guide, complement to target genomic regions, and a sequence coding for Cas9.

In the experiment where is tested the role of single enhancer in the SE, ESC cells were transfected with plasmids and the expression of gene under SE regulations was measured by PCR.

CRISPR-Cas9 application.

In article experiment, each enhancer of SE is deleted and the clone with specific deleted enhancer are tested for gene expression.

If we combined the experiments about luciferase assay, OCT4 silencing and genome editing with CRISPR-Cas9 we observe that E3 is important for Prdm4 gene expression and E3 is necessary for luciferase activity, OCT4 binds E3.

In order to verify that enhancers of SE may form chromatin loop, the chromatin immunoprecipitation with ligated regions are tested using ChIA-PET.

In the upper part of figure, there are fragments that indicate the long range interactions. These results are compared with ChIP versus H3K27Ac in clone wild type or clone with deletion of enhancer 3. Deletion of E3 induces a H3K27Ac signal reduction, therefore E3 is important for chromatin remodeling.

Signal transduction pathways: cell regulation

The terminal TFs of the Wnt (TCF3), TGF-b (SMAD3), and LIF (STAT3) signaling pathways, which play essential roles in transcriptional control of the stem cell state, were among the TFs whose **binding pattern to SE constituents** was most similar to that of OCT4, SOX2, and NANOG at SE constituents

The results of transcriptional profiling and gene set enrichment analysis in ESCs confirm this prediction; **super-enhancer associated genes were found enriched among the genes whose expression exhibited the most profound changes** after pathway stimulation or perturbation (Wnt: $p < 0.01$; TGF- b: $p < 0.01$; LIF: $p < 0.01$).

Cancer cells frequently acquire super-enhancers at oncogenes and dysregulation of signaling pathways is a hallmark of cancer, leading us to investigate whether these two features may be linked. Colorectal cancer tumorigenesis depends on hyperactivation of the Wnt pathway and the authors wanted to investigate whether the acquisition of super-enhancers provides for TCF binding and Wnt responsiveness at associated genes. The high density of H3K27Ac signal that is characteristic of super-enhancers was used to identify super-enhancers in normal human colon tissue and in the colorectal cancer cell (CRC) line HCT-116.. **Oncogene-associated super-enhancers, not present in normal colon cells**, including those at the c-MYC locus, **were found in the colorectal cancer cell line** (Figure 4A, Figure S4B), indicating that they are acquired in the tumor cells.

In order to test the role of single enhancer in the SE in proximity of Myc, the luciferase assay is performed upon Wnt stimulation and after Wnt inhibition. Wnt stimulation induced an increase of luciferase activity, while Wnt inhibition decreased signal.

High signal of H3K27Ac is observed in the regions that are identified as SE in the comparison between normal colon tissue and HCT-116, colonretal cell lines.

Wnt stimulation induces an upregulation of genes that are associated with SE. the bars indicated the genes linked to SE and the colour scale indicated the expression. Wnt inhibition downregulates the genes associated with SE as shown the high density of SE-gene near the colour blue, that marked the low expression of genes.

These data represent the based for therapy and the authors made a Syros Pharmaceutycal.