Lesson 5- 23 May 2018

In the previous lesson, we have showed that SE is a platform for several TFs that are activated from oncogenic signaling cascade. For studying the role of SE, we can used:

- Luciferase assay for testing the ability of SE in the transcription activation
- CRISPR-Cas9 system for testing whether SE is important for transcription regulation

In the article of Mansour et al., 2014, the mutation in the non coding regions creates a SE that activates the oncogene TAL-1.

Somatic acquisition of **insertions in a non-coding region** near *TAL1* established a **super-enhancer** that activated *TAL1* transcription in T-ALL cells. The goal of their study was to assess if genomic alterations affecting *cis*-regulatory elements could activate monoallelic *TAL1* transcription in TAL1+ T-ALL cells lacking *TAL1d* or *TAL1/TCRAD* lesions.

ChIP-Seq revealed **aberrantly high density** and breadth of histone H3 lysine 27 acetylation (**H3K27Ac)** from –20 kb through + 10 kb of the ***TAL1* TSS in Jurkat cells** , but not in other T-ALL cell lines or normal human hematopoietic stem cells.

The alignment of genome sequence for two cell lines and patients show a mutation in the sequence.

TAL1 expression is tested in these sample. Why is necessary PCR assay?

These sequences were actually known to correspond to the *TAL1* 5′ proximal enhancer (–10 kb) and 'Jurkat enhancer' (–8 kb), a *cis*-regulatory element that TAL1 and other transcription factors bind, to promote *TAL1* transcription in Jurkat cells. Genomic sequencing of Jurkat cells indicated a monoallelic 12-base pair (bp) insertion at sequences within the –8 kb TAL1 **ChIP-Seq peak**. **Monoallelic insertions of 2–18 bp were found in one out of eight other T-ALL lines (Molt3) and eight out of 146 primary T-ALL samples,** but not in normal cells (controls) from two of the T-ALL patients. These sequence data indicate that **'mutation of the *TAL1* enhancer' (MuTE)** can be acquired in immature T cells or in cells that differentiate into the T lymphoid lineage. The H3K27Ac profile of the Jurkat MuTE suggested that this genomic alteration might create a *TAL1* super-enhancer. Consistent with this

idea, *TAL1* mRNA was expressed from only MuTE alleles in T-ALL cells from five patients with allele-specific polymorphisms in their *TAL1* 3' UTRs.

Moreover, **each MuTE insertion** created at least one **predicted binding site for the MYB transcription factor**.

Genomic fragments spanning **MuTE insertions activated the expression of a luciferase** reporter gene in Jurkat cells, **while MYB knock-down lowered the reporter activity**.

ChIP-Seq demonstrated that MYB could bind to the MuTE insertions in Jurkat and Molt3 cells, but not to corresponding 'normal' regions in TAL–T-ALL cell lines, or a primary *TAL1d* T-ALL. ChIP-Seq reads from Jurkat and Molt3 cells further indicated that MYB bound –8 kb of the *TAL1* TSS on only MuTE alleles, highlighting that MuTE establishes a *cis*-regulatory element that drives oncogenic TAL1 transcription.

Myb associates this region in cooperation with other transcription factors.

Immunoprecipitation assay using antibodies against Myb and westen blot using antibodies against TAL1 show that Myb associates with TAL1. The same result is obtained if IP anti- TAL1 blot anti-Myb (not shown in slide).

CRISPR-Cas9 assay and ChIP-Seq analysis: Deletion of the wild type allele had no effect on endogenous TAL1 mRNA levels, but deletion of the mutant allele completely abrogated endogenous TAL1 expression.

CRISPR-Cas9 assay and ChIP-Seq analysis: Deletion of the wild type allele had no effect on H3K27ac signal and MYB binding, but deletion of the mutant allele completely abrogated H3K27ac signal and MYB binding.

Specific mechanism for Jurkat cells.

Of note, an attempt to delete the MuTE insertion in Jurkat cells using **CRISPR/Cas9 genomic editing**, proved unsuccessful due to problems expanding targeted clones, and presumably reflecting the dependence of Jurkat cellular proliferation on TAL1 protein expression. Consistent with this notion, retrovirus-driven *TAL1* expression has enabled the isolation of MuTE-deleted Jurkat lines, which express lower levels of endogenous *TAL1* mRNA relative to parental Jurkat cells, demonstrating that MuTE is the causative lesion that drives oncogenic *TAL1* transcription.

Indeed, ChIP-Seq of these lines has shown loss of MYB binding and H3K27Ac at the *TAL1* locus, demonstrating that the Jurkat *TAL1* super-enhancer requires the MuTE-inserted MYB-binding sites for efficient *TAL1* transcription. This study thus established the paradigm that **somatic mutation of a non-coding region can introduce binding sites for a transcription factor.** The binding of the transcription factor, in turn, establishes a super-enhancer, and this is significant because it can then activate oncogenic transcription of a nearby proto-oncogene.

Many recent reports describe evidence that **specific chromosome structures play important roles in gene control.** A core principle that has emerged from these studies is that **genes and their regulatory elements typically occur together within specific DNA loop structures, which we have called "insulated neighborhoods."**

Tumor cell gene expression programs are typically driven by somatic mutations that alter the coding sequence or expression of proto-oncogenes, and identifying such mutations in patient genomes is a major goal of cancer genomics.

Dysregulation of proto-oncogenes frequently involves mutations that bring transcriptional enhancers into proximity of these genes.

Transcriptional enhancers normally interact with their target genes through the formation of DNA loops, which typically are constrained within larger **CCCTCbinding factor (CTCF) cohesin**–mediated loops called insulated neighborhoods , which in turn can form clusters that contribute to **topologically associating domains (TADs)**.

This recent understanding of chromosome structure led us to hypothesize that silent proto-oncogenes located within insulated neighborhoods might be activated in cancer cells via loss of an insulated neighborhood boundary, with consequent aberrant activation by enhancers that are normally located outside the neighborhood.

To test this hypothesis, we used chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) to map neighborhoods and other cis-regulatory interactions in a cancer cell genome.

ChIA-PET generates a high-resolution (~5 kb) chromatin interaction map of sites in the genome bound by a specific protein factor. Cohesin was selected as the target

protein because it is involved in both CTCF-CTCF interactions and enhancer-promoter interactions and has proven useful for identifying insulated neighborhoods.

Such CTCF-CTCF loops have been called insulated neighborhoods because disruption of either CTCF boundary causes dysregulation of local genes due to inappropriate enhancer-promoter interactions.

Consistent with this, the Jurkat chromosome structure data showed that the majority of cohesion associated enhancer-promoter interactions had end points that occurred within the CTCF-CTCF loops. **These results provide an initial map of the three-dimensional (3D) regulatory landscape of a tumor cell genome**.

The relationship between genes that have been implicated in T-ALL pathogenesis and the insulated neighborhoods. The majority of genes implicated in T-ALL pathogenesis, as curated from the Cancer Gene Census and individual studies, were located within the insulated neighborhoods identified in Jurkat cells.

Both active oncogenes and silent proto-oncogenes are located within insulated neighborhoods.

TAL1 can be activated by deletions that fuse a promoter less TAL1 gene to the promoter of STIL, and this was observed in many patient deletions

CRISPR/Cas9-mediated deletion of the TAL1 neighborhood boundary in human embryonic kidney (HEK-293T) cells induces the activation of TAL1 proto-oncogene. , In wild type cells TAL1 is silence as evidenced by low H3K27Ac (histone H3 acetylated Lys27) occupancy and RNA-seq.

Deletion of a ~400–base pair (bp) segment encompassing the boundary CTCF site, caused a factor of 2.3 induction of the TAL1 transcript, which suggests that the integrity of the neighborhood contributes to the silent state of TAL1.

Testing of the role of the neighborhood boundary in other cells and in cancer.

In these cancers, a considerable fraction of the mutated neighborhood boundary CTCF sites were affected by multiple mutations.

This article open the way to the definition of new regulatory unit.

Gene regulatory elements and their target genes generally occur within **insulated neighborhoods**, which are chromosomal loop structures formed by the interaction of two DNA sites bound by the CTCF protein and occupied by the cohesin complex.

Individual chromosomes are partitioned into **megabase- sized TADs**, regions with relatively high intradomain DNA interaction frequencies as measured by Hi-C chromosome conformation capture data (Dixon et al., 2012; Nora et al., 2012). These TADs, which have similar boundaries in all human cell types examined, have been proposed to constrain enhancer-gene interactions because most DNA contacts occur within the TADs (Dixon et al., 2012, 2015). This structuring of the genome helps explain why enhancer-gene interactions rarely occur between chromosomes and tend to be constrained within megabase-sized domains.

Enhancer-bound proteins are constrained such that they tend to interact only with genes within these CTCF-CTCF loops. The subset of CTCF sites that form these "loop anchors" thus function to insulate enhancers and genes within the loop from enhancers and genes outside the loop. For these and other reasons, these CTCF-CTCF DNA loops have been called "insulated neighborhoods."

The vast majority of enhancer-gene interactions occur within insulate neighborhoods

In a dozen loci and in multiple cell types, CRISPR/Cas9 deletion of CTCF binding sites at the anchors of insulated neighborhoods has been shown to produce changes in the expression of genes within the neighborhoods and immediately adjacent to the deleted neighborhood boundary.

The finding that cancer cells can activate oncogenes through somatic mutations or epigenetic modifications that disrupt insulated neighborhood boundaries provides additional evidence that neighborhood loop anchors have functional insulating properties.

While different cell types share very similar insulated neighborhood boundaries, the enhancer-gene interactions that occur within these neighborhoods are cell-type specific because enhancer activity is cell-type specific.

TADs are megabase-sized domains with relatively high DNA interaction frequencies and are identified using a Hidden Markov Model-based analysis of Hi-C chromosome conformation capture data (Dixon et al., 2012; Nora et al., 2012). Two

observations argue that TADs are generally composed of, and likely structured by, insulated neighborhoods.

Insulated neighborhoods are a major structuring component of TADs.

Comparison of insulated neighborhoods with loop domains and CTCF contact domains in the same cell type suggests extensive overlap between these structures. For example, 70% of loop domains and 54% CTCF contact domains have the same boundaries as an insulated neighborhood in human lymphoblastoid cells (Rao et al., 2014; Tang et al., 2015). Differences in experimental and analytical methods can explain many of the differences in loop structures reported by various studies; indeed, similarities among loop structures are more evident when data are analyzed with increasing stringency.

Insulated neighborhood models provide a new approach to identify the target genes of disease-associated enhancer variation. Tens of thousands of non-coding genetic variants have been linked with various human diseases and traits in genome-wide association studies (GWASs), and the majority of these variants occur in enhancers (Ernst et al., 2011; Farh et al., 2015; Hnisz et al., 2013; Maurano et al., 2012). The identification of the target genes of these variants is challenging because proximity-based assignment has proven, in some cases, to be inaccurate. Mapping interactions between enhancers and promoters in disease- relevant cells improves the accuracy of the assignment (Grubert et al., 2015; McGeachie et al., 2016; Pomerantz et al., 2009), but this is not always feasible. Because insulated neighborhoods tend to be shared by different cell types, existing maps of insulated neighborhoods should allow investigators to develop a hypothesis regarding the potential target genes of enhancer-associated variation.

Insulated neighborhoods:

- are structural and functional units of gene control
- are used during development to control the diverse cell identities that contribute to complex animals.
- form the mechanistic basis of higher-order chromosome structures, such as topologically associating domains (TADs)

- genetic and epigenetic perturbations of neighborhood boundaries contribute to disease.