Lesson 6- 240518

In the previous lesson, we have discussed the roles of three-dimensional chromosome structure and powerful cis-acting elements (super-enhancers) in regulating gene transcription. In the first step, somatic acquisition of **insertions in a non-coding region** near *TAL1* established a **super-enhancer** that activated *TAL1* transcription in T-ALL cells. The mutation creates a site for the transcription binding sites Myb that remodels the chromatin and induce TAL1 expression.

In the second step, **disruption of boundaries** in the TAL1 Insulated Neighborhood **Activates TAL1** Transcription in T-ALL Cells. CRISPR/Cas9-mediated deletion of 400 bp spanning this CTCF site in human embryonic kidney cells or primary human T cells resulted in higher levels of TAL1 transcripts. This deletion was also found to permit interactions between sequences normally compartmentalized within the TAL1 or adjacent insulated neighborhood.

Cell identity—more specifically, the identity of one or another differentiated cell type—is controlled in large part by the action of transcription factors (TFs) that recognize and bind specific sequences in the genome and thereby regulate gene expression. While nearly half of all of the TFs encoded in the human genome are expressed in any one cell type, a small number of **master TFs**, sometimes called **lineage regulators**, are sufficient to establish control of the gene expression programs that define cell identity. Thus, the control of transcriptional programs that characterize **normal differentiated cell states** is dominated by these master TFs, which are expressed at high levels in selected cell types, tend to co-occupy most enhancers together with other master TFs, and typically regulate their own genes through an autoregulatory loop that forms the **core transcriptional regulatory circuitry of a cell**.

The **master TFs** of any one cell type can be found at the enhancers of a majority of the **active cell-type-specific genes** and may thus account for much of the organization of **cell type- specific gene expression programs**.

The master TFs bind cooperatively to enhancer DNA elements and **recruit coactivators and the transcription apparatus**. These TFs can activate transcription from the enhancer elements themselves, producing enhancer RNAs (eRNAs) that bind certain TFs and cofactors and contribute to enhancer maintenance and dynamics. Enhancers, which tend to be cell type specific because they are generally established by cell-type-specific master TFs, have been mapped in a broad spectrum of human tissue types by using epigenetic marks associated with enhancer activity.

Bound by master TFs, **clusters of enhancers known as superenhancers (SEs)** regulate genes that play prominent roles in cell identity or specialized cellular function. Enhancer-associated proteins and RNAs, including TFs, cofactors such as Mediator, chromatin regulators, signaling factors, RNA polymerase II (RNAPII), enhancer-associated chromatin marks (H3K27Ac), and eRNAs, are all found at especially high density at the constituent enhancers of SEs.

The constituent enhancers of SEs physically associate with one another and can function as independent or interdependent components of these large transcription-regulating complexes to drive high-level expression of their associated genes.

Enhancers and super-enhancers become physically juxtaposed to target gene promoters by **looping of the chromatin** and, having become so, stimulate transcription from these promoters.

Although **enhancers** can activate any gene, they **are physically and functionally constrained to act within insulated neighborhoods**. Insulated neighborhoods are **chromosomal loop structures formed by the interaction of two DNA sites bound by the CTCF protein and occupied by the cohesin complex**. These chromosomal neighborhoods engender specific enhancer-gene interactions and are essential for **normal gene activation and repression**. The CTCF-CTCF loops that form insulated neighborhoods are the mechanistic basis of higher-order chromosome structures, such as **topologically associating domains (TADs)**, and form a chromosome scaffold that is largely preserved throughout development.

Normal cell states depend on signals received from the tissue microenvironments. Much of this contextual information is delivered by signaling pathways to SEs and, to a lesser extent, to typical enhancers.

SEs have been shown to integrate input from Wnt, TGFb, and LIF signaling pathways operating within embryonic stem cells (ESCs). This signal integration is thought to be a consequence of the ability of master TFs to recruit signal-activated TFs to enhancer sites previously established by the master TFs.

Enhancers that are responsive to various types of afferent signaling are thought to be highly dynamic because the activity of signaling TFs is linked to their destruction. Some TFs may recognized by specific ubiquitin ligases operating in the nucleus, leading to their proteasome-mediated destruction. Remarkably, ubiquitylation and degradation of many transcriptional activators occurs at enhancer/promoter sites and can be required for efficient transcription.

DNA is packaged into nucleosomes, which consist of histones that are substrates for chromatin regulators that can modify various amino acid residues or bind in a modification-dependent manner to these histones. The positioning of nucleosomes on DNA can also influence gene control, for example, by limiting

access of TFs to regulatory sequences, and ATP-dependent remodeling complexes influence transcriptional states by mobilizing nucleosomes. The common functional theme of these regulators is that they facilitate maintenance of positive or negative gene expression states.

DNA methylation contributes to gene control at three levels. **Methylation of enhancer and promoter sites** can prevent TF binding and thus **silence genes**. **Methylation of CTCF loop anchor sites** prevents CTCF binding and can thus alter insulated neighborhood structure. Cytosine methylation and hydroxymethylation present spatially positioned chemical motifs that can be recognized by chromatin-associated proteins (e.g., MECP2), thereby influencing transcriptional regulation.

In summary, normal cells have transcriptional programs that are established and maintained by master TFs that regulate genes by binding specific enhancer elements, which in turn interact with genes within insulated neighborhoods. The maintenance of normal cell states depends on the tissue environment, and such information is delivered by signaling pathways ultimately to enhancers. Maintenance of cell identity and dynamics of cell states also depend on a large number of histone readers, writers, and erasers, as well as regulators of DNA methylation, that together ensure chromatin states are appropriate for positive and negative gene regulation.

Transcriptional dysregulation arises in cancer from disease- defining genetic alterations either **indirectly**, via **mutation of signaling factors converging on transcriptional control**, or **directly**, **via genetic alterations in gene control factors themselves**. Cancer-associated genetic alterations can affect proteins participating in nearly all levels of transcriptional control, including trans-factors (TFs, signaling proteins, cofactors, chromatin regulators, and chromosome structuring proteins) and cis-elements (enhancers, promoters, and insulators).

The TFs that are deregulated in cancer cells and have the potential to produce profound changes in gene expression programs can be considered to fall into three classes:

- **master TFs** involved in organizing cell identity,
- **proliferation control TFs** that amplify transcriptional output
- **signaling TFs involved** in dynamic changes in the control machinery occurring in response to extracellular signals.

The activation of a master TF that is normally expressed in early development, such as the pluripotency TF OCT4, or activation of a master TF that is normally expressed early in a specific lineage, such as TAL1 in T cells, can alter core regulatory circuitry and activate additional genes that are normally expressed in more embryonic states.

Genes encoding the MYC and P53 proliferation control TFs, a classic oncogene and tumor suppressor gene, are among the most frequently mutated genes

in cancer. MYC can have profound effects because it can function to amplify the entire gene expression program, and P53 is a powerful tumor suppressor because of its ability to arrest progress through the cell cycle or induce apoptosis.

Dysregulation of signaling pathways is a common feature of cancer cells; a dysregulated signaling TF can profoundly change the gene expression program through its binding to enhancers occupied by master TFs and dysregulated signaling can even stimulate super-enhancer formation. These TFs signal to RNAPII through transcriptional cofactors, defined here as regulatory components that do not bind directlyto DNA in a sequence-specific manner. Exemplifying this class of transcriptional signaling proteins are the components of the Mediator complex, which is recruited to enhancer-promoter regions by TFs in the context of transcription activation. Genetic alterations of Mediator-complex-encoding genes are observed frequently in prostate cancer and in many uterine myomas. Beyond these discrete diseases, the genome-wide activities of Mediator in gene control would be expected to function broadly in all cancer- associated transcription. Interestingly, few cancer-associated alterations are identified in the core RNAPII complex itself , suggesting that coordinated transcriptional signaling upstream of polymerase favors the neoplastic cell state more than alterations of this core complex.

Efficient transcriptional signaling from enhancers to promoters is often chromatin dependent, mediated by specialized transcriptional cofactors that physically associate with or biochemically modify the genome to reinforce gene activation or repression. Chromatin regulators function globally, so their dysregulation can also have profound effects on the gene expression program of cells. In some tumors, chromatin regulators have become fused to transcriptional cofactors, producing gene-specific effects, such as those observed in acute lymphoblastic leukemia cells with MLL-AF4 fusions.

Insulated neighborhoods contribute to proper positive and negative gene control, so alterations in chromosome-structuring proteins that establish and maintain insulated neighborhood boundaries would be expected to have profound effects on a cell's overall gene expression program. Cancer genome sequencing has revealed that **somatic mutations occur in CTCF and cohesin coding sequences in various solid tumors and leukemias**, and it seems likely that these mutations contribute to oncogenesis by altering insulated neighborhoods throughout the genome, perhaps rendering chromatin generally more permissive to oncogenic transcriptional signaling.

Tumor cells acquire SEs at oncogenic driver genes, and they do so through many different mechanisms:

-Translocation
- Focal amplification
- Small insertion

Mutations that alter insulator sequences of oncogene-containing insulated neighborhoods appear to make important contributions to the dysregulation of gene expression observed in some cancers.

The concept of **oncogene addiction** refers to the behavior of cancer cells that exhibit an absolute **dependence on oncogenes** that were initially acquired during **multi-step tumorigenesis** and remain critical to the ongoing proliferation and viability of these cells long after they have progressed to a fully neoplastic state. As of late, this concept has been extended to include other changes acquired during tumor progression that fostered the early development of a tumor and continue to be absolutely essential to its continued growth. Included among these are the **dysregulated transcriptional programs** operating in certain tumor cells, yielding the concept of transcriptional addiction.

Dysregulated transcriptional programs may be target for drug discovery:

- Increased turnover of oncogenes
- Interfering with cooperation between TFs and cofactors
- Targeting chromatin remodeling enzymes
- Genome editing of specific regulatory regions
- Activation of proteasome degradation machinery