TSS MODULE PRIMER:

REVIEW OF TRANSCRIPTION, PROMOTER STRUCTURE AND CHROMATIN PACKAGING

JAMIE SIDERS, MEG LAAKSO & WILSON LEUNG

Eukaryotic organisms are comprised of many distinct cell types, each of which possesses an identical nuclear genome. However, each cell type (e.g., liver, kidney, neuron) has a unique morphology and function. This is made possible by the fact that each cell type transcribes a distinct set of genes. Genes within a given cell type that are transcribed are said to be ***expressed*** and the complement of all the RNAs transcribed are aptly known as the ***transcriptome***. The transcriptome includes all types of RNA, including messenger RNA (mRNA) and functional RNA molecules such as ribosomal RNA (rRNA). Expression of the correct set of genes at the correct time is necessary for proper organismal development, cellular function and maintenance of homeostasis in cells. Transcription is therefore a highly regulated process in eukaryotes.

These TSS modules aim to familiarize students with the process of identifying the TSS and defining the core promoter region for a gene. This primer therefore includes some biological concepts that may helpful to review prior to working your way through the four TSS Modules, including: a brief review of transcription, transcription initiation, promoter structure, and chromatin.

REVIEW OF TRANSCRIPTION

Transcription is the biological process by which a *template* strand of genomic DNA is used to generate a single-stranded RNA molecule known as the transcript. Transcription begins at a position in the genomic DNA known as the transcription start site (TSS). The TSS is found within a region of the genomic DNA known as the **core promoter**. The core promoter is bound by the major enzyme responsible for transcription – RNA Polymerase II (RNA Pol II).

It is RNA Pol II that synthesizes the RNA transcript by incorporating ribonucleotides that are complementary to the bases found in the template strand of genomic DNA. RNA Pol II does not act alone, however. There are cellular proteins known as transcription factors that are critical in ensuring the RNA Pol II enzyme is recruited to the core promoter and which help to regulate transcription. **Figure 1** shows a basic overview of transcription, with special emphasis on RNA Pol II and the core promoter region.

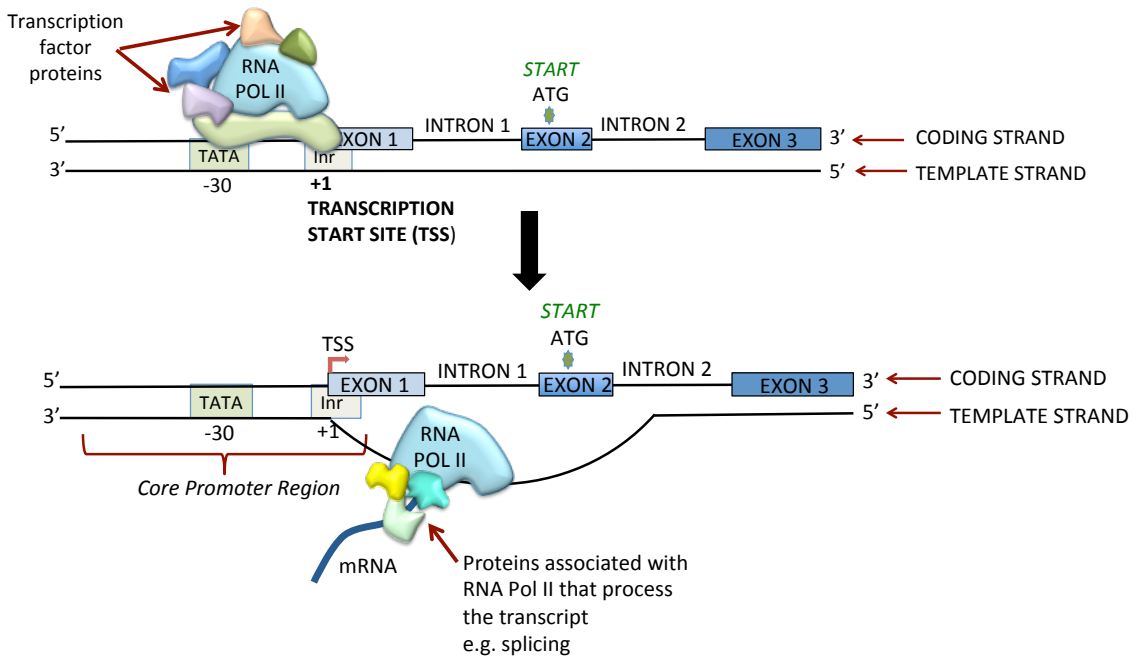

Figure 1 Overview of transcription

RNA PROCESSING

Transcripts that can be translated into proteins are known as messenger RNAs (mRNAs) and are processed in three ways prior to export to the cytoplasm where translation occurs. The three mRNA processing events are:

1. **5' Cap**: a 7 methyl-guanosine is added to the 5' end of the RNA transcript
2. **Splicing**: Intronic regions are removed via splicing. Conserved nucleotide sequences found in the genomic DNA and corresponding transcript, including splice donor (GU) and splice acceptor (AG) sites are critical for faithful removal of introns by the proteins responsible for splicing. Alternative patterns of splicing can result in different gene isoforms.
3. **Poly-A Tail:** A string of non-templated adenines (~30-200) is added to the 3' end of the transcript.

Processing of the transcript in these three ways results in a completed or 'mature' mRNA molecule. **Figure 2** below summarizes eukaryotic gene structure from genomic DNA to mature transcript.

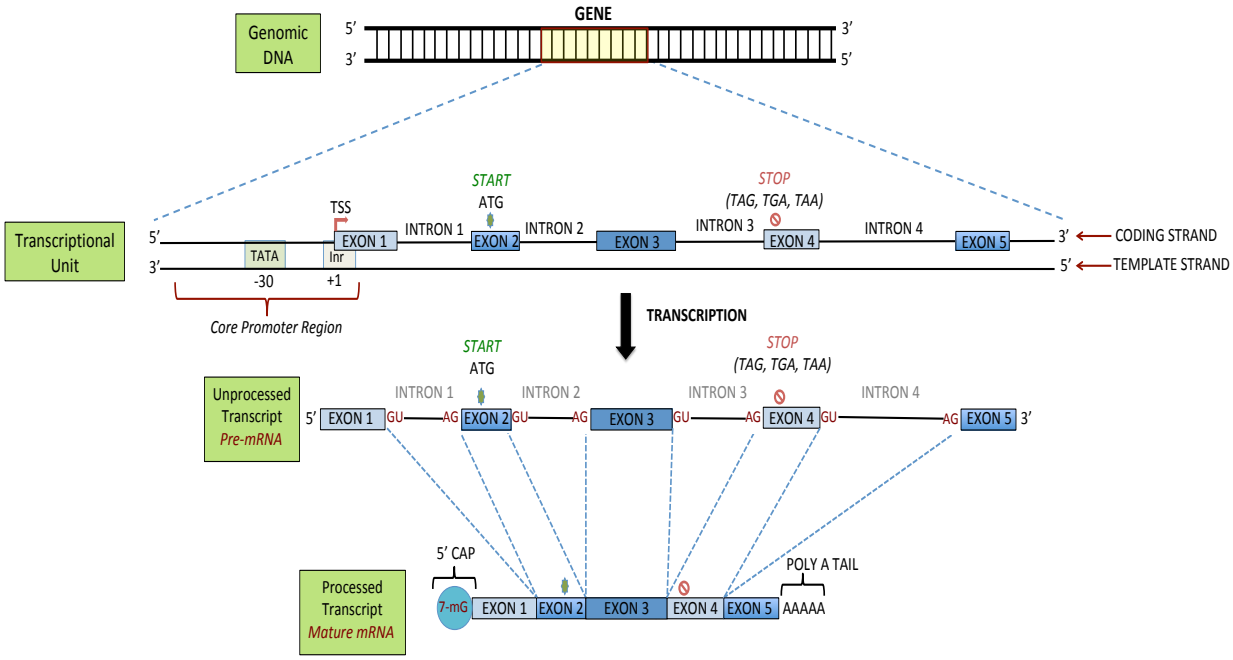

Figure 2 Summary of eukaryotic gene structure from genomic DNA to the final processed transcript

Q1: On Figure 2, use brackets and labels to indicate the following parts of the final processed transcript: 1. 5' untranslated region (5' UTR), the coding sequence (CDS) and the 3' untranslated region (3' UTR).

THE CORE PROMOTER AND PROMOTER STRUCTURE

The DNA sequence surrounding the first transcribed exon of a gene is required for recruitment of RNA Pol II to the genomic DNA and is known as the **core promoter (Figure 1)**. The core promoter for a given gene consists of a unique complement of *specific nucleotide sequences* referred to as DNA motifs. Two such motifs, known as the TATA box and Inr motif, are shown in Figure 1 and Figure 2, to help you visualize the position of these nucleotide sequences relative to the start of transcription. The nucleotide position on the genomic DNA that corresponds to the start of transcription is known as the **transcription start site (TSS)** and for any given gene, the TSS position is designated as position +1, and all other core promoter motif positions are designated relative to the +1 TSS position. The DNA sequences that make up these motifs are critical because transcription factor proteins directly bind to these bases and help to recruit and properly position RNA pol II on the genomic DNA (**Figure 1**). You will learn more about motifs in TSS Module 2. Once RNA Pol II has been recruited and properly positioned at the TSS on the genomic DNA, the process of transcription initiation and elongation can begin.

Promoters can be classified into one of three major categories: Peaked, Broad, and Intermediate (**Figure 3**). Classification is based on the number and distribution of the transcription start sites (TSSs) utilized by the RNA polymerase. This is readily seen in Figure 3, where the location of transcription start sites for each promoter type are shown as arrows and peaks in the figure. Relative levels of transcripts generated from each TSS are demonstrated by the size and thickness of the arrow and by the height of the peaks. 'Promoter shape' refers to the distribution of the TSSs for a given promoter. Peaked promoters have narrow shape because they possess only a *single* TSS. Broad promoters have a broad shape resulting from the fact that there are multiple TSSs within the promoter that are utilized by RNA Pol II. Given that peaked promoters have a single TSS, transcription from peaked promoters is sometimes referred to as focused transcription, whereas transcription from broad promoters is referred to as dispersed transcription. Genes that exhibit tissue- or development-specific gene expression often have peaked promoters, while broad promoters are often associated with genes that are expressed ubiquitously in the organism.

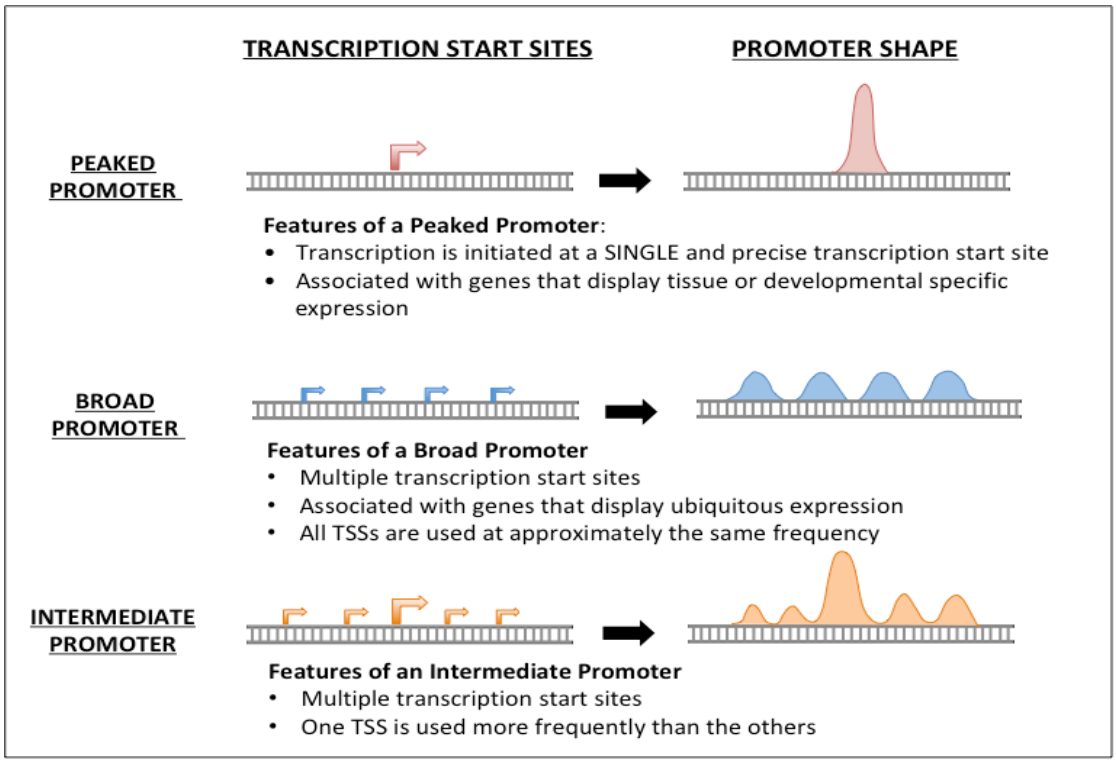

Figure 3 Three major types of core promoters: Peaked, Broad, and Intermediate

The classification of core promoters is rarely cut and dried. Recent analysis has shown that most genes in the fruit fly *Drosophila melanogaster* have multiple TSSs and that promoter type may be classified as "Intermediate". Intermediate promoters have multiple TSSs, but some TSSs are used at higher frequencies than others (Figure 3, bottom panel).

REVIEW OF CHROMATIN PACKAGING

It is important to remember that the process of transcription occurs on DNA that is packaged into chromatin. Chromatin refers to the complex of DNA, histone proteins and non-histone proteins that interact with each other to form an intact chromosome within the cell nucleus.

The amino acids of histone proteins can be chemically modified, which allows the DNA to be wound tighter or looser around the histone octamer. Areas of the genome in which the DNA is packaged comparatively loosely is referred to as euchromatin (**Figure 4**). DNA that is packaged tightly around the histone octamer is referred to as heterochromatin. Euchromatic regions of the genome are generally transcriptionally active, whereas heterochromatic regions of the genome are comparatively transcriptionally inactive.

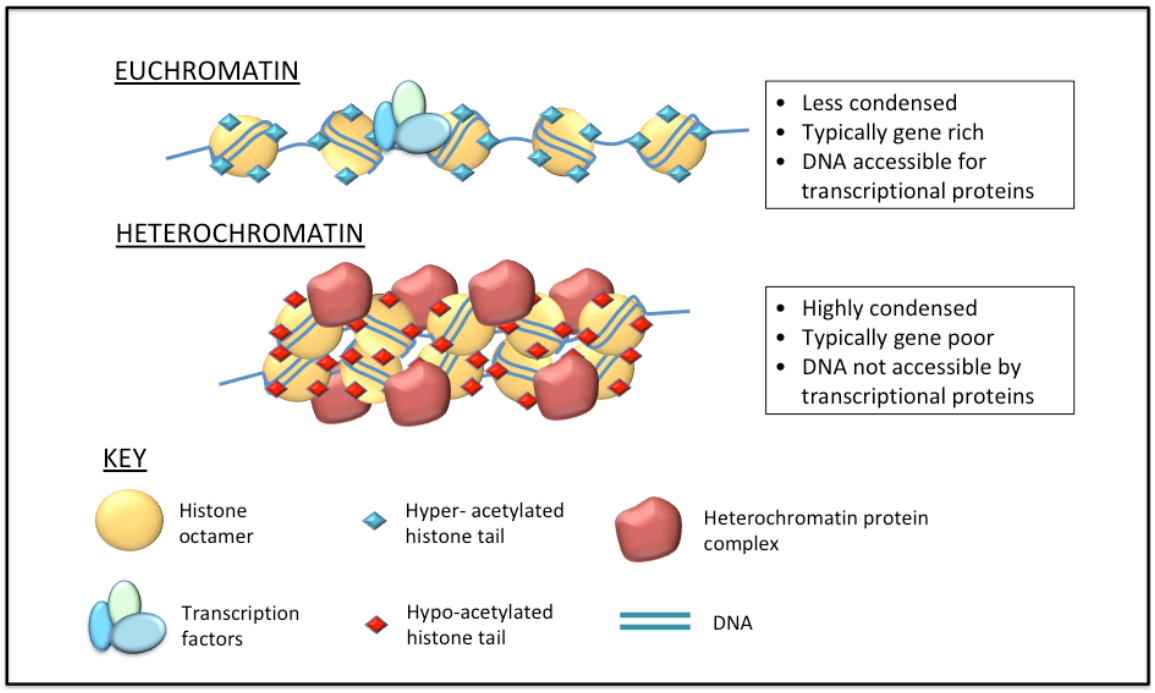

Figure 4 Characteristics of euchromatin versus heterochromatin

In the TSS modules you will see that characterization of chromatin packaging in a particular area of the genome can be used as evidence to assist in the process of TSS annotation.