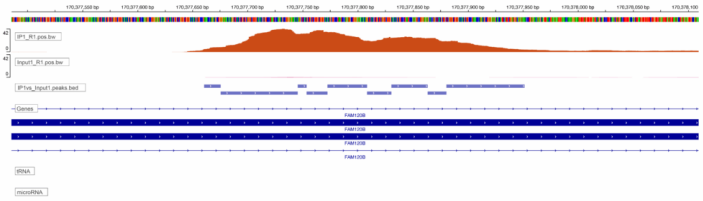
What are peaks?

Designing effective RNA therapeutics often requires understanding not just which RNAs are present, but which proteins interact with them. RNA-binding proteins (RBPs) can control an RNA's processing and function, and these controls can change in health and disease. Identifying how these RBPs exert their effects on their target RNAs is critical for determining how to modulate these interactions to produce an intended therapeutic outcome.

Fortunately, drug developers can gain insight into these interactions through sequencing and peak calling-based approaches. This process first involves preparing two RNA libraries. The first library, called the IP, is immunoprecipitated using antibodies specific to the RBP and contains the RNA targets the RBP was bound to. The second library, called the input, is generated without an antibody and contains a size-selected pool of total RNA. Drug developers can sequence these libraries, which can be used to determine the location and prevalence of RBP interactions through peak calling.



Peak calling

Drug developers use peak calling algorithms to identify regions on RNA where an RBP interacts. In practice, this is a multi-step process that involves many specialized algorithms. Sequencing reads first need to be cleaned of any extraneous sequences introduced by the library preparation workflow. Alignment algorithms then map these reads onto the transcriptome, identifying where these RNA sequences originated from. Visually, this results in regions of an RNA with enriched read coverage, called "clusters", and represents putative RBP interaction sites.

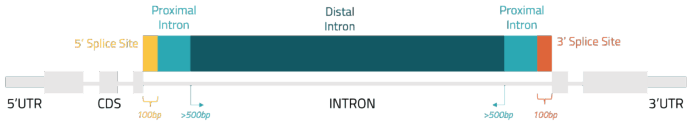
Once read clusters are identified, drug developers use peak calling algorithms to distinguish true binding interactions from false ones. Components used in the library preparation workflow may bind to RNA nonspecifically, resulting in false signal. For each cluster, peak calling algorithms compare the level of enrichment in the IP and input libraries. Clusters that are significantly enriched in the IP over the input are deemed "peaks" and represent high-confidence RBP-RNA binding sites.

Peak annotation

After locating the binding sites, drug developers must determine what gene is present in that region. They can do this through peak annotation.

Peak annotation begins with a high-quality reference genome. Based on the reference genome, drug developers identify different regions of a gene, such as introns, UTRs, and coding sequences. When overlayed with the peak data, developers can see which regions of a gene contain peaks. This is important because RBPs may have different functions on gene expression depending on where on the gene it binds.

In some cases, there can be multiple isoforms of a gene with overlaps in the annotation of UTRs, introns, and coding sequences. When this occurs, researchers can use a hierarchy to determine the most likely bound feature, such prioritizing coding regions over UTRs.



Peaks at Eclipsebio

At Eclipsebio, we use peak calling and annotation in our eCLIP technologies. Our RBP-eCLIP assay locates where and how proteins bind to RNA, and our miR-eCLIP+ assay does the same for miRNAs and siRNAs. Our m6a-eCLIP assay reveals methylated bases. Using our years' worth of reference sequence data and motifs, we can also help support effective peak annotation.

Interested in revealing RNA insights through peaks? Contact Eclipsebio today.