Annotation Source • ENCODE project • UCSC genome browser Genome Annotation • Bed12 (transcripts)

Peak File

- DNA-binding sites
- RNA-binding sites

Bed6 Format

- Unique peak ID
- Store original information

Bedtools, intersection

Peak + Transcripts (DNA)

- Multiple features of same peak
- Multiple transcripts of same gene
- Multiple genes of same peak

Bed6

Peak + Transcripts (RNA)

- Multiple features of same peak
- Multiple transcripts of same gene
- Multiple genes of same peak

Peak-Gene Pairwise (DNA)

- Priority feature (TSS (<=1kb), TSS (1-2kb),
 5' UTR, CDS, ..., TTS (<=1kb), ...)
- Transcripts with closest distance
- Priority gene class (protein_coding, IncRNA, miRNA, snoRNA...)

Peak-Gene Pairwise (RNA)

- Priority feature (CDS, codon, 5'UTR, 3' UTR, Exon, Intron...)
- Transcripts with maximum overlap length
- Priority gene class (protein_coding, IncRNA, miRNA, snoRNA...)

Another Round of Annotation (optional)

- Cpg islands
- Microsatelline
- Transposons
- Repeat elements (Alu...)
- •

- snoRNA
- miRNA
- tRNA and rRNA
- GWAS catalog
- •

Peak-Gene + Extra annotation

- Comprehensive annotation
- Customized analysis