ChIPseq

染色质免疫共沉淀技术（Chromatin Immunoprecipitation，ChIP），抗体特异结合DNA结合蛋白，免疫沉淀，可以用于捕获蛋白质（如转录因子，组蛋白修饰）的DNA靶点。

ChIP与二代测序结合称为ChIP-seq技术。分为四步，

\* cross-linking，交联

\* sonication，超声打碎

\* IP，免疫沉淀

\* sequencing，测序

DNA和蛋白质交联(cross-linking)，超声(sonication)将染色体随机打断，利用抗原抗体的特异性识别(IP)，把目标蛋白相结合的DNA片段沉淀下来，反交联释放DNA片段，最后测序(sequencing)。

##########

peak calling

##########

peak，protein binding site。

##########

peak annotation

##########

peak annotation，binding site的相关基因注释。

注释的结果文件中，包含两种注释：

\* genomic annotation：annotation这列。

注释的peak的位置，peak可以落在UTR、内含子、外显子

\* nearest gene annotation：剩余的其他列。

最近基因是peak相对于转录起始位点的距离，不管peak是落在内含子或者别的什么位置上，即使它落在基因间区上，我都能够找到一个离它最近的基因（即使它可能非常远）。

第三种注释：

##########

BED文件

##########

BED（Browser Extensible Data）文件，peak calling的输出。

3个必须字段：

\* chrom，染色体名称

\* chromStart，染色体起始位点。起始于0。

\* chromEnd，染色体终止位点

可选的9个字段：

\* name，名字

\* score，分值(0-1000)。用于genome browser展示时上色。

\* strand，正负链。对于ChIPseq数据来说，一般没有正负链信息。

\* thickStart，画矩形的起点

\* thickEnd，画矩形的终点

\* itemRgb， RGB值

\* blockCount，子元件（比如外显子）的数目

\* blockSizes，子元件的大小

\* blockStarts，子元件的起始位点

一般情况下只用前5字段。这也是做peak calling的MACS的输出。

##########

ChIPseeker包

##########

##########

ChIPseq的几个问题

##########

1、为什么用TxDb.Hsapiens.UCSC.hg19.knownGene，而不用hg38？

用什么基因组，这取决于你最初fastq用BWA/Bowtie2比对于某个版本的基因组，你最初用了某个版本，后面就得用相应的版本，不能混，因为不同版本的位置信息有所不同。

2、为什么ChIPseeker支持所有物种

（1）首先，Bioconductor提供了30个TxDb包，可以供我们使用，这当然只能覆盖到一小部分物种

（2）从UCSC或者Ensembl获得，ChIPseeker支持所有物种，就是因为UCSC和ensembl上所有的基因组都可以被ChIPseeker支持。

可以使用GenomicFeatures包函数来制作TxDb对象：

① makeTxDbFromUCSC： 通过UCSC在线制作TxDb

② makeTxDbFromBiomart: 通过ensembl在线制作TxDb

③ makeTxDbFromGRanges：通过GRanges对象制作TxDb

④ makeTxDbFromGFF：通过解析GFF文件制作TxDb