```r
setwd("~/Ting_signature_analyses/")

annotation_methylation_CpGs <- read.delim("Final.Tumor.sig.annotation.txt")
annotation_methylation_CpGs$CHR_hg38 <- substr(annotation_methylation_CpGs$CHR_hg38, 4, 6)


if (!exists("pchic")){
  load("pchic.RData")
  pchic <- pchic[, c(1:10)]
  List_Promoter <- unique(paste(pchic$baitChr, pchic$baitStart, sep = "_"))
  colnames(pchic)[c(1:5, 6:10)] <- rep(c("chr", "start", "end", "ID", "Name"), 2)
  PCHiC_bed <- unique(rbind(pchic[, c(1:3, 5)], pchic[, c(6:8, 10)]))
  PCHiC_GRange <- GRanges(
    seqnames = PCHiC_bed$chr,
    IRanges(start = PCHiC_bed$start, end = PCHiC_bed$end),
    Gene_Pchic = PCHiC_bed$Name,
    start_fragment = PCHiC_bed$start,
    end_fragment = PCHiC_bed$end
  )
  PCHiC_GRange$ID <- paste(PCHiC_bed$chr, PCHiC_bed$start, sep = "_")
  colnames(pchic) <- c("chr_bait", "start_bait", "end_bait", "ID_bait", "Name_bait", "chr_oe", "start_oe
  pchic$IDbait <- paste(pchic$chr_bait, pchic$start_bait, sep = "_")
  pchic$IDoe <- paste(pchic$chr_oe, pchic$start_oe, sep = "_")
  # pchic <- pchic[, c("IDbait", "Name_bait", "IDoe", "Name_oe")]
}

if (!exists("Blueprint_network")){
  Blueprint_network <- read.csv("BLUEPRINT_fragments_good.tsv", sep = "\t")
  Blueprint_network <- dplyr::select(Blueprint_network, "chr", "start", "end", "type", "ensembl", "gene_
  Blueprint_network <- Blueprint_network %>% separate_rows(., gene_names, sep = " ")
}


if (!exists("Final.Tumor.RNAseq.sig")){
  Final.Tumor.RNAseq.sig <- read.delim("Final.Tumor.RNAseq.sig.txt")
  Final.Tumor.RNAseq.sig <- Final.Tumor.RNAseq.sig[,"Gene"]
}

Blueprint_Granges <- GRanges(
  seqnames = Blueprint_network$chr,
  ranges = IRanges(start = Blueprint_network$start, end = Blueprint_network$end),
  Gene_name = Blueprint_network$gene_names,
  Promoter = Blueprint_network$type
)
```

```r
#### Overlapping between Chromatin fragment from PCHiC data and Blueprint data


overlaps_Blueprint_pchic <- findOverlaps(PCHiC_GRange, Blueprint_Granges)
match_hit_BP_Pchic <- data.frame(mcols(PCHiC_GRange[queryHits(overlaps_Blueprint_pchic),]), data.frame(

message("=== Number of nodes in BP network ===")
```

```
## === Number of nodes in BP network ===
```

```r
length(unique(match_hit_BP_Pchic$ID))
```

```
## [1] 253148
```

```r
message("== Number of promoter in BP network ===")
```

```
## == Number of promoter in BP network ===
```

```r
match_hit_BP_promoter_pchic <- match_hit_BP_Pchic[match_hit_BP_Pchic$Promoter == "P",]
length(unique(match_hit_BP_promoter_pchic$ID))
```

```
## [1] 20817
```

```r
Genes_BP <- unique(match_hit_BP_Pchic$Gene_name)
Genes_BP_Promoter_signature <- Genes_BP[Genes_BP %in% Final.Tumor.RNAseq.sig]

message("== Number of Promoter in signature ==")
```

```
## == Number of Promoter in signature ==
```

```r
length(Genes_BP_Promoter_signature)
```

```
## [1] 1211
```

```r
#### Overlapping between CpGs from signature and PCHiC data


CpGs_GRanges <- GRanges(
  seqnames = annotation_methylation_CpGs$CHR_hg38,
  ranges = IRanges(start = annotation_methylation_CpGs$Start_hg38, end = annotation_methylation_CpGs$End
  chr_cpg = annotation_methylation_CpGs$CHR_hg38
)

overlaps_CpGs_pchic <- findOverlaps(PCHiC_GRange, CpGs_GRanges)
match_hit_CpGs_Pchic <- data.frame(mcols(PCHiC_GRange[queryHits(overlaps_CpGs_pchic),]), data.frame(mcol

message("=== Number of CpGs sig nodes ===")
```

```
## === Number of CpGs sig nodes ===
```

```r
length(unique(match_hit_CpGs_Pchic$ID))
```

```
## [1] 811
```

```r
#### Overlapping between the overlapped CpGs in PCHiC data and Blueprint data


CpGs_pchic_GRanges <- GRanges(
  seqnames = match_hit_CpGs_Pchic$chr_cpg,
  ranges = IRanges(start = match_hit_CpGs_Pchic$start_fragment, end = match_hit_CpGs_Pchic$end_fragment
  chr_cpg = match_hit_CpGs_Pchic$chr_cpg
)

overlaps_CpGs_Blueprint <- findOverlaps(Blueprint_Granges, CpGs_pchic_GRanges)
match_hit_CpGs_Blueprint <- data.frame(mcols(Blueprint_Granges[queryHits(overlaps_CpGs_Blueprint),]), da

message("=== CpGs nodes in BP promoter ===")
```

```
## === CpGs nodes in BP promoter ===
```

```r
CpGs_promoter <- match_hit_CpGs_Blueprint[match_hit_CpGs_Blueprint$Promoter == "P", "Gene_name"]
length(CpGs_promoter)
```

```
## [1] 442
```

```r
message("=== promoter in CpGs sig ===")
```

```
## === promoter in CpGs sig ===
```

```r
CpGs_unique_promoter <- unique(CpGs_promoter)
length(CpGs_unique_promoter)
```

```
## [1] 337
```

```r
#### Promoters from the overlaps between PCHiC blueprint and CpGs


Genes_sig_CpGs_sig <- CpGs_promoter[CpGs_promoter %in% Final.Tumor.RNAseq.sig]
message("=== Number of nodes from CpGs sig that are in gene signature list ===")
```

```
## === Number of nodes from CpGs sig that are in gene signature list ===
```

```r
length(Genes_sig_CpGs_sig)
```

```
## [1] 28
```

```r
message("=== Number of genes from CpGs sig that are in gene signature list ===")
```

```
## === Number of genes from CpGs sig that are in gene signature list ===
```

```r
length(unique(Genes_sig_CpGs_sig))
```

```
## [1] 17
```

```r
#### Neighbor of the CpGs signatures

Neighbor_network <- unique(rbind(pchic[pchic$IDbait %in% match_hit_CpGs_Pchic$ID,],pchic[pchic$IDoe %in

colnames(Neighbor_network) <- rep(c("chr", "start", "end", "ID", "Gene_name"), 2)
Neighbor_nodes <- unique(rbind(Neighbor_network[,c(1:5)], Neighbor_network[,c(6:10)]))

Promoter_neighbors <- match_hit_BP_promoter_pchic[match_hit_BP_promoter_pchic$ID %in% Neighbor_nodes$ID
message("=== Number of neighbor nodes in promoter ===")
```

```
## === Number of neighbor nodes in promoter ===
```

```r
length(unique(Promoter_neighbors$ID))
```

```
## [1] 2451
```

```r
message("=== Number of promoter of genes in neghbor CpGs sig ===")
```

```
## === Number of promoter of genes in neghbor CpGs sig ===
```

```r
Promoter_neighbors_genes <- unique(Promoter_neighbors$Gene_name)
length(Promoter_neighbors_genes)
```

```
## [1] 3614
```

```r
#### Promoter that are both in Gene sig and neighbor of CpGs sig
```

```
Promoter_neighbors_genes_sig <- Promoter_neighbors_genes[Promoter_neighbors_genes %in% Final.Tumor.RNAs
message("=== Number of of promoter of genes sig in neghbor CpGs sig ===")
```

```
## === Number of of promoter of genes sig in neghbor CpGs sig ===
```

```
length(unique(Promoter_neighbors_genes_sig))
```

```
## [1] 146
```