

Hmm-based Identification of Transcription factor footprints (HINT) ATAC-Seq Processing Pipeline

I. Preparation of files before running HINT

*note: any file highlighted in **bold** is provided

1. Create BED file for *APP^{V717I}* vs. NDC comparison to input into HINT using all possible peaks
 - a. R code located in **ATAC-seq_DiffBind_chipenrich_Pipeline_final.R** file
#53,598 all possible total peaks here
>ATACSeq.report.V717I_full
#Convert to a data frame
>ATACSeq.report.V717I.out_full_df <-
as.data.frame(ATACSeq.report.V717I.out_full)
#Select columns 1-3 (chr or seqnames, start, end)
>ATACSeq.report.V717I.out_full_df <-
ATACSeq.report.V717I.out_full[, c(1:3)]
#Export as a BED file, e.g. 210511.dATACpeaks_full.V717I.bed (to use for HINT!)
>rtracklayer::export.bed(ATACSeq.report.V717I.out_full_df, con =
"210511.dATACpeaks_full.V717I.bed", format = "bed")
2. Create BED file for *PSEN1^{A79V}* vs. NDC comparison to input into HINT using all possible peaks
 - a. R code located in **ATAC-seq_DiffBind_chipenrich_Pipeline_final.R** file
#53,345 all possible peaks here
>ATACSeq.report.A79V_full
#Convert to a data frame
>ATACSeq.report.A79V.out_full_df <-
as.data.frame(ATACSeq.report.A79V_full)
#Select columns 1-3 (chr or seqnames, start, end)
>ATACSeq.report.A79V.out_full_df <-
ATACSeq.report.A79V.out_full[, c(1:3)]
#Export as a BED file, e.g. 210511.dATACpeaks_full.A79V.bed (to use for HINT)
>rtracklayer::export.bed(ATACSeq.report.A79V.out_full_df, con =
"210511.dATACpeaks_full.A79V.bed", format = "bed")
3. Create BED file for *PSEN2^{N141I}* vs. NDC comparison to input into HINT using all possible peaks
 - a. R code located in **ATAC-seq_DiffBind_chipenrich_Pipeline_final.R** file
#53,345 possible peaks found
>ATACSeq.report.N141I_full
#Convert to a data frame
>ATACSeq.report.N141I.out_full_df <-
as.data.frame(ATACSeq.report.N141I_full)
#Select columns 1-3 (chr or seqnames, start, end)

Hmm-based Identification of Transcription factor footprints (HINT) ATAC-Seq Processing Pipeline

```
>ATACSeq.report.N141I.out_full_df <- ATACSeq.report.N141I.out_full_df[,  
c(1:3)]  
#Export as a BED file, e.g. 210511.dATACpeaks_full.N141I.bed  
>rtracklayer::export.bed(ATACSeq.report.N141I.out_full_df, con =  
"210511.dATACpeaks_full.N141I.bed", format = "bed")
```

4. Merge three replicate .bam files for NDC into one combined .bam file using the **merge_NDC.sh** script
5. Merge three replicate .bam files for *APP^{V717I}* mutation into one combined .bam file using the **merge_V717I.sh** script
6. Merge three replicate .bam files for *PSEN1^{A79V}* mutation into one combined .bam file using the **merge_A79V.sh** script
7. Merge three replicate .bam files for *PSEN2^{N141I}* mutation into one combined .bam file using the **merge_N141I.sh** script
8. Gather the motifs from the CISBP database under the MEME suite
9. Add the collection of CIS-BP motifs under cisbp_hint.tar.gz to the following path:
/home/prvaldes/rgtdata/motifs/
 - a. (base) [prvaldes@tscc-login12 motifs]\$ tar -zxvf cisbp_hint.tar.gz
10. Create motif logos from CIS-BP database
 - a. Run **setupLogoData.py** to generate logos for all available repositories
 - i. (base) [prvaldes@tscc-3-21 rgtdata]\$ python setupLogoData.py – all from the /home/prvaldes/rgtdata/ folder
 - ii. >>> CREATING logos for ['cisbp_hint', 'uniprobe_secondary', 'jaspar_plants', 'hocomoco', 'swissregulon', 'uniprobe_primary', 'jaspar_vertebrates']
>>> cisbp_hint

II. Running the HINT program

11. Run TF footprinting for *APP^{V717I}* vs. NDC comparison using the **RGT-HINT-ATAC-AD_V717I_NDC_footprinting.sh** file
 - a. (base) [prvaldes@tscc-2-14 HINT]\$ bash RGT-HINT-ATAC-AD_V717I_NDC_footprinting.sh
12. Run TF footprinting for *PSEN1^{A79V}* vs. NDC comparison using the **RGT-HINT-ATAC-AD_A79V_NDC_footprinting.sh** file
 - a. (base) [prvaldes@tscc-2-14 HINT]\$ bash RGT-HINT-ATAC-AD_A79V_NDC_footprinting.sh

Hmm-based Identification of Transcription factor footprints (HINT) ATAC-Seq Processing Pipeline

13. Run TF footprinting for *PSEN2*^{N141I} vs. NDC comparison using the **RGT-HINT-ATAC-AD_N141I_NDC_footprinting.sh** file
 - a. (base) [prvaldes@tscc-2-14 HINT]\$ bash RGT-HINT-ATAC-AD_A79V_NDC_footprinting.sh
14. Perform motif matching for *APP*^{V717I} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_V717I_NDC_motif_matching_cisbp_full.sh** file
 - a. (base) Phobes-MacBook-Pro:Shell_Scripts phoebevaldes\$ bash RGT-HINT-ATAC-AD_V717I_NDC_motif_matching_cisbp_full.sh
15. Perform motif matching for *PSEN1*^{A79V} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_A79V_NDC_motif_matching_cisbp_full.sh** file
 - a. (base) Phobes-MacBook-Pro:Shell_Scripts phoebevaldes\$ bash RGT-HINT-ATAC-AD_A79V_NDC_motif_matching_cisbp_full.sh
16. Perform motif matching for *PSEN2*^{N141I} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_N141I_NDC_motif_matching_cisbp_full.sh** file
 - i. (base) Phobes-MacBook-Pro:Shell_Scripts phoebevaldes\$ bash RGT-HINT-ATAC-AD_N141I_NDC_motif_matching_cisbp_full.sh
17. Perform differential analysis for *APP*^{V717I} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_V717I_NDC_differential_cisbp_full.sh** file
 - i. (base) [prvaldes@tscc-3-0 Scripts_2021]\$ bash RGT-HINT-ATAC-AD_V717I_NDC_differential_cisbp_full.sh
18. Perform differential analysis for *PSEN1*^{A79V} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_A79V_NDC_differential_cisbp_full.sh** file
 - i. (base) [prvaldes@tscc-3-3 Scripts_2021]\$ bash RGT-HINT-ATAC-AD_A79V_NDC_differential_cisbp_full.sh
19. Perform differential analysis for *PSEN2*^{N141I} vs. NDC with the CISBP motif database using the **RGT-HINT-ATAC-AD_N141I_NDC_differential_cisbp_full.sh** file
 - i. (base) [prvaldes@tscc-3-3 Scripts_2021]\$ bash RGT-HINT-ATAC-AD_N141I_NDC_differential_cisbp_full.sh