Binding site definition on IH-AB FLAG-PURA overexpression data set

Melina Klostermann

9/25/23

Table of contents

1 What was done in this report?

In this script I define PURA binding sites from the IH-AB FLAG-PURA iCLIP data set. The following steps were performed in this order:

1. Definition of Binding sites and filtering
    - PureCLIP peak calling (on merged sample1+2 sample3+4)
    - Global filter of pureclip sites (>5 cl events on pureclip site)
    - Local (genewise) filter of pureclip sites (> 75% percentile of gene)

- Binding site definition (5nt, Matrix approach, discard BS where center is not max pureclip, has less then 2 sites with crosslinks)
- Reproducibility filtering (soft boundary: 0.05% percentile of cl per binding site per sample, repro in min 3 of 4 samples)
- Comparison with old BS definition pipeline

2. Analysis of biniding behaviour

- Analysis of bound gene types
- Analysis of bound gene regions

2 Input

```
raw_path <- "/Users/melinaklostermann/Documents/projects/PURA/01_raw_data/oe-imb_koenig_2019

# crosslink sites in bw format
##########################################

#all 4 sample merged by Anke
bw_all_plus_path <- paste0(raw_path, "merged/bw/imb_koenig_2019_11_allsamples.v2uniqMD.duprm
bw_all_minus_path <- paste0(raw_path, "merged/bw/imb_koenig_2019_11_allsamples.v2uniqMD.dupr

# single samples
bw_1_plus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample1.v2uniqMD.duprm.plus.
bw_1_minus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample1.v2uniqMD.duprm.minu

```

```
bw_2_plus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample2.v2uniqMD.duprm.plus.
bw_2_minus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample2.v2uniqMD.duprm.minu

bw_3_plus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample3.v2uniqMD.duprm.plus.
bw_3_minus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample3.v2uniqMD.duprm.minu

bw_4_plus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample4.v2uniqMD.duprm.plus.
bw_4_minus_path <- paste0(raw_path,"bw/all/DR/imb_koenig_2019_11_sample4.v2uniqMD.duprm.minu

#
bw_all_samples_path <- "bw_4samples.RData"
bw_merges_path <- "bw_merges.RData"


# pureclip calls
###################################
#(obtained by running pureclip on pseudo samples 1u2 and 3u4 see below)
pureclip_path <- "/Users/melinaklostermann/Documents/projects/PURA/02_R_new_pip/01-BS_def/01


# gencode annotation v31
#################################
# this is the filtered annotation as used in molitor et al
annotation <- readRDS("/Users/melinaklostermann/Documents/projects/PURA/Molitor-et-al-2022/a

anno_txdb <- makeTxDbFromGRanges(annotation)

```