**Process to reproduce Fig. 1e for paper “Defining chromatin accessibility and nucleosome positioning in single cells with simultaneous measurements”**

**Binbin Lai**

Preparation: Download the scripts, executable file and annotation file:

seperate\_reads\_by\_length.py

get\_subgroup\_gene\_rpkm.py

extprog\_readcount\_on\_tss (Linux x86 version; otherwise, download source codes in src directory and type “make -f Makefile\_prog\_readcount\_on\_tss” to generate executive file on your machine)

draw\_profile.r

refFlat\_wf\_mm9.txt

#Step1: Download 48 bed (alignment) files for NIH3T3 single cell MNase fragments from GEO GSE96688

From:

GSM2538320\_GA5875\_3T3\_MNase\_0\_0\_mapq10\_noDup.bed.gz

To

GSM2538367\_GA9232\_3T3\_0\_0\_mapq10\_noDup.bed.gz

#Step2: Filter subnucleosomal fragments (length <80 bp) for all 48 cells:

e.g. for sample GA5875:

python seperate\_reads\_by\_length.py \ GSM2538320\_GA5875\_3T3\_MNase\_0\_0\_mapq10\_noDup.bed GA5875.les80.bed

#Step3: Pool all the subnucleosomal fragments:

cat \*les80.bed > combined.les80.bed

#Step4: Download NIH3T3 RNA-seq from published data: GEO [GSE39524](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39524) GSM970853. Align raw fastq to mm9 using Bowtie2. Calculate RPKM for each gene. Here we provide final output RPKM file:

NIH3T3\_rpkm\_refGene.f.txt

#Step5: Separate genes into 4 subgroups by RPKM values:

python get\_subgroup\_gene\_rpkm.py

This step output four files:

gene\_rpkm0-0.1.txt

gene\_rpkm0.1-1.txt

gene\_rpkm1-10.txt

gene\_rpkm10.txt

#Step6: Calculate average tag (subnucl.) density (RPKM) around TSS for each subgroup of genes derived from Step5:

Parameters:

-upstream 500 bp

-downstream 500 bp

-bin size 10 bp

-smoothed

-position shift: 0 bp (using center)

Command:

./extprog\_readcount\_on\_tss -U refFlat\_wf\_mm9.txt \ -b combined.les80.bed -u 500 -d 500 -w 10 -f 0 -t 1 -m 1 -g \ gene\_rpkm10.txt -p combined\_les80\_at\_rpkm10 -3 y

./extprog\_readcount\_on\_tss -U refFlat\_wf\_mm9.txt \ -b combined.les80.bed -u 500 -d 500 -w 10 -f 0 -t 1 -m 1 -g \ gene\_rpkm1-10.txt -p combined\_les80\_at\_rpkm1-10 -3 y

./extprog\_readcount\_on\_tss -U refFlat\_wf\_mm9.txt \ -b combined.les80.bed -u 500 -d 500 -w 10 -f 0 -t 1 -m 1 -g \ gene\_rpkm0.1-1.txt -p combined\_les80\_at\_rpkm0.1-1 -3 y

./extprog\_readcount\_on\_tss -U refFlat\_wf\_mm9.txt \ -b combined.les80.bed -u 500 -d 500 -w 10 -f 0 -t 1 -m 1 -g \ gene\_rpkm0-0.1.txt -p combined\_les80\_at\_rpkm0-0.1 -3 y

Output files:

combined\_les80\_at\_rpkm0-0.1.countaveprofile.txt

combined\_les80\_at\_rpkm0.1-1.countaveprofile.txt

combined\_les80\_at\_rpkm1-10.countaveprofile.txt

combined\_les80\_at\_rpkm10.countaveprofile.txt

# Step 7: Draw profiles using R:

Rscript draw\_profile.r \

combined\_les80\_at\_rpkm0-0.1.countaveprofile.txt \ combined\_les80\_at\_rpkm0.1-1.countaveprofile.txt \ combined\_les80\_at\_rpkm1-10.countaveprofile.txt \ combined\_les80\_at\_rpkm10.countaveprofile.txt \ combined\_les80\_at\_4groupgenes.png

The output figure combined\_les80\_at\_4groupgenes.png is the same as Fig. 1d. in the paper.