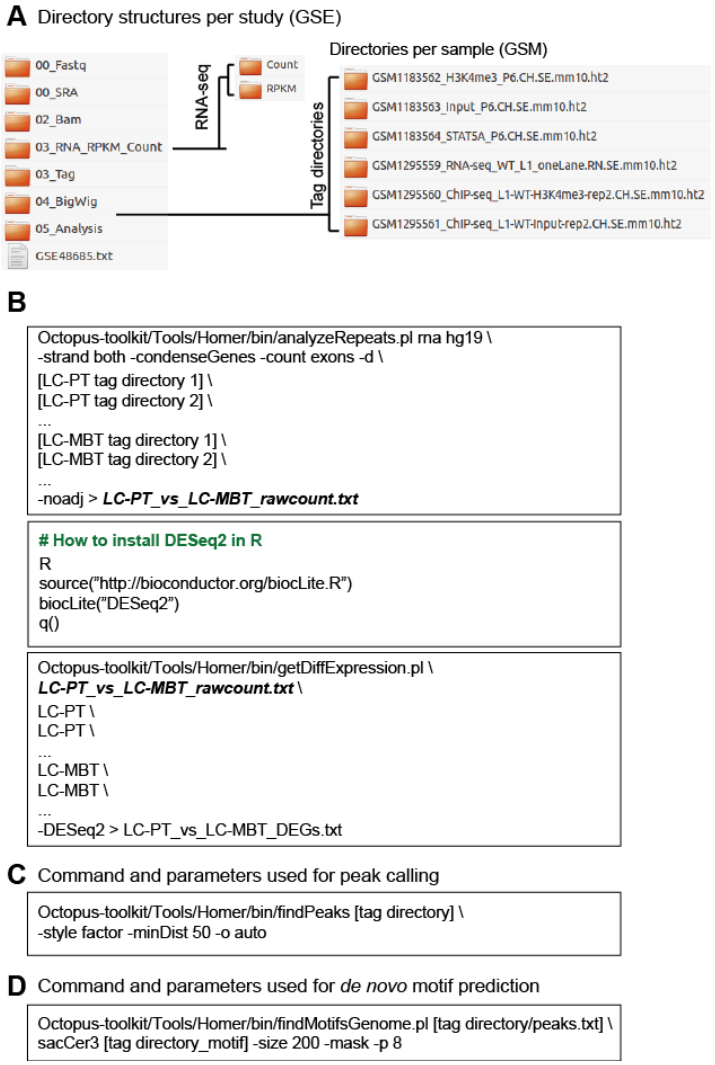
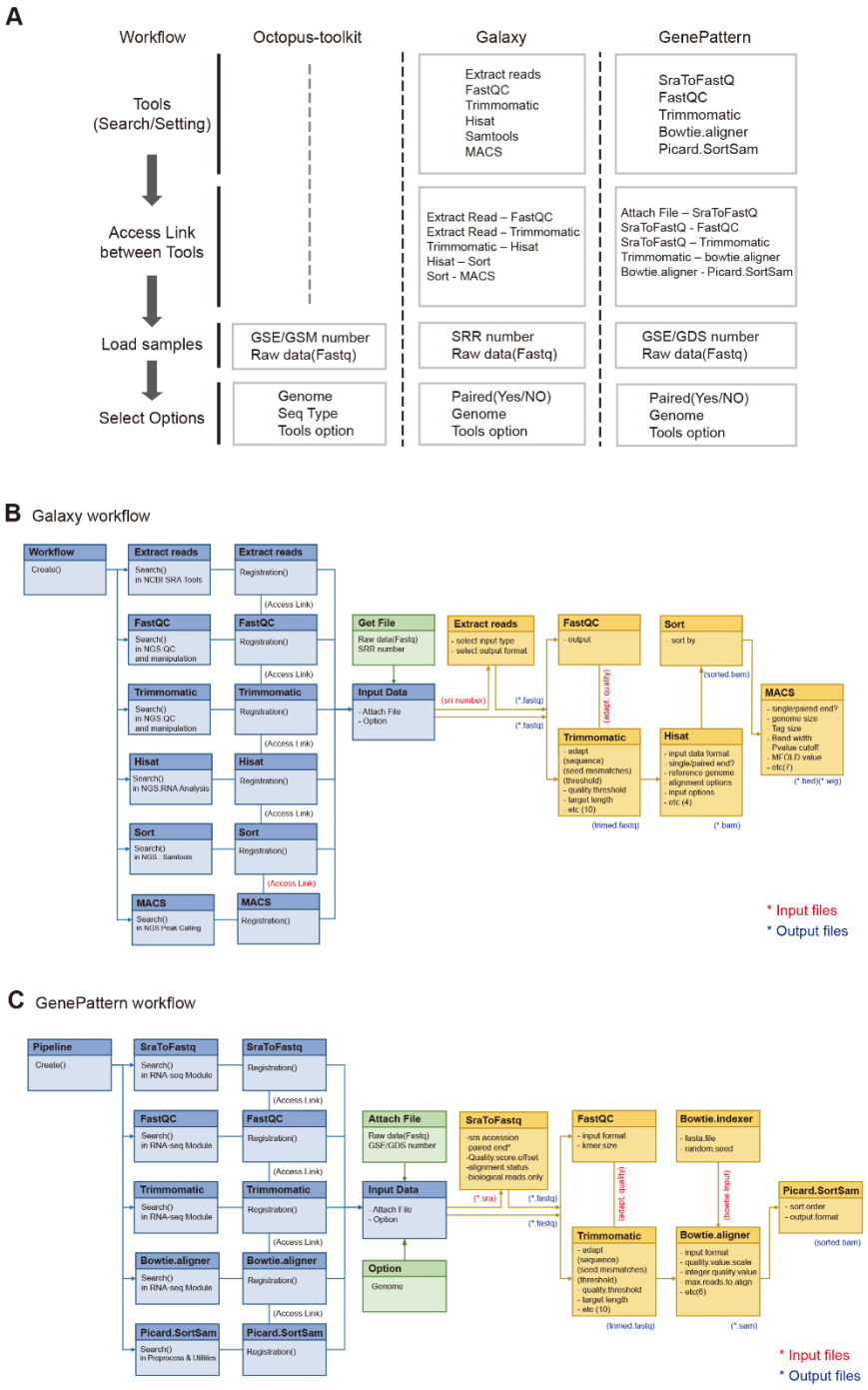


Supplementary Figure 1. Octopus-Toolkit output and commands used in this study.

(A) Octopus-Toolkit generates several directories and stores processed files in different directories. For each study (GSE accession number), SRA files are downloaded into the 00_SRA directory. After the file conversion from SRA files to FASTQ files, the converted files are stored in the 00_Fastq directory. FastQC output is stored in the 01_Fastqc directory. BAM files and BigWig files (for visualization) are placed in the 02_Bam and 04_BigWig directories, respectively. The tag directories generated by HOMER are stored in the 03_Tag directory. For RNA-seq data, the number of mapped reads and normalized expression values (RPKM) of genes can be checked with text files in the 03_RNA_RPKM_Count directory. (B) Commands used to identify differentially expressed genes using DESeq2. (C) Command used to identify peaks for ChIP-seq data. (D) Command used to detect motif sequences in the identified peaks.



Supplementary Figure 2. Workflow for analyzing epigenomic and transcriptomic NGS data in Octopus-toolkit, Galaxy, and GenePattern.

(A) A detailed workflow of epigenomic and transcriptomic NGS data analysis. (B, C) Workflow of steps needed for analyzing epigenomic and transcriptomic NGS data using (B) Galaxy and (C) GenePattern. File format in red and blue indicate input and output files for each step, respectively.

Remove the Batch Effect using EdgeR package in R

```
# EdgeR

# Install : packages
source("http://bioconductor.org/biocLite.R")
biocLite("edgeR")

# Load : packages
library("edgeR")

# Load : Read count file
ds = read.table("Pten-Hetp_ReadCount.txt",header = T,row.names=1)
matrix_ds = ds
group = c(0,0,1,1)
libSizes = as.vector((colSums(ds)))

# Remove Batch Effect

# removeBatchEffect
matrix_ds = removeBatchEffect(matrix_ds, group)

# remove minus(-) value
ifa = matrix_ds > 0
rowNum = c()
for(i in 1:nrow(ifa)){
  rowNum = c(rowNum,all(ifa[i,]))
}
remove_matrix_ds = matrix_ds[rowNum,]

libSizes = as.vector((colSums(remove_matrix_ds)))

# Using EdgeR
d = DGEList(counts = remove_matrix_ds, group = group,lib.size = libSizes)
d <- calcNormFactors(d)
d <- estimateCommonDisp(d)
d <- estimateTagwiseDisp(d)
de.com <- exactTest(d)

# Save : Result of EdgeR
results <- topTags(de.com,n = length(ds[,1]))
write.table(as.matrix(results$table),file="Result_removeBatchEffect_EdgeR.txt",sep="\t",quote = F)
```

***Pten-Hetp_ReadCount.txt (Input file)**
Input file is generated using data in Octopus-toolkit/Result/GSExxx/03_RNA_RPKM_Count/Count/Total_mm10.txt (https://github.com/kangk1204/Octopus-toolkit2/blob/master/doc/_templates/Pten-Hetp_ReadCount.txt)

Supplementary Figure 3. EdgeR command to remove batch effect using Octopus-toolkit output in R.

Command used to remove batch effect using EdgeR in R. The input file, highlighted in red, was generated using the output file from Octopus-toolkit in the 03_RNA_RPKM/Count directory. The format of input file consists of gene symbol and read count of each sample.