Software\_requirement

-R (<https://www.r-project.org/>) V3.3.1

-library GenomicRanges 1.26.4 (<https://bioconductor.org/>)

Limma 3.30.13 (<https://bioconductor.org/>)

edgeR 3.16.5 (<https://bioconductor.org/>)

DESeq2 1.14.1 (<https://bioconductor.org/>)

The «Code.r» script requires 5 input files to run. As an example, data obtained on chromosome 22 for each file is provided.

This code is designed

-to identify genes under the control of CGI/promoter

-to integrate differential analysis (ctrl vs tumor) for methylation (at CpG islands) and expression data

-to calculate the correlation between CNV and expression data.

The output file «result.txt» summarizes these informations per gene as well as the mean methylation of the CGI/promoter and the expression level per gene and sample.

Input files:

1- «gencode.v19.annotation.txt» is used to know the gene position and is derived from the gtf file obtained on the GENCODE web site (<https://www.gencodegenes.org/human/release_19.html>)

2- «hg19\_cpg\_island.txt» contains the CpG islands position on hg19 genome and was obtained from UCSC (<http://genome.ucsc.edu/cgi-bin/hgTables>)

3- «450k\_methylation.txt», obtained with the GenomeStudio software from Illumina, contains the filtered βvalues for each sample (ctrl & tumor).

4- «RNAseq.txt», obtained with htseq-count script (<https://htseq.readthedocs.io>), contains the count for each gene and for each sample (ctrl & tumor)

5- «CNV.txt», obtained with Chromosome Analysis Suite (ChAS) from ThermoFisher, contains the CNV fragments for each samples.