**R tutorial for:**

**Detecting Differentially Methylated Loci Using Cosine Similarity Model Base on Case-Control Designs**

Input: A file with DNA methylation *β*-values

Output: A file with calculated scores and p-values of all loci, etc

Required R package: None

**Instructions**

1. Download the following sample dataset and R code document:

* Sample\_data.txt
* csDML\_RCode.docx

1. Save these files into your working directory.
2. Run the R code in the document and find “csDML\_Output.csv” in your working directory.

**Sample dataset**

The sample data analyzed in this tutorial is a subset of a 450K DNA methylation dataset of tumor and adjacent normal tissues of breast invasive carcinoma (BRCA) from The Cancer Genome Atlas (TCGA) which contains

1. 64 tumor and normal-adjacent pairs.
2. first 100 CpG sites on chromosome 1.
3. data has been preprocessed, all CpGs with null beta values has been deleted.

**Explanation of the output**

The output file has following information:

* array.ILMNID = the probe ID
* array.CHR = chromosome where a CpG site locates
* distance\_x = the average methylation level of a CpG site in the controls
* distance\_y = the average methylation level of a CpG site in the cases
* skewness = the skewness estimate of a CpG site
* score = defined score of a CpG site
* PValue = p-value of a CpG site