**R tutorial for:**

**Detection of differentially methylated regions with mean and variance combined signals**

**Ya Wang (yw2453@columbia.edu)**

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

Output: A file with identified DMRs and p-values, etc

Required R package: bumphunter

**Instructions**

1. Download the following sample dataset and R script:

* Sample\_data.RData
* R\_code.R

1. Save these files into your working directory.
2. Run the R script and find “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. 30 tumor and normal-adjacent pairs.
2. first 1000 CpG sites on chromosome 6.
3. data has been normalized, corrected for type II bias and ordered by their genomic location.

**Explanation of the output**

The R function “regionFinder” from the bumphunter package (Jaffe, et al., 2012) is used to identify candidate DMRs and generate the output table.

The output file has following information:

* chr = chromosome where a candidate DMR locates
* start = the genomic location of the first CpG site in a candidate DMR
* end = the genomic location of the last CpG site in a candidate DMR
* area = the strength of evidence of a candidate DMR
* cluster = the cluster index of a candidate DMR
* L = number of CpG sites in a candidate DMR
* P.valueArea = raw p-value of a candidate DMR without multiple comparison adjustment
* fwerArea = family-wise error rate of a candidate DMR

**Reference**

Jaffe A.E.*, et al.* Bump hunting to identifying differentially methylated regions in epigenetic epidemiology studies. International Journal of Epidemiology. 2012;41(1):200-209.

Wang Y, Teschendorff AE, Widschwendter M, Wang S. Detection of differentially methylated regions with mean and variance combined signals. 2016; submitted.