Result comparison of different method: MACS, MEDIPS,BALM

Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus

1, MACS Result stablilty

By default, robust linear regression [3, 4] is applied to normalize data with different sequencing depth since we assume that all experiments were performed under the same controllable conditions. A normalization factor is calculated by performing robust linear regression between tags count of 10kb bins of the two samples. Then the samples are adjusted to have comparable enrichment levels using the normalization factor.

***A). Simulated peaks (binding sites)***

(a) Randomly generate *n* binding motifs (e.g. 10 nt in length) on the genome; (b) For each motif, generate a certain number of artificial fragments according to the ChIP sample; (c) Randomly define each fragment’s length, varying from 100-300 nt; (d) Randomize each fragment’s position around the motif; (e) One end of each fragment was randomly selected and its coordinates was recorded as one tag.

Omics association study mediated by epigenetic emissary CpG-SNP

CpGI在整个human基因组0.696%的中(21,842,742, 3,137,144,693, GRCh37 )。

CpGI shore

根据其定义为CpGI上下游各2k的区域，但实际上95.6%(27431, 28691,UCSC,2012)的CpGI长度小于2K，为此我重新定义了CpGI上下游各CpGI的区域为CpGI shore区域。在这种情况下CpGI shore从长度上和CpGI具有相同的分布。 另外除去和CpGI有overlap的CpGI shore区域，adaptive CpGI shore(A-CpGI shore) 共有54937个区域（2445个被删除，注意如果采用2k作为阈值，会有6824个区域被删除）。