Brief Description of the "TF Enrichment" Method

1. We downloaded the "wild type" MSCs DNase-seq from ENCODE and treated/aligned to hg19.
2. We called peaks (open chromatin) on the MSCs using MACS.
3. The sequence from the peaks called on the DNase-seq were used as input to the deep network. The network will compute "DNA motifs".
4. We associated the DNA motifs to TFs using TOMTOM. All the "*de novo*" motifs, i.e. motifs that were not assigned to any known TF were removed from the analysis since our method is not ready with those yet.
5. We converted the motifs obtained using the Deep Net to regular Position Weight Matrices (PWMs), which are the structures we use to "find motifs".
6. We applied the PWMs in the following set of regions:
    a. Regulatory regions of upregulated genes.
    b. Regulatory regions of downregulated genes.
    c. Regulatory regions of the genes associated to hypermethylated sites.
    d. Regulatory regions of the genes associated to hypomethylated sites.
    e. Hypermethylated CpGs.
    f. Hypomethylated CpGs.
7. We performed the enrichment test by testing each of the aforementioned regions against a random background composed by random genomic regions with same length of the regions and 100 times more regions (for statistical significance) using Fisher's Exact Test.
8. The heatmaps were generated using the p-values from the enrichment test. Heatmaps were generated using regions in pairs: (a,b), (c,d) and (e,f).

Obs1: I can generate the nice graphs with the motif logos as soon as I talk to Danilo or Wagner and the Figures are established.

Obs2: Matt can generate "activation distribution plots", which are important on the field, which you can put in the supplementary material.