

List of Figures

1	Mean and standard deviation of the raw methylation data	2
2	Eigengene methylation profile of the modules	2

1 Preprocessing

1.1 Gene expression array data

Gene expression array data for 344 AML patients was downloaded from GEO (GSE 14468). Normalization was performed in [?]. Each probe was annotated to the gene with nearest transcription start site and the corresponding gene symbol was provided in the file. The gene symbols of the probes are used to query the Gene Ontology [?] for biological process annotation. We are interested in genes related to methylation and transcriptional regulation. To remove irrelevant genes, we filter for genes that have keyword “methylation” and “transcript” in their biological process annotation. In the end, 3417 out of 17788 remained for the downstream analysis.

1.2 Methylation array data

Methylation array data for 344 AML patients was downloaded from GEO (GSE 18700). Each probe was annotated to the nearest transcription start site allowing for a maximum distance of 5 kb from the TSS. Any probes that were mapped to a gene are filtered for downstream analysis and in total 22759 out of 25626 probes are left after the filtering step. The The distribution of the mean and standard deviation are shown below in Figure 1.

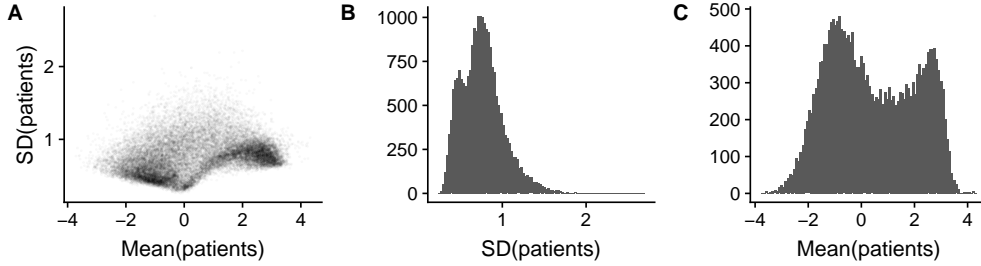


Figure 1:

Successive normalization were performed on the methylation array [?] until the absolute difference between Frobenius norm of consecutive iterations is below 10^{-8} . PhenoGraph were applied on the filtered on the normalized dataset [?] which constructs a nearest-neighbor graph ($k = 30$) with weight between each probe pair as Jaccard similarity coefficients between the corresponding nearest neighbor set and performs Louvain community detection algorithm by maximizing the modularity on the graph. 20 communities (modules) were obtained as a result. The “eigengene” profile was computed for each module as their first principal component, shown in the Figure 2.

Figure 2: Eigengene methylation profile of the modules. Each black line represents a methylation level of a probe across patients. The blue line shows the eigengene expression profile.