

## List of Figures

1	Mean and standard deviation of the raw methylation data . . . . .	2
---	---	---

# 1 Preprocessing

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

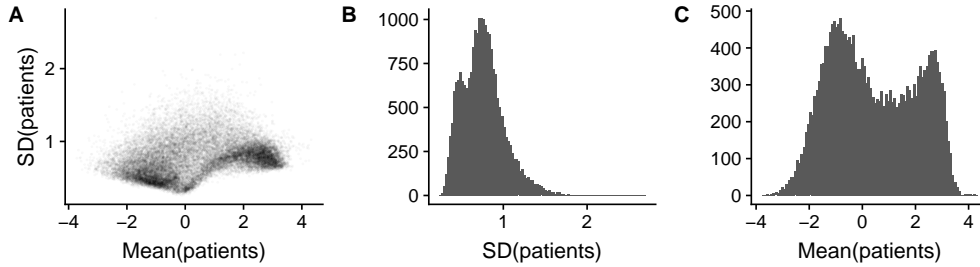


Figure 1:

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 ??.

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.