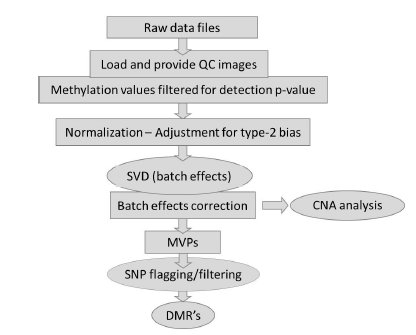
**Genome-wide DNA methylation profile for ESCA based on MH450K beadchip**

**Method**

****

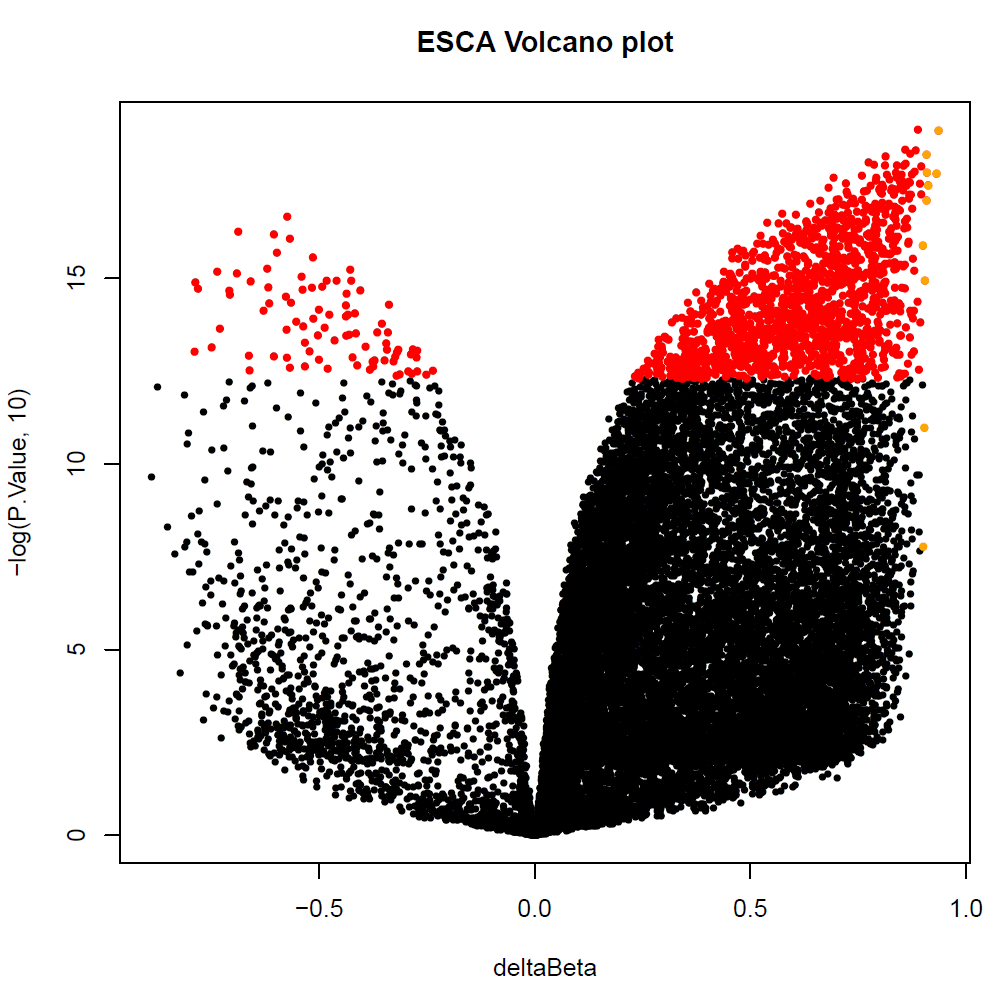
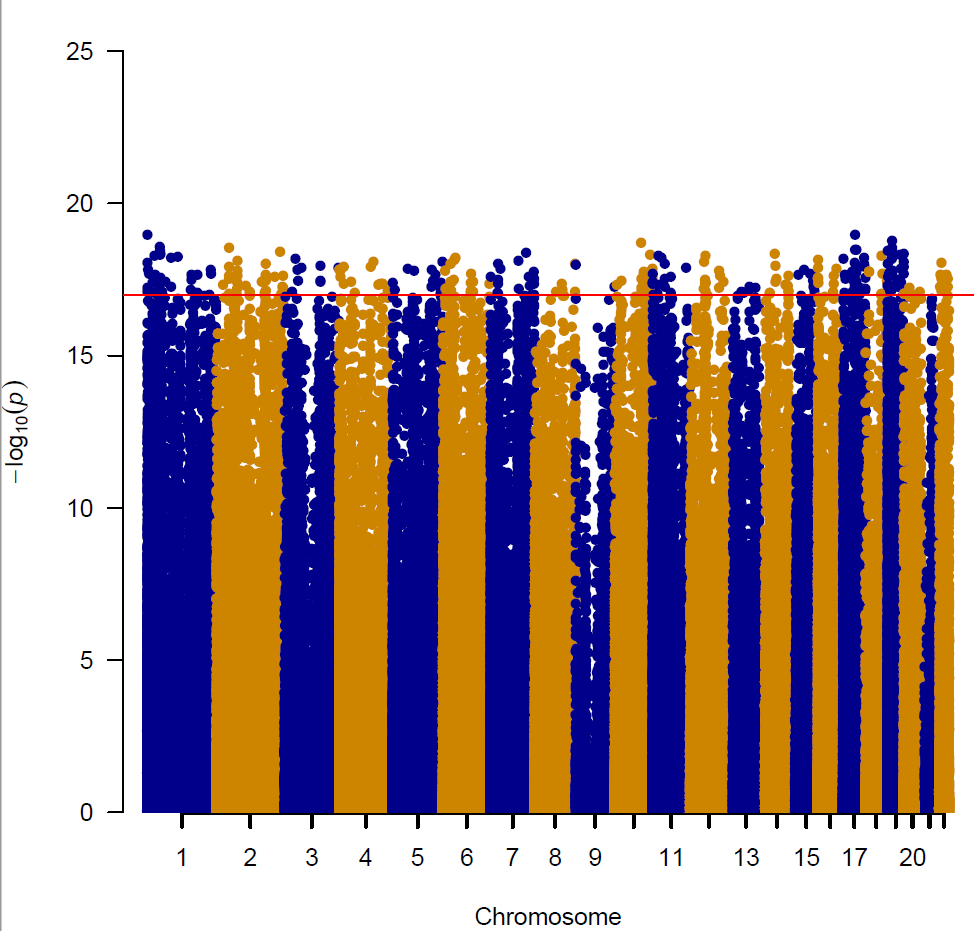
Methylation profile of 10 ESCA sample and 1 normal sample were measured with Methylation 450K Beadchip. Microarray analysis was followed by ChAMP with empirical threshold and setting. Filtering probes with a detection p-value above 0.01 in one or more samples has removed 1800 probes from the analysis. Filtering probes with a beadcount <3 in at least 5% of samples, has removed 2542 from the analysis. Filtering probes with SNPs as identified in Nordlund et al, has removed 28753 from the analysis. Filtering probes that align to multiple locations as identified in Nordlund et al, has removed 8478 from the analysis. Filtering probes on the X or Y chromosome has removed 11147 from the analysis. The analysis will proceed with 432792 probes and 12 samples. Batch effect correction was applied with Comat algorithm and normalizing data with BMIQ. High Methylation Variable site (MVS), differential methylation site (DMS) and differential methylation region (DMR) were reported and conducted further statistic analysis. [Principal component analysis](https://en.wikipedia.org/wiki/Principal_component_analysis)(PCA) and cluster analysis based on genome-wide methylation data were applied to measure the similarity of the samples. Heatmap based on differential DMS were applied to show the magnitude of the difference between cancer and normal samples. volcano plot was applied to show the difference of methylation and P-value distribution. manhattan plot was applied to show the DMRs distribution among whole human genome.

**Result**

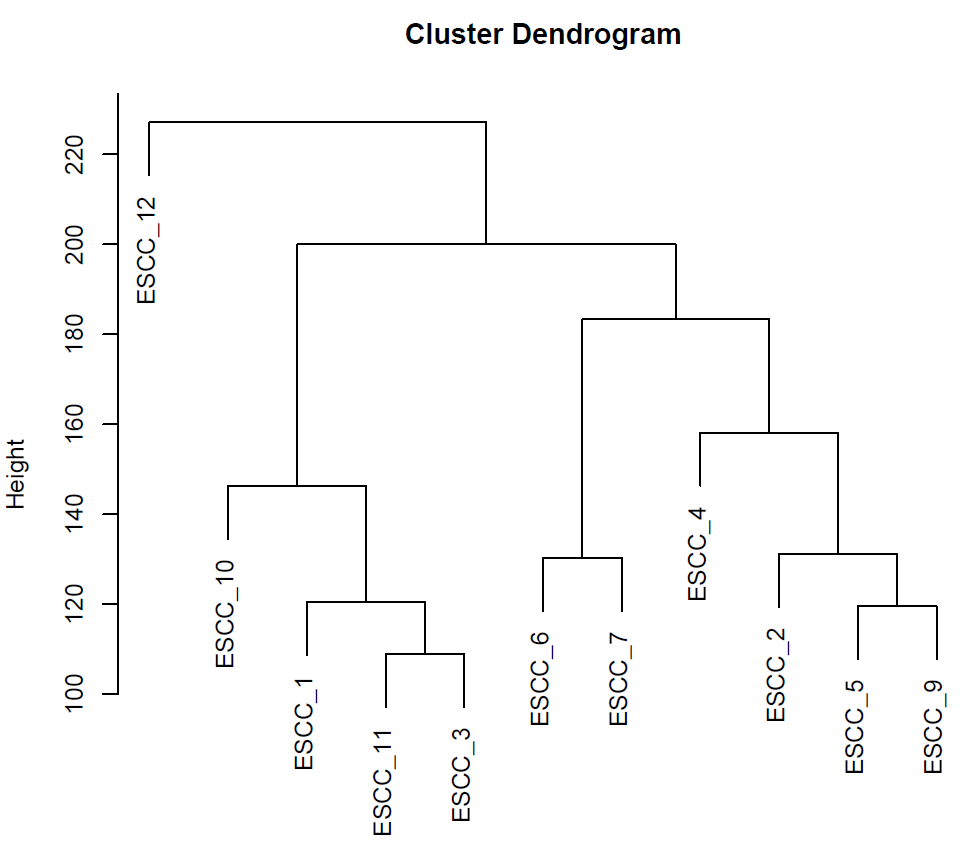
95830 significant MVS, 3509 significant DMRs (including 19504 CpGs) were identified with a BH multiple test correction of adjusted P-value <0.05 (Figure 1A). Manhattan and Volcano analysis showed the methylation aberrance were dispersing among the whole genome without any chromosome unbalance or preference (Figure 1A, 1B, 1F). PCA and cluster analysis showed significant difference between cancer and normal group and genome-wide methylation could powerfully represent the similarly of the samples (Figure 1C and 1D). Heatmap analysis showed DMS could distinguish cancer samples from normal samples (Figure 1E). Large number of DMRs were distributed in promoter, enhancer, TSS nearby regions (Figure.1G),

Figure 1. Bioinformatics analysis to MH450K in ESCA dataset.

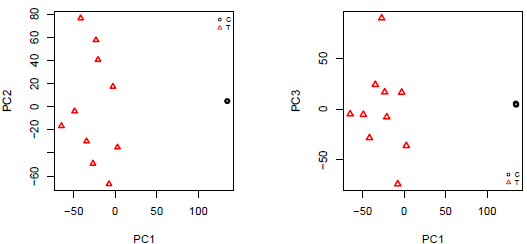
A. Manhattan plot B. Volcano plot



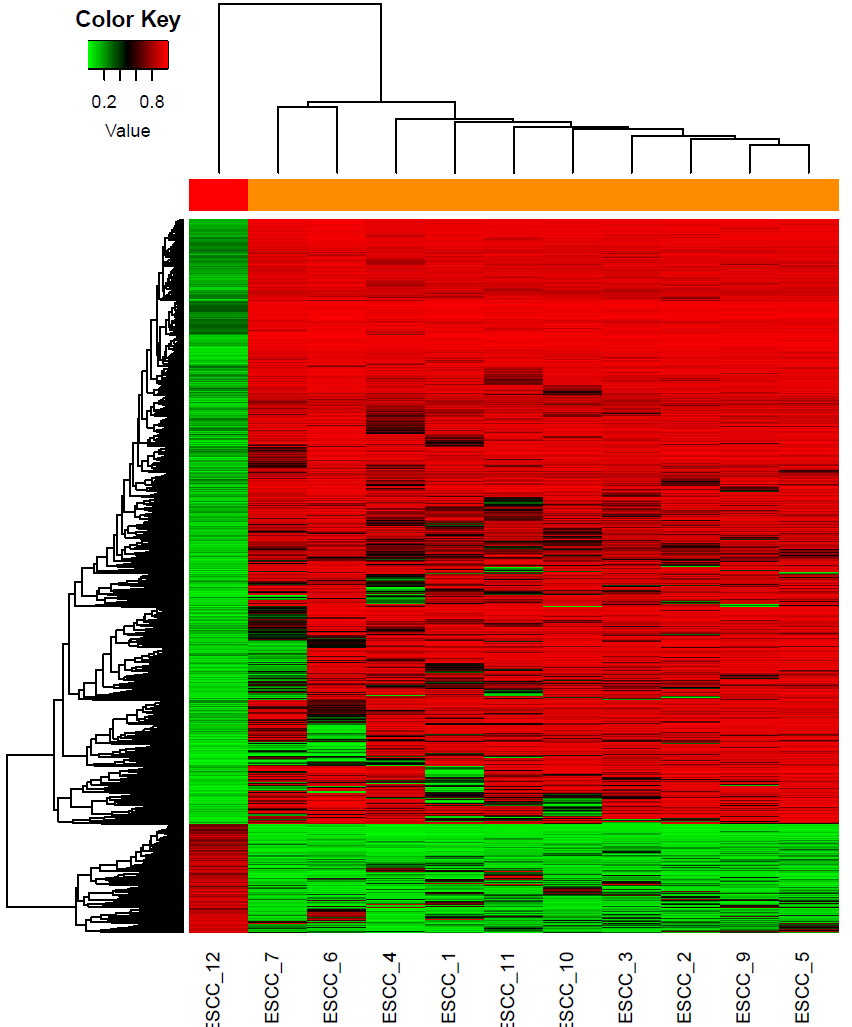
C. Cluster tree plot



D. PCA plot



E. Heatmap plot



F. DMS distribution in Genomic G. DMS in Genomic elements

