Identification of Non-Invasive Tumor-specific Biomarkers for Hepatocellular Carcinoma through Multi-transcriptomics Data Integration

Zahra Golshan[1#], Hanieh Fadakar[2#], Saeedeh Salehi[3], Yasaman Najafi[4+], Sanaz Akbari[5+], Maryam Momeni[6], Alieh Gholaminejad[7]*

[1] Department of Applied Mathematics, School of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran.

[2] Department of Medical Laboratory Sciences, School of Paramedical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran.

[3] Department of Immunology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran.

[4] Department of Biology, School of Sciences, Razi University, Kermanshah, Iran

[5] Department of Microbiology, School of Sciences, Karaj Islamic Azad University, Alborz, Iran.

[6] Department of Biotechnology, School of Biological Science and Technology, University of Isfahan, Isfahan, Iran.

[7] Regenerative Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran.

[#] These authors contributed equally to this work and share first authorship.

[+] These authors contributed equally to this work.

Emails:

z.golshan@modares.ac.ir, 0009-0000-3186-7414

haniyehfadakar46@gmail.com, 0009-0006-0560-0154

Salehi_saeedeh@yahoo.com, 0000-0003-1767-9315

Yasamannajafi7676@gmail.com, 0009-0008- 4246-0773

Sanazakbari022@gmail.com , 0000-0001-9155-8728

Maryammomeni23@yahoo.com, 0000-0003-3294-3516

a.gholaminejad@res.mui.ac.ir, Al_gholami@yahoo.com, 0000-0003-0329-2937

Correspondence:

Alieh Gholaminejad, Ph.D.

Regenerative Medicine research center, Isfahan University of Medical Sciences, Hezar Jerib Avenue, Isfahan, 81746-73461, Iran.

Tel/Fax: +98-3136687087
Email: a.gholaminejad@res.mui.ac.ir, Al_gholami@yahoo.com

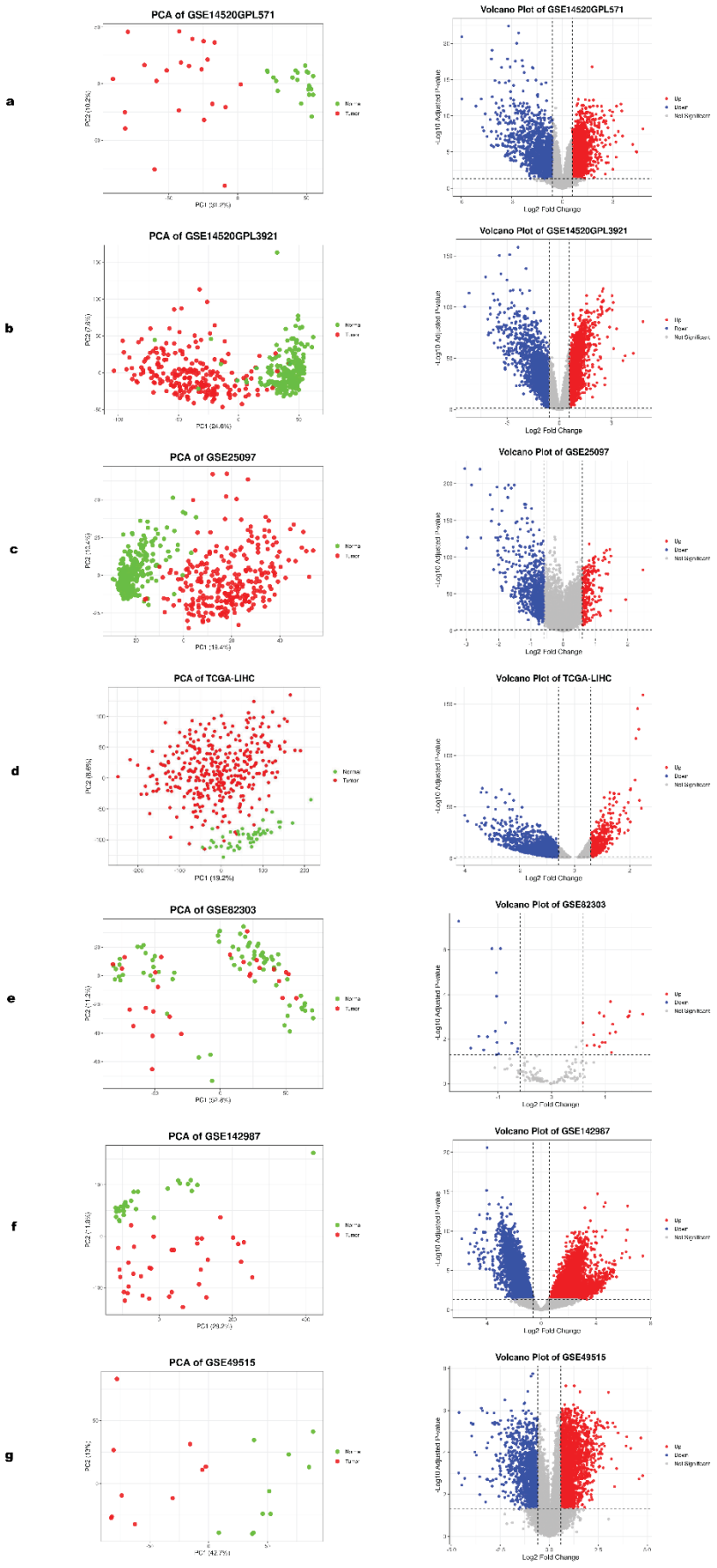

a
PCA of GSE14520GPL571
Volcano Plot of GSE14520GPL571
b
PCA of GSE14520GPL3921
Volcano Plot of GSE14520GPL3921
c
PCA of GSE25097
Volcano Plot of GSE25097
d
PCA of TCGA-LIHC
Volcano Plot of TCGA-LIHC
e
PCA of GSE82303
Volcano Plot of GSE82303
f
PCA of GSE142987
Volcano Plot of GSE142987
g
PCA of GSE49515
Volcano Plot of GSE49515
Normal
Tumor
Up
Down
Not Significant
-Log10 Adjusted P-value
Log2 Fold Change

Fig 1. Principal Component Analysis (PCA) and volcano plots of gene expression profiles across seven datasets. Panels **a-g** represent the datasets GSE14520-GPL571, GSE14520-GPL3921, GSE25097, TCGA-LIHC, GSE82303 (exosomal miRNAs), GSE142987 (plasma), and GSE49515 (PBMC) respectively. Each panel contains two plots: a PCA plot (left) and a volcano plot (right). The PCA plots depict the distribution of tumor (red) and normal (green) samples based on global expression patterns, highlighting the overall variance and clustering between groups. The volcano plots illustrate the results of differential expression analysis, with the x-axis representing the $\log_2$ fold change and the y-axis showing the $-\log_{10}$ adjusted p-value. Genes with $|\log_2 FC| > 0.585$ and adjusted p-value < 0.05 were considered significantly differentially expressed. Upregulated genes are shown in red, downregulated genes in blue, and non-significant genes in grey. These visualizations collectively demonstrate the extent of transcriptomic alterations between tumor and normal samples across multiple platforms and sample types

Fig 2 : GSE14520-GPL3921

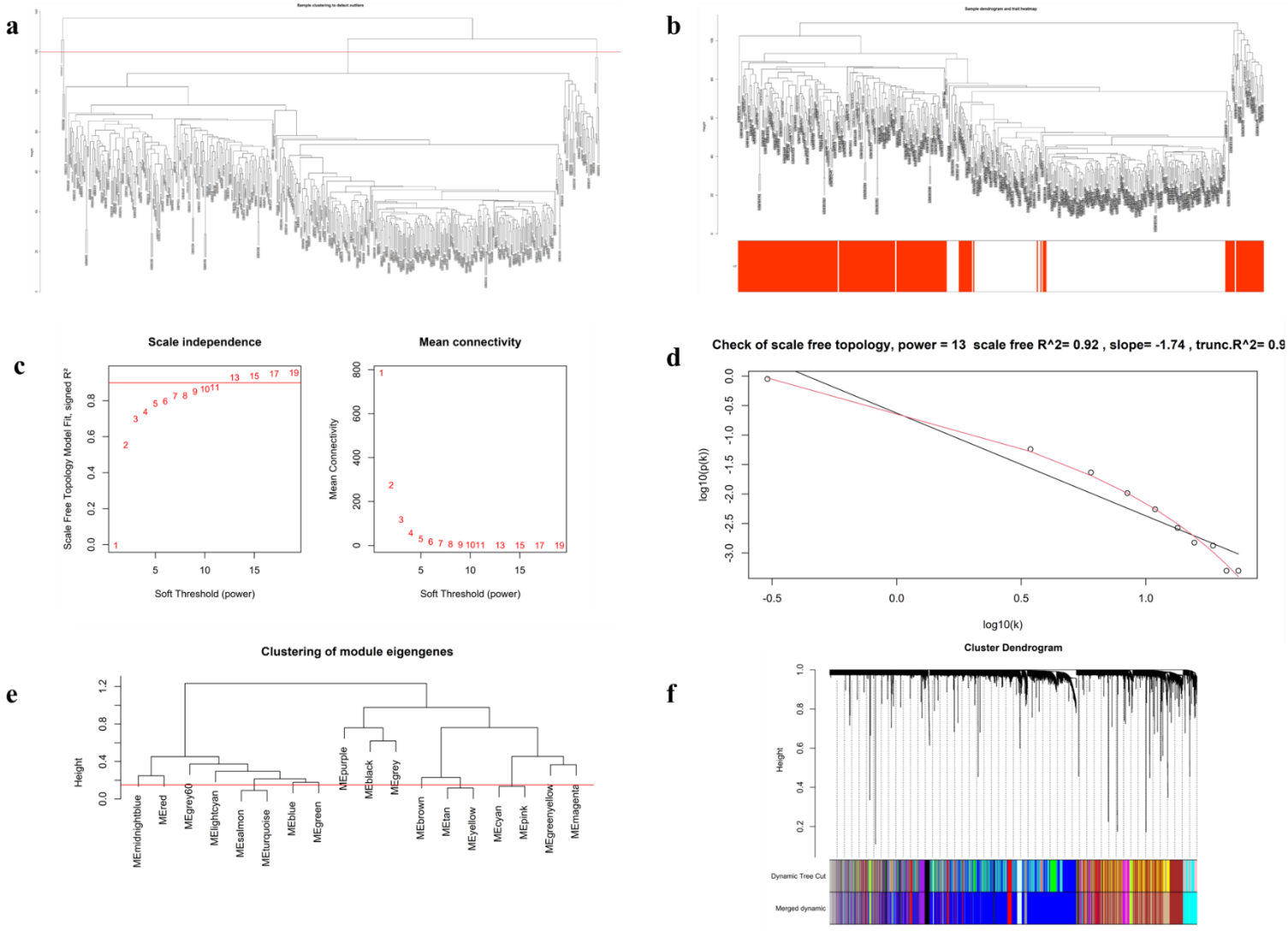


Fig 3 : GSE14520-GPL571

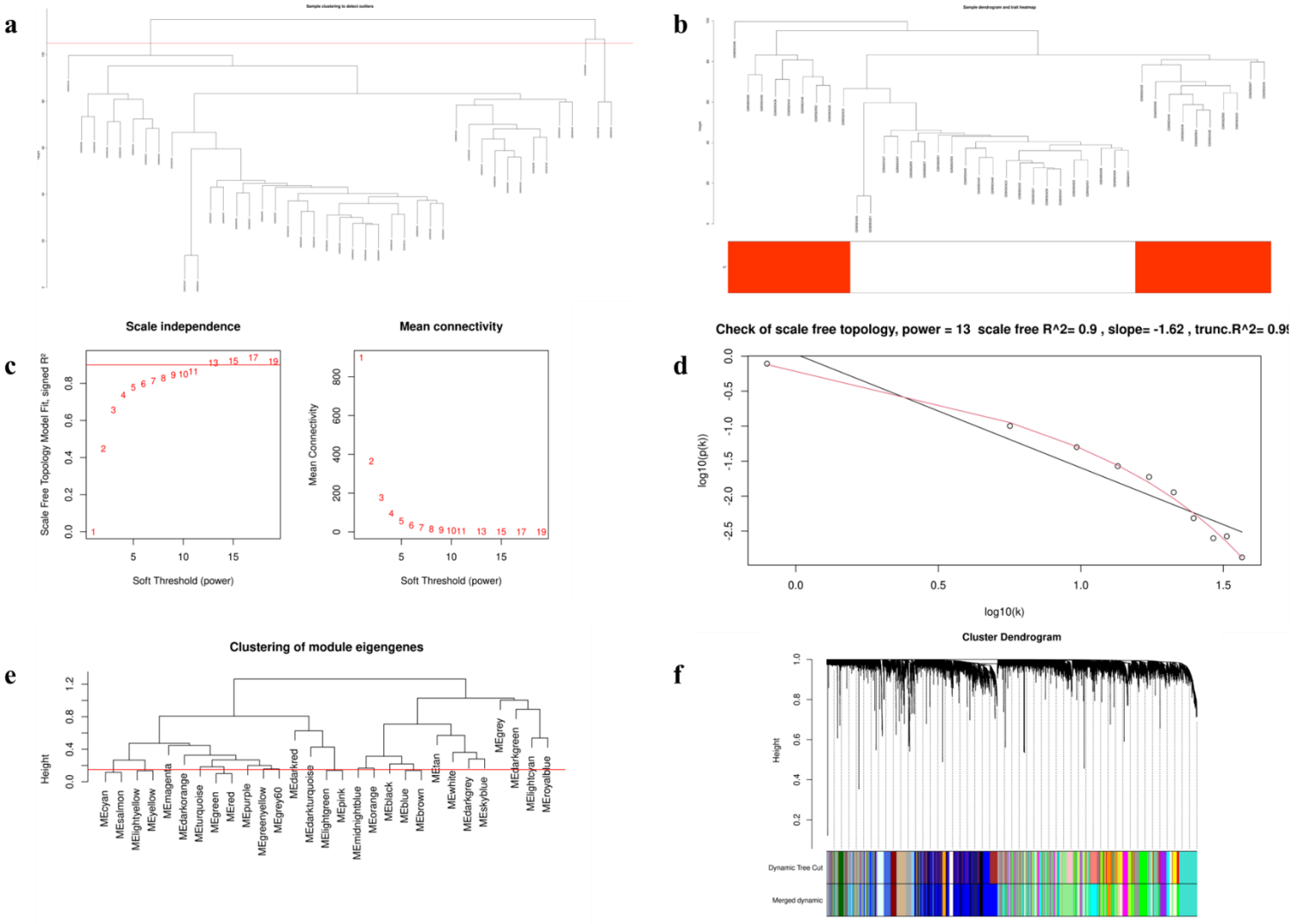


Fig 4 : GSE25097

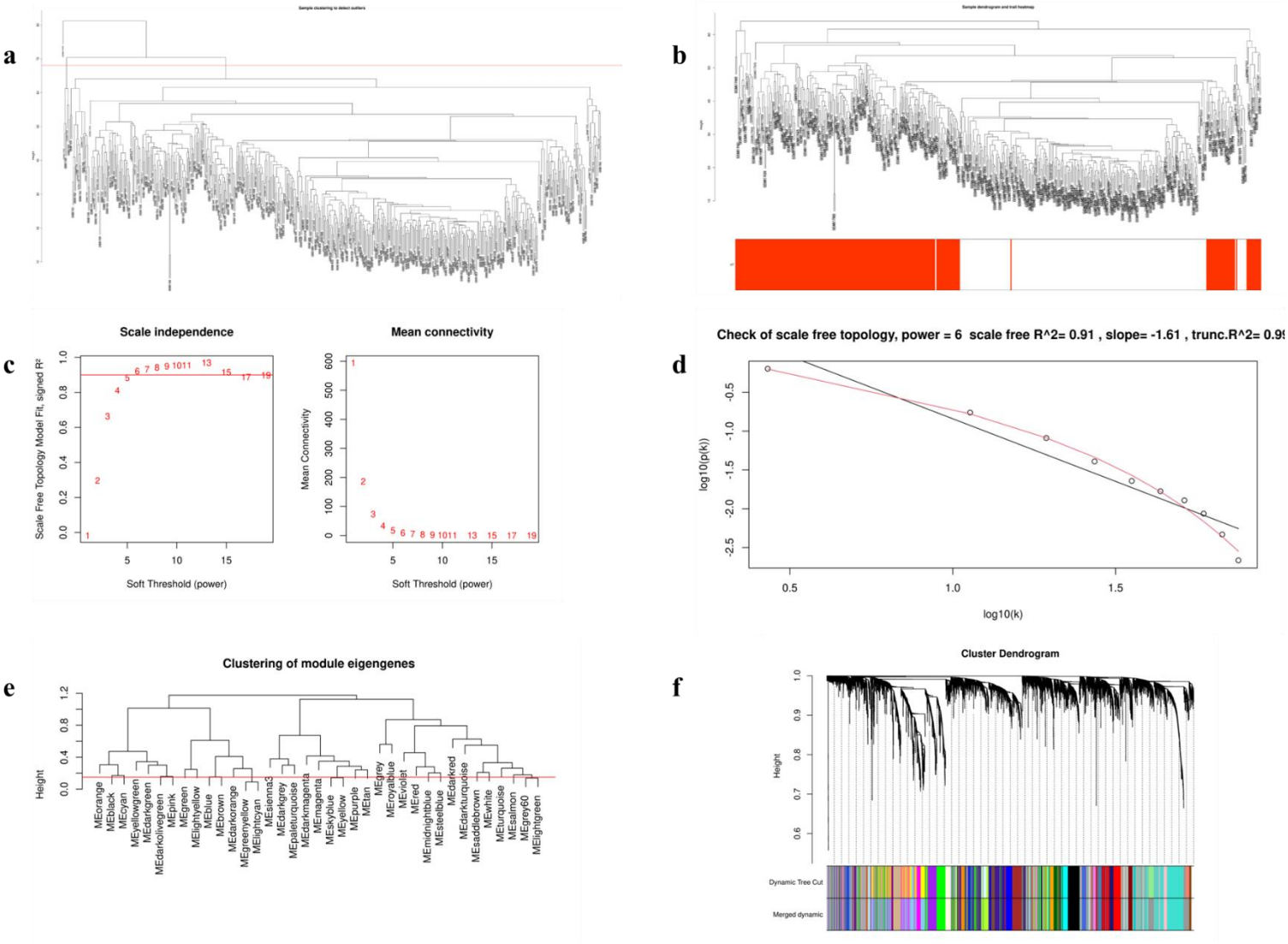


Fig 5 : LIHC-TCGA

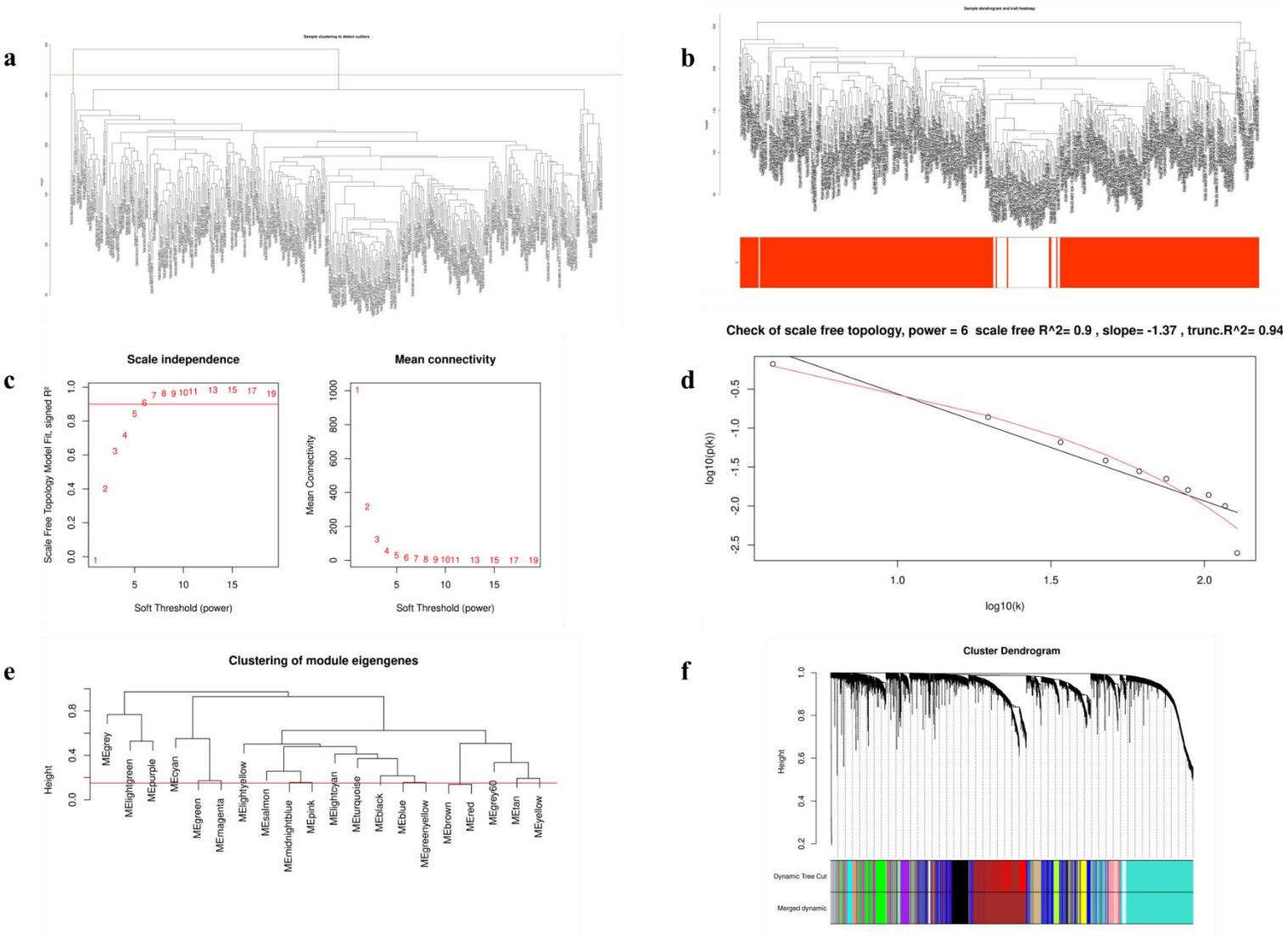


Figure 2-5: a. Sample Clustering: Hierarchical clustering of samples based on gene expression, with outliers identified and excluded using a red threshold line. **b. Sample Dendrogram and Trait Heatmap:** Sample dendrogram grouped by gene expression similarities, with a heatmap showing the correlation between clinical traits and sample clusters. **c. Soft-Thresholding:** Analysis of the scale-free topology fit index for various power values (β) to determine the optimal threshold for network construction. **d. Scale-Free Topology:** Validation of the chosen soft-thresholding power by evaluating the network's scale-free topology through mean connectivity at different thresholds. **e. Clustering of Module Eigengenes:** Clustering of module eigengenes to visualize relationships between gene modules with similar expression patterns. **f. Gene Dendrogram:** Gene dendrogram with color-coded modules identified by dynamic tree cut, reflecting co-expressed gene sets and illustrating the hierarchical clustering of genes based on topological overlap, identifying gene clusters with similar expression.