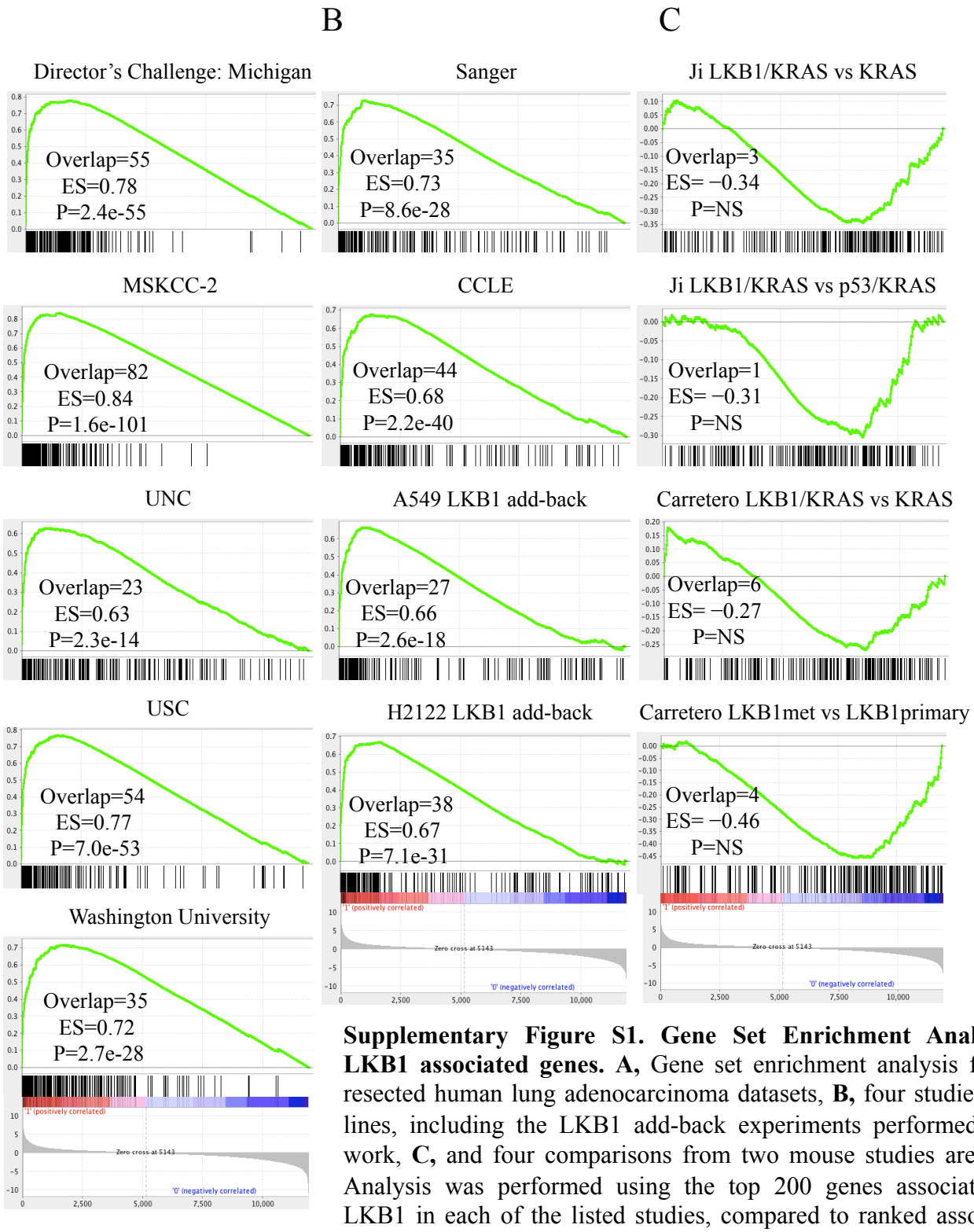
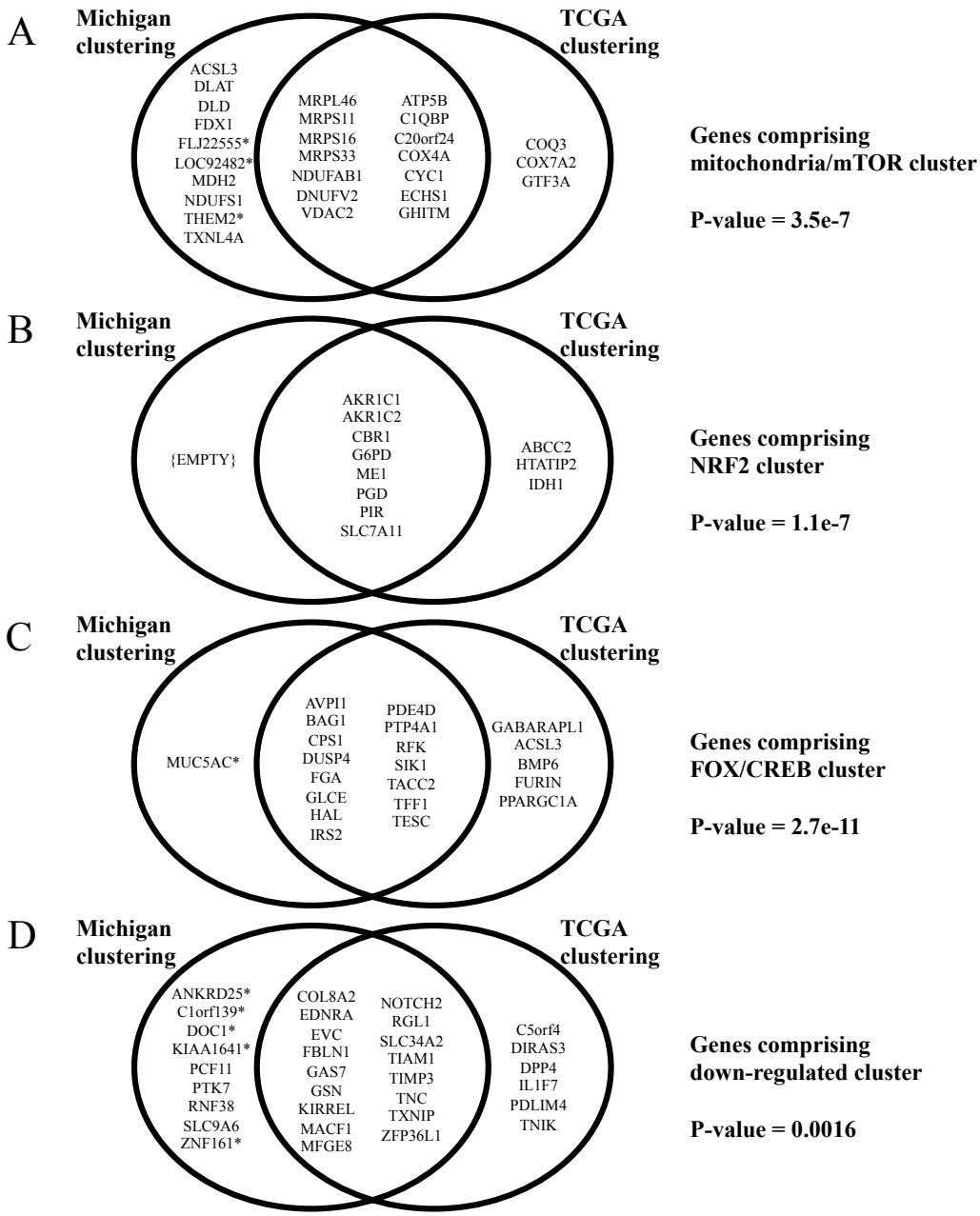
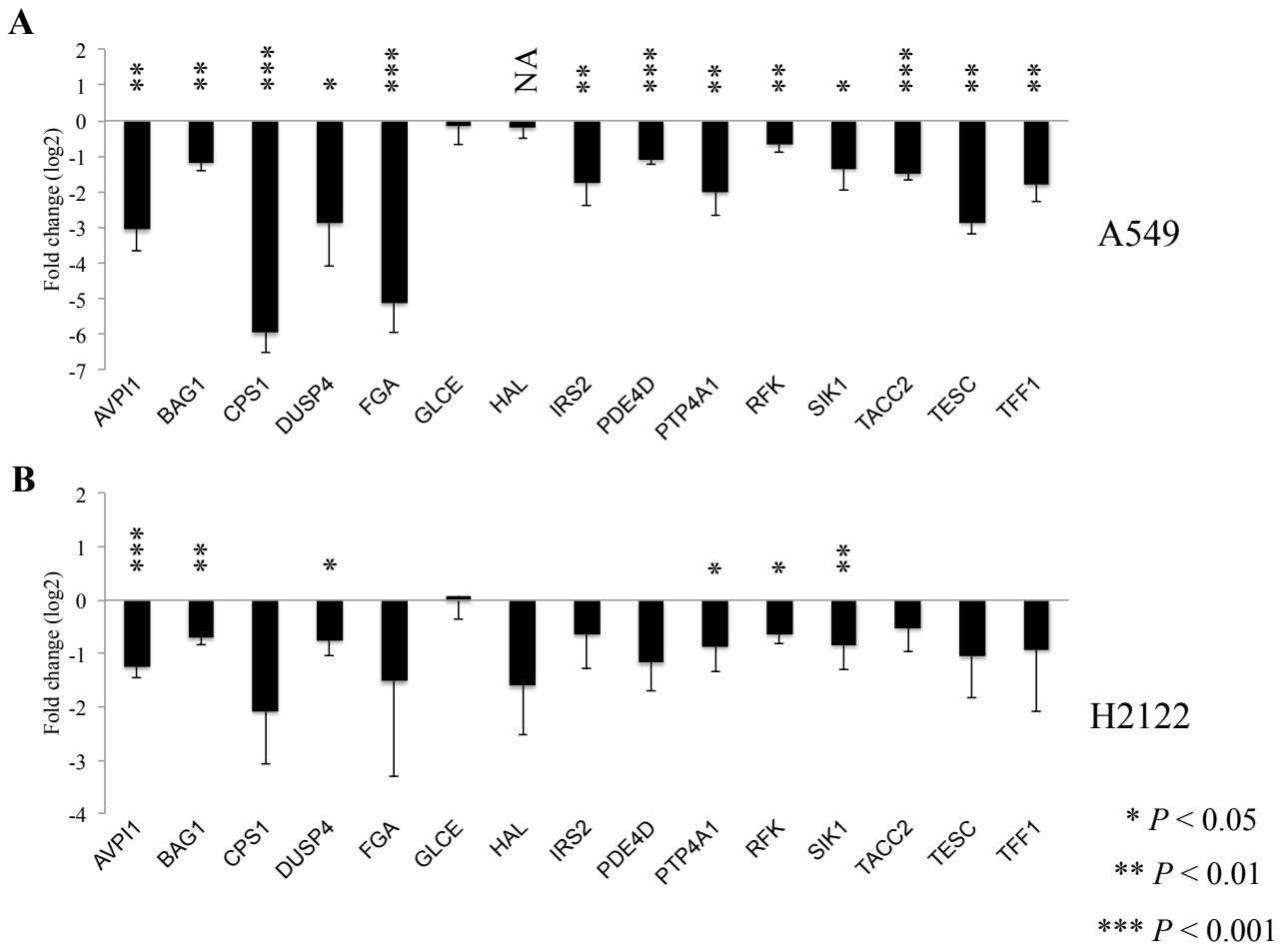


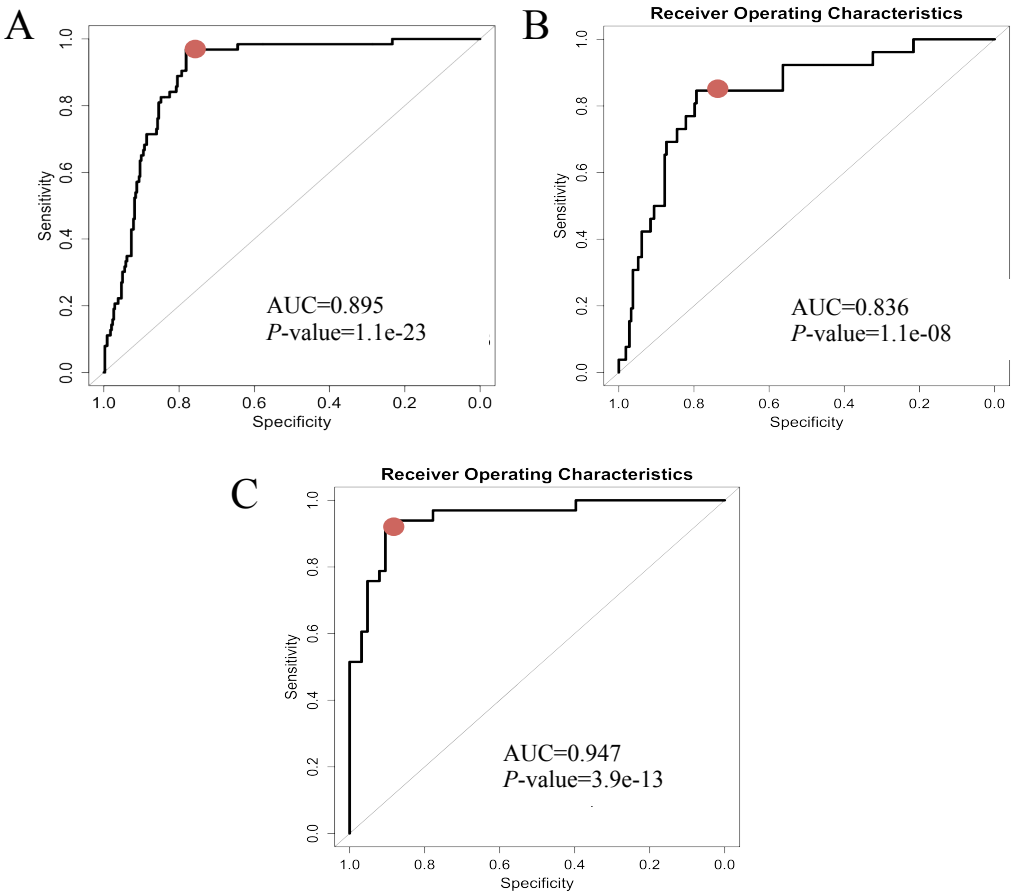
Supplementary Figure S1. Gene Set Enrichment Analyses of LKB1 associated genes. A, Gene set enrichment analysis from six resected human lung adenocarcinoma datasets, **B,** four studies of cell lines, including the LKB1 add-back experiments performed in this work, **C,** and four comparisons from two mouse studies are shown. Analysis was performed using the top 200 genes associated with LKB1 in each of the listed studies, compared to ranked associations with LKB1 mutations from the TCGA (x-axis). The distribution of the 200 genes is shown as black hash marks along the x-axis. The enrichment score is plotted on the y-axis, and the maximum enrichment score is given for each plot. Also shown are the number of overlapping genes with the top 200 genes in the TCGA study, and the p-value for the significance of this overlap, calculated using the hypergeometic test. The number of overlapping genes expected by chance is four.



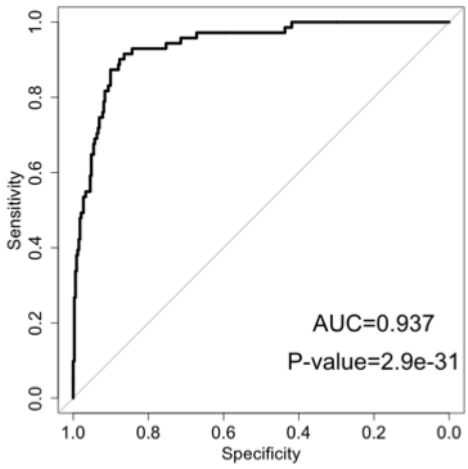
Supplementary Figure S2. Genes comprising each of the four clusters identified in hierarchical clustering. Hierarchical clustering of LKB1-associated genes was performed for the Director's Challenge Michigan cohort (129 genes, 178 tumors) and TCGA lung adenocarcinomas (110 genes, 446 tumors). Four clusters of correlated genes were observed in each patient cohort, as shown in Figure 2 B,C of the main text. The genes comprising these clusters are shown for each dataset (**A-D)**, with overlapping genes indicated. Asterisks indicate genes represented by probesets in Michigan data (Affymetrix U133A microarray platform) that were not included in TCGA RNA sequencing data. P-values represent the results of the Fisher's exact test to show the significance of overlap of the two gene lists considering genes represented in both studies and selected from 69 over-expressed genes (**A-C)** or 41 under-expressed genes (**D**).



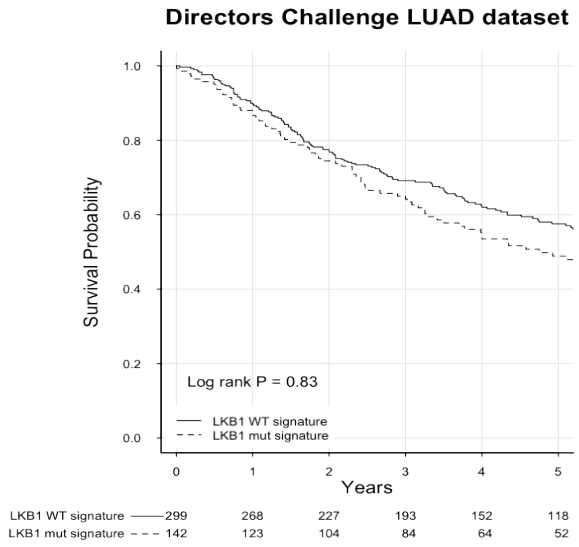
Supplementary Figure S3. Expression of wild-type LKB1 in A549, H2122, or HeLa cell lines decreases the expression of the genes in the CREB transcriptional node.
Microarray gene expression was measured in triplicate after stable expression of pBABE vector or LKB1-wild-type in **A,** A549, or **B**, H2122,. For each gene comprising the CREB transcriptional node the average change in gene expression (log base 2) is plotted comparing LKB1 wild-type to pBABE control. Error bars represent standard deviations; P-values represent the results of student's t-test comparing these groups. NA indicates minimal basal expression.



Supplementary Figure S4. Receiver operating characteristics for LKB1-loss score in resected lung adenocarcinomas and NSCLC cell lines. Receiver operating curves to show the relationship of sensitivity and specificity of the LKB1 loss score for detecting LKB1 mutations in **A,** the TCGA lung adenocarcinomas (67 LKB1 mutations; 406 total), **B**, pooled analysis of other resected lung adenocarcinomas (26 LKB1 mutations; 239 total), or **C**, NSCLC cell lines (39 LKB1 mutations; 85 total). AUC refers to area under the curve, and the P-value reflects the results of the Mann-Whitney U test. The red dot indicates the location of the cutoff score of 0.2 used in this work.



Supplementary Figure S5. Receiver operating characteristics showing association between expression of NRF2 cluster and KEAP1 mutations in TCGA data. Relationship between sensitivity and specificity is shown for the detection of KEAP1 mutations (n = 71; total = 403) with the expression of the NRF2 cluster. AUC refers to area under the curve, and the P-value reflects the results of the Mann-Whitney U test.



Supplementary Figure S6. Association between patient outcome and LKB1 loss.
Resected LUAD tumors from the Director's Challenge Consortium (n=441), were classified as LKB1-loss or LKB1 WT using the LKB1-classifier score. Kaplan-Meier curves were used to plot cumulative events for these two groups for overall survival, P-values represent the results of the log-rank test; the number of evaluable tumors remaining are given at yearly intervals below each plot.