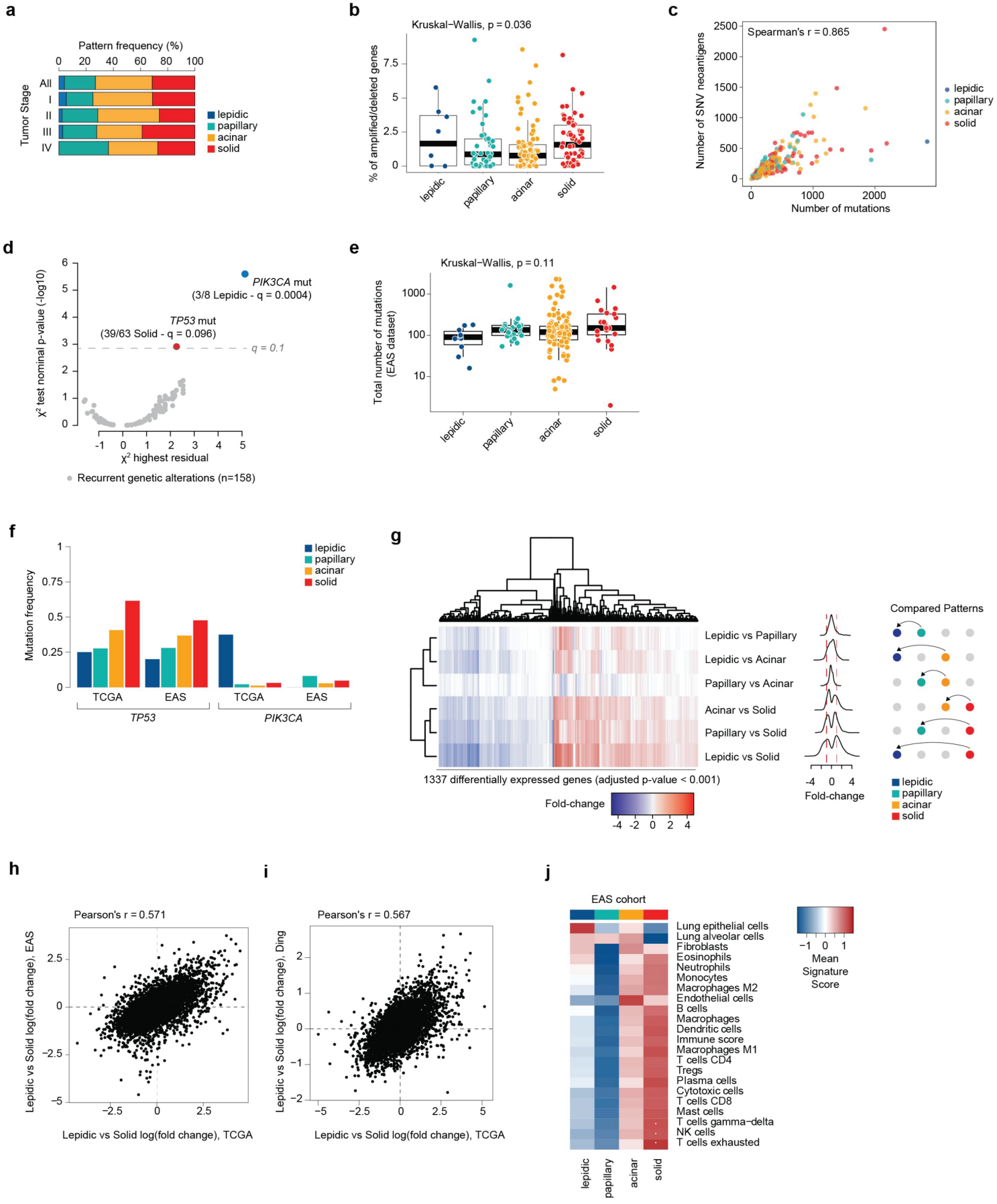
Supplementary Fig. S1



Supplementary Fig. S1

a) Frequency of TCGA samples annotated to each pattern for all tumor stages and for each tumor stage separately.

b) Boxplot comparison of the percentage of amplified or deleted genes (Y-axis) predicted by the GISTIC algorithm (GISTIC value = 2 or = -2) in each sample annotated for each of the 4 histologic patterns.

c) Comparison of total number of mutations per sample (X-axis) and number of predicted neoantigens by NetMHC (Y-axis) in the TCGA cohort. Samples are color coded by prevalent pattern.

d) Volcano plot showing Chi-square test p-value of the association between somatic alterations and histologic patterns (Y-axis) and the highest Chi-square residual (X-axis). Significant alterations are highlighted and color coded.

e) Total number of somatic coding mutations (Y-axis, log10 scale) in EAS samples (colored points) stratified by histologic pattern classification (X-axis).

f) Mutation frequency across patterns for selected genes (TP53 and PIK3CA) in TCGA and EAS datasets.

g) Gene expression log2-fold changes between each pair of patterns (right) for significantly differentially expressed genes represented as heatmap (left) and density distributions (center), with dashed lines highlighting fold changes of +/-2.

h,i) Scatter plots showing gene expression log2-fold changes for all genes between lepidic and solid patterns in TCGA dataset (X-axis) and EAS **(h)** and Ding **(i)** datasets (Y-axis).

j) Mean mRNA signature scores for multiple cell types (rows) within each histologic pattern subtype (columns) in EAS dataset. Values are normalized by rows (Z-scores). Sorting of cell types is the same as for TCGA dataset (**Fig. 1**).