Additional File 2 - Iorio et al.: Dissecting the genomic heterogeneity of cancer hallmarks' acquisition with SLAPenrich Legends of Supplementary Figures and Tables

Supplementary Figure S1: Visualization of enriched pathways, core-components and differential pathway enrichment analysis for the LUAD case study. (A) Heatmap summarizing the status of the genes belonging to a pathway enriched at the population level in the case study of the lung adenocarcinoma dataset. Genes and patient samples (respectively on rows and columns) have been permuted with a dedicated function in order to highlight mutual exclusivity trends in the observed somatic alterations. (B) Heatmap showing a sub-set of genes (on the rows) shared by multiple significantly enriched pathways (on the columns), together with a bar plot diagram (on the right) showing the percentages of patient samples where each gene is altered. These figures are automatically generated by SLAPenrich. (C) Visual output of the differential enrichment analysis function using the case study lung adenocarcinoma dataset in input, and stratifying patients based on their smoking status. The heatmap on the left shows the alteration status of the top/bottom 10 most positively/negatively differentially enriched pathways between the groups of smokers vs non-smokers (on the column); the heatmap in the centre shows enrichment significance of individual pathways in the two sub-populations, and the barplot shows corresponding differential sample level enrichment scores.

Supplementary Figure S2: Differential pathway enrichment analysis results: non-mucinous vs mucinous bronchioloalveolar LUAD patients. Results from a differential SLAPenrich analysis obtained contrasting two sub-pulations of LUAD patients based on their bronchioloalveolar type (non-mucinous vs. mucinous).

Supplementary Figure S3: Pathway enrichment significance correlation between SLAPenrich and PathScan.

(A) Number of enriched pathways detected by SLAPenrich, PathScore and both methods across cancer types (left barplot), and overlap significance (right barplot); (B) Percentages of tissue specific high-confidence cancer driver genes included in the top 10, 20, 50 and 100 enriched pathways according to SLAPenrich and PathScore across cancer types (first 4 barplots), and in whole set of statistically significantly enriched pathway (FDR < 5% for SLAPenrich and adjusted p-value < 0.05 for PathScore; (C) Percentages of tissue specific high-confidence cancer driver genes included in the top *k* enriched pathways according to SLAPenrich and PathScore. For SLAPenrich, all the possible k values are considered; PathScore does not output results for all the tested pathways but only for the significantly enriched one, therefore in this case *k* ranges from 1 to the least significantly enriched pathway.

Supplementary Figure S4: Enriched pathways versus sample size, downsampled analyses, and covered known cancer genes. (A) Number of significantly enriched pathway at the population versus the number of samples available in the analysed cohorts, across cancer type. (B) Number of significantly enriched pathway at the population level across 5 different cancer types (with more than 350 samples), indicated by different colors, and down-sampled trials. In each of this trials, for each cancer type and 50 different iterations, a set of n samples is randomly selected and a SLAPenrich analysis is performed on this sub-set of data. Average number of SLAPenriched pathway (and standard deviations) are reported. n = 800, 400, and 250 for BRCA and n = 250 for the other four cancer type. For four of the tested tissues there is no tendency for increased number of samples to produce more SLAPenriched pathways. A mild dependency trend is observable for BRCA only, with a continuously increasing average number of enriched pathways as a function of sample size up to 800 samples, that plateaus above this size, with a very similar number of enriched pathways when analysing 1,132 samples or across 50 analysis on 800 pathways. (C) Each bar quantifies the ratio of high-confidence cancer genes contained in at least one pathway enriched at the population level (covered pathways), across cancer types. Different contained colored bars indicate the ratio of the genes included in covered pathways associated to different hallmarks, one colored bar per hallmark. The white bar at the top indicate the ratio of genes included in covered pathways associated to multiple hallmarks.

Supplementary Figure S5: Hallmark heterogeneity across cancer types. Heatmaps showing pathways enrichments at the population level across cancer types for individual hallmarks. Color intensities correspond to the enrichment significance. Cancer types and pathways are clustered using a correlation metric. See also Figure 4.

Supplementary Figures S6, S7: Impact of known cancer genes' mutations on the results. Heatmaps showing, for each enriched-pathway/cancer-type, the ratio between the number samples harbouring mutations in known cancer genes belonging to the pathway under consideration and the total number of samples harbouring mutations in any gene belonging to the pathway under consideration. See also Figure 6.

Supplementary Figure S7: Hallmark signature analysis to discover new cancer driver networks. In each row, first circle plots shows pathway enrichments at the population levels when considering all the somatic variants (bars on the external circle) and when considering only variants not involving known high-confidence cancer driver genes; second

circle plot shows similarly a comparison between the hallmark signatures resulting from SLAPenrich analysis including (bars on the external circle) or excluding (bars on the internal circle) the variants involving known high-confidence cancer genes. The bar plot shows a comparison, in terms of true-positive-rate (TPR) and positive-predicted-value (PPV), of the SLAPenriched pathways across the two analysis and, finally, the scatter plots on the right shows a comparison between the resulting hallmark signatures.

Supplementary Table S1: KEGG pathways enriched in the LUAD case study dataset

Supplementary Table S2: SLAPenrich/PathScan results' comparison

Supplementary Table S3: SLAPenrich/PathScan results' comparison

Supplementary Table S4: Differential enrichment analysis comparing LUAD Smokers vs. Non-Smokers patients (Results)

Supplementary Table S5: Differential enrichment analysis comparing LUAD Mucinus vs. Non-Mucinus BAC types (Results)

Supplementary Table S2: Keywords used to manually curated the mapping between genes, pathways and hallmarks

Supplementary Table S3: Manually curated mapping between genes, pathways and hallmarks