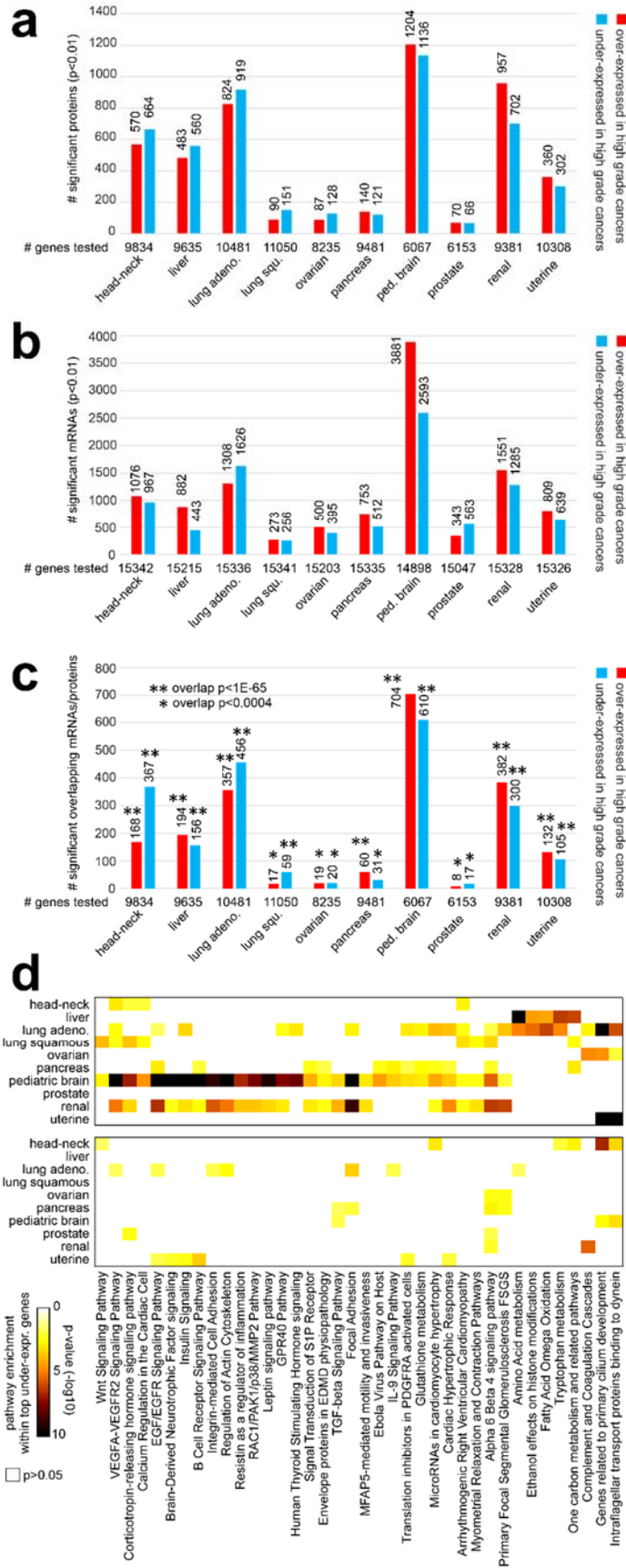


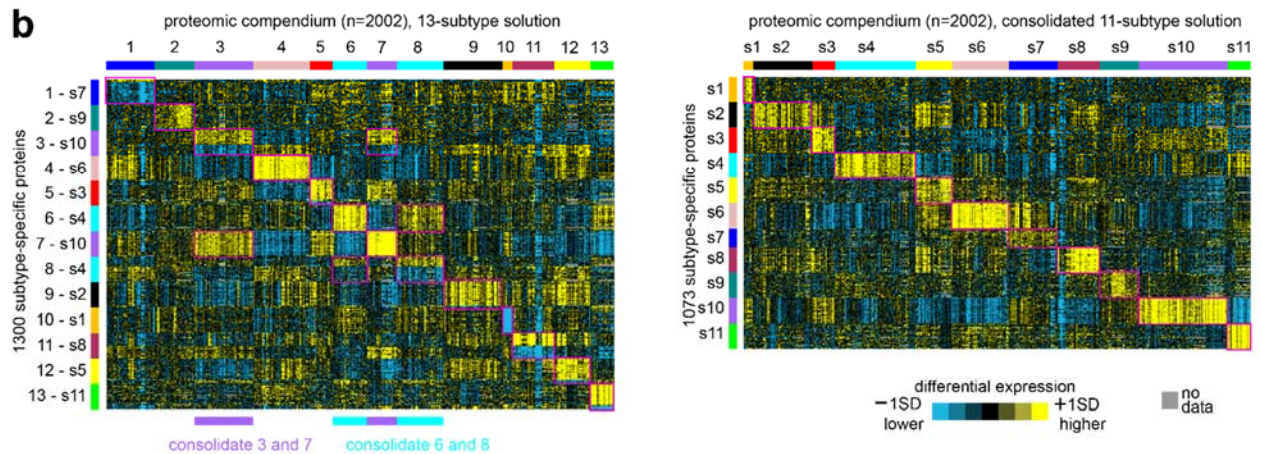
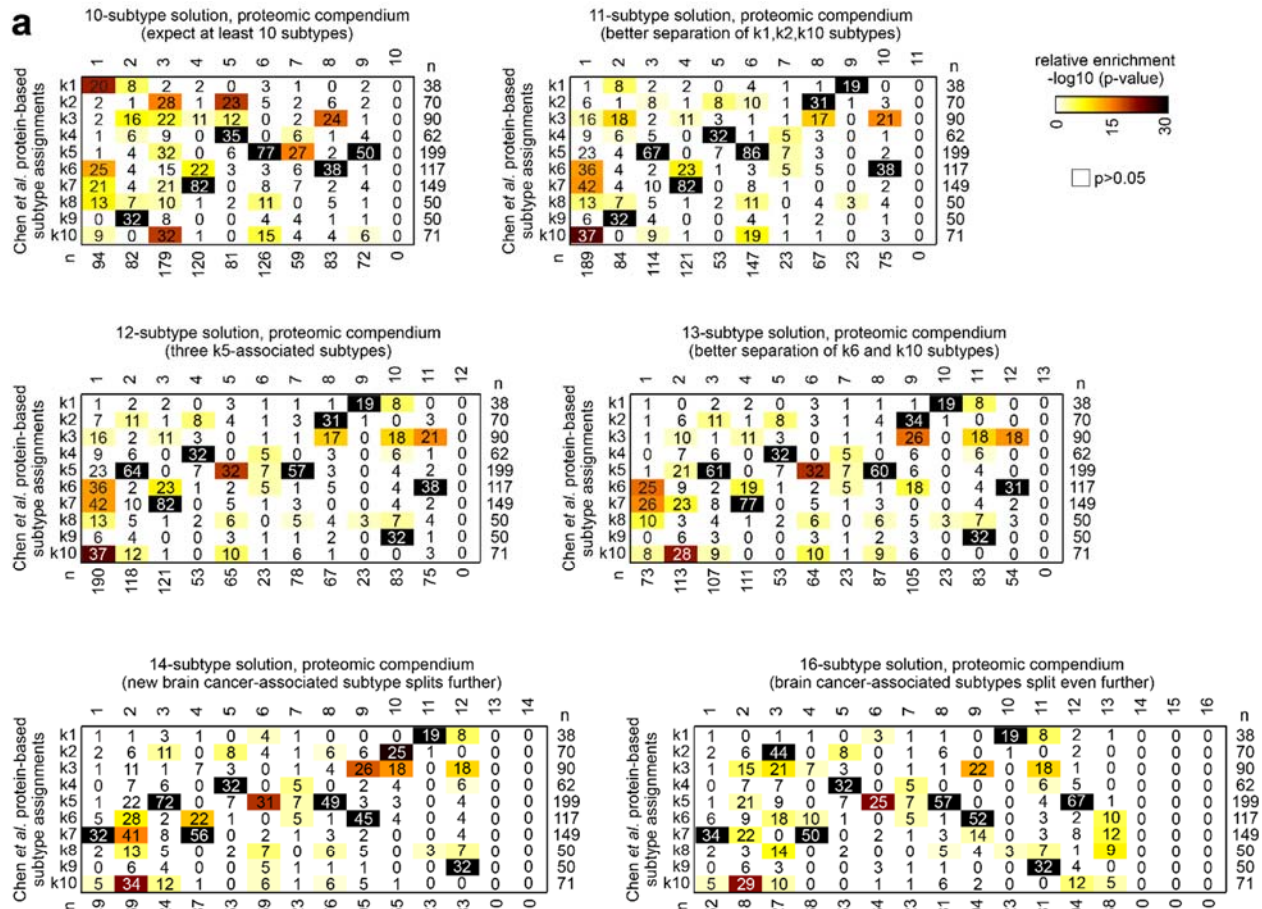
Supplementary Information

Zhang, Chen, Chandrashekar et al. "Proteogenomic characterization of 2002 human cancers reveals pan-cancer molecular subtypes and associated pathways"



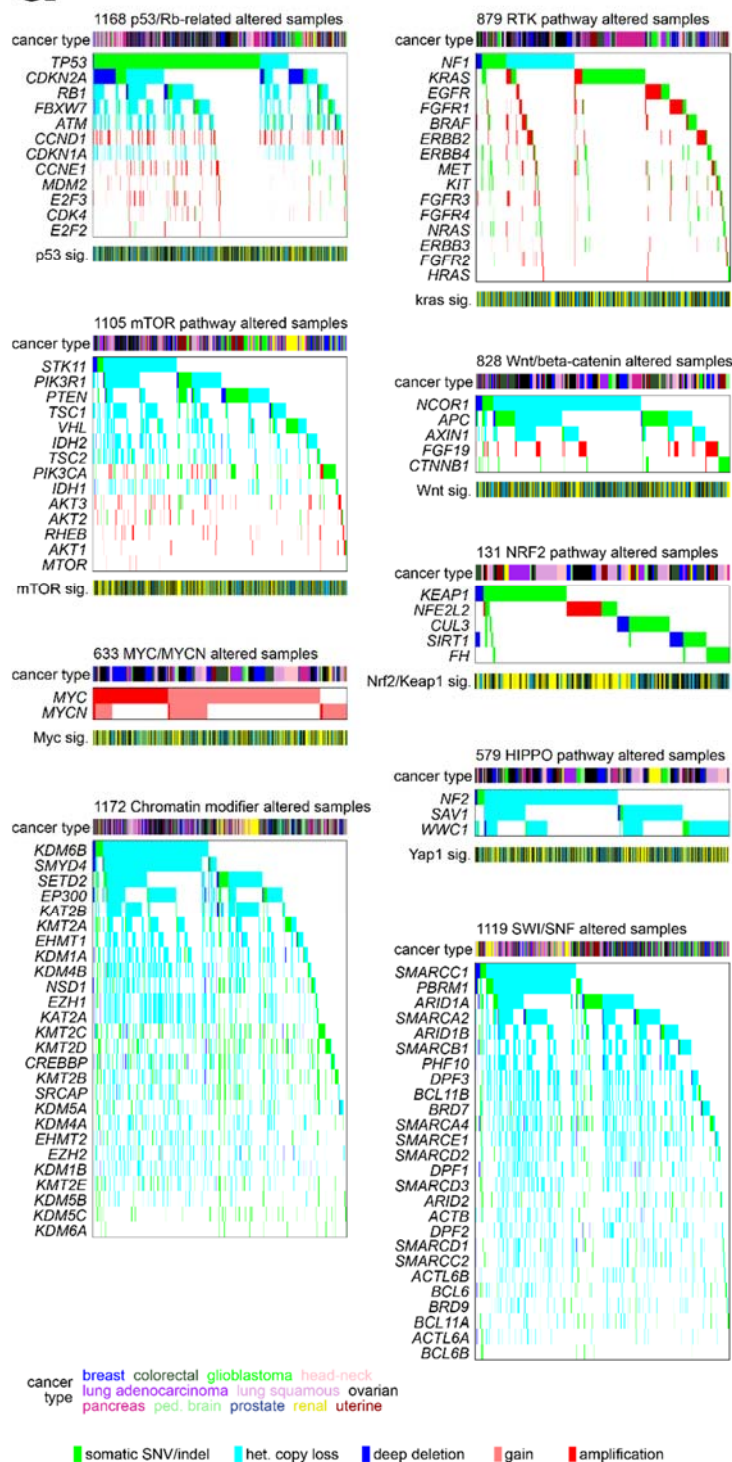
Supplementary Figure 1, related to Figure 2. Additional information on proteomic and transcriptomic signatures of high-grade cancers. (a)

For each indicated cancer type, numbers of top differentially expressed proteins ($p < 0.01$, Pearson correlation using log-transformed data), associated with higher cancer grade. The numbers of unique genes tested are indicated for each comparison (involving proteins for which measurements were made in over half of samples profiled). **(b)** Similar to part a, but for differentially expressed mRNAs. The numbers of unique genes tested are indicated for each comparison. **(c)** Overlapping top protein and mRNA features ($p < 0.01$ for protein and $p < 0.01$ for mRNA, based on shared gene) by comparisons according to grade. P values for significance of overlap by one-sided Fisher's exact tests. **(d)** Significance of enrichment (by one-sided Fisher's exact test) for wikiPathway gene sets with the respective sets of proteins and mRNAs under-expressed ($p < 0.01$, Pearson) with tumor grade for each cancer type represented. The set of pathways represented were significant within the under-expressed proteins for at least one cancer type with $FDR < 10\%$ and for at least two cancer types with $p < 0.01$.

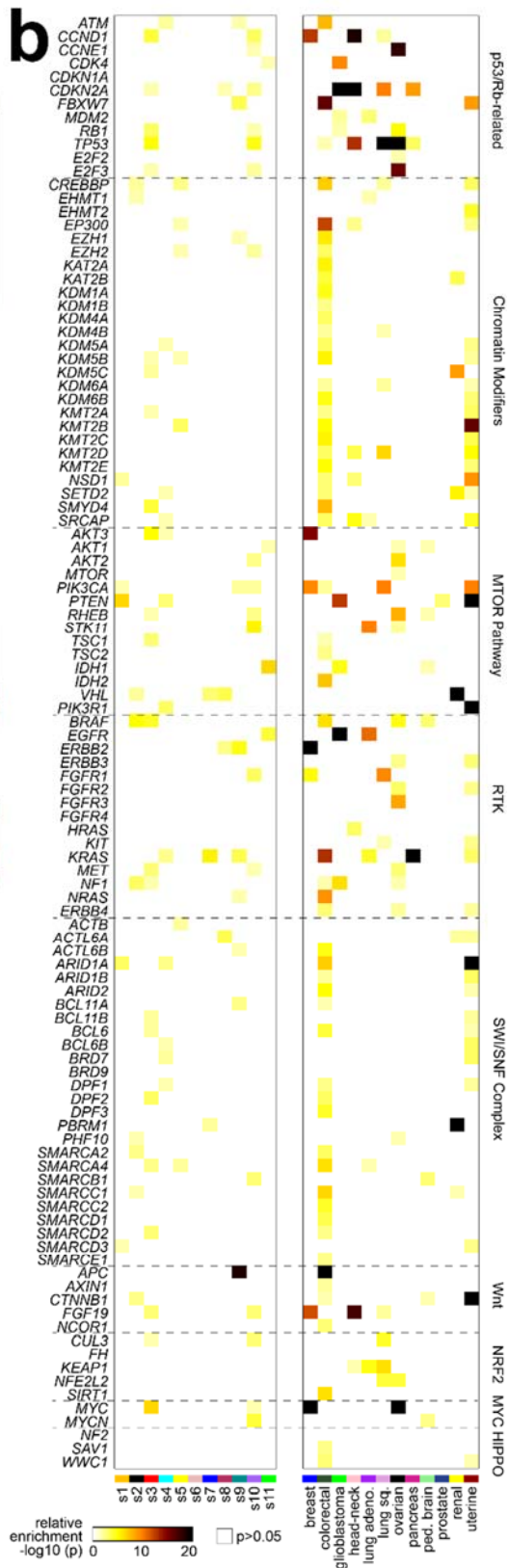


Supplementary Figure 2, related to Figure 3. Determining the number of *de novo* pan-cancer proteomic subtypes. (a) Consensus ward linkage hierarchical clustering identified $k = 2$ to $k = 20$ subtypes, based on proteomic data from 2002 human tumors (referred to here as the “proteomic compendium”). The significance of overlap of previously identified pan-cancer proteomic subtypes¹ (k1-k10, rows) with subtypes obtained from the current proteomic compendium dataset. Subtype solutions from 10 subtypes to 16 subtypes are considered here. P-values by one-sided Fisher’s exact test. We expected at least ten subtypes, based on previous studies. With an 11-subtype solution, we observed better separation of the previously identified k1, k2, and k10 subtypes. With a 12-subtype solution, we observed three distinct k5-associated subtypes. With a 13-subtype solution, we observed better separation of k6 and k10 subtypes. With successive subtype solutions (14- and 16-subtype solutions represented here), the brain cancer-associated subtype (“s11” subtype in main Figure 3) splits further and further, but with no distinctive biology being associated with these subdivisions. The 13-subtype solution was therefore selected and explored further below. **(b)** Across 2002 tumor proteomic profiles, differential expression patterns (values normalized within each tissue-based cancer type; SD, standard deviation from the median) for a set of 1300 proteins (left) and for another set of 1073 proteins (right) found to best distinguish between the respective subtypes considered (top ~100 over-expressed proteins for each subtype). On the left, the 13-subtype solutions from consensus ward linkage hierarchical clustering (part a) is considered, where visually, subtypes 3 and 7 look very similar to each other, and subtypes 6 and 8 look similar to each other. Therefore, we both consolidated subtypes 3 and 7 into one subtype (“s10”) and consolidated subtypes 6 and 8 into another subtype (“s4”) to arrive at the final 11 subtype solution, represented on the right.

a



b



Supplementary Figure 3, related to Figure 7. Somatic alteration patterns across cancers by pathway annotation. (a) For the pathways from main Figure 7, somatic alteration events involving each gene included in the pathway are represented. Where applicable, the corresponding pathway signature patterns are shown (for p53 signature, based on mRNA; for the other signatures, based on protein). **(b)** For each cancer type and pan-cancer proteomic subtype, significance of enrichment (one-sided Fisher's exact test) of mutation events for each gene within the given proteomic subtype/cancer type versus the rest of the tumors.

References

1. Chen, F., Chandrashekar, D., Varambally, S. & Creighton, C. Pan-cancer molecular subtypes revealed by mass-spectrometry-based proteomic characterization of more than 500 human cancers. *Nat Commun* **10**, 5679 (2019).