S5  Appendix

For transcriptome data, the pre-processing steps included:

- Like in [2] gene-level expected counts were upper-quartile-normalized to 1000.
- log₂ transformation.

For proteome data, the pre-processing steps included:

- log₂ transformation.
- Normalization by median substraction.
- Filtering out proteins which were detected in less than 50% of samples.
- For clustering only: imputation of missing values using the R package impute [3]. For differential expression analysis, we used unimputed values.

For phosphoproteome data, the pre-processing steps included:

- log₂ transformation.
- Normalization by median substraction.
- Filtering out proteins which were detected in less than 50% of samples.
- Batch correction with the R package edgeR.
- For clustering only: imputation of missing values using the R package impute [3]. For differential expression analysis, we used unimputed values.

The CNA data was obtained at the gene level from the study by Ng et al. [2]. The copy number status was derived from the log-ratio and takes values from 2 to −2, which denote [1]:

- 2: amplification
- 1: copy gain indicates a low-level gain
- 0: copy number neutral
- -1: shallow deletion, indicating a heterozygous deletion
- -2: deep deletion, indicating a homozygous deletion

This way, despite the ordinal nature of the CNA data, the range of the values justifies the normal approximation.
References

