Identification of *MYCN* non-amplified neuroblastoma subgroups points towards molecular signatures for precision prognosis and therapy stratification

Supplementary Materials

Table of contents

Supplementary Methods

1. Data Collection

We searched the keywords "(MYCN) OR (MYC) OR (MYCN amplification) AND (Neuroblastoma) AND (Homo sapiens)" and publication dates before 31/05/2020 in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) and European Bioinformatics Institute (EBI) ArrayExpress (AE). Three hundred and five experiments were identified in the initial screening. Then, we included only datasets that are neuroblastoma samples and contain *MYCN* status information with at least 2 biological replicates. Datasets generated on platforms other than Illumina, Affymetrix or Agilent or detected genes less than 10,000 were excluded to avoid the technical mismatch between different platforms. Twenty datasets with 3,853 samples passing these inclusion criteria were described in **Supplementary Table 1**.

2. Data Preparation

Raw microarray files were downloaded and imported into the R environment (v4.0.2). The normalisation of raw data depended on the generated platform. Affymetrix datasets were performed by the rma function in affy (v1.66.0)[1] or oligo(v3.11)[2] packages. Agilent microarrays were normalised using the normalise BetweenArrays function. Illumina datasets were standardised by neqc function in limma (v3.44.3)[3]. Microarray probe IDs were mapped to gene symbol according to the GPL annotation files provided in NCBI. Probes mapped to multiple gene symbols were removed and genes mapped to multiple probe IDs were summarised by calculating the mean.

3. Quality Control

To ensure reliability, we filtered out low-quality datasets by performing leave-one-out-cross-validation. Nineteen out of 20 datasets were training data for optimising the *MYCN*-amplification signature and evaluating the prediction performance of this signature on the one left. Two datasets (GSE73537 and GSE53371) were filtered out due to the low area under receiver operating characteristic (ROC) curve (AUROC) score. The generation of a *MYCN*-amplification signature and the validation of this signature on the dataset left were run through MetaIntegrator package (v2.1.3)[4] with FDRThresh = 0.05 and effectSizeThresh = 1.3.

4. Consensus Clustering

Qualified datasets (18 in total) were merged into one dataset using gene symbols as references. The ComBat function in sva (v3.36.0)[5] was then used to correct for the technical batch effect. The merged dataset was randomly split into a training cohort (n = 2,160) and a testing cohort (n = 925) in a 7:3 ratio using the caret package[6]. Median absolute deviations were calculated for each gene and the top 50% most variant genes (n = 5,792) were extracted for consensus clustering (ConsensusClusterPlus, v1.52.0)[7]. The consensus was performed using K-means with Euclidean distance, 80% item resampling (pItem), 100% gene resampling (pFeature) and 10,000 iterations to generate the robust consensus clusters. For TARGET RNA-seq cohort, we performed hierarchical clustering with spearman distance. The delta area plot and the Cluster-Consensus score suggested a matrix with k = 3 displayed the highest stability within clusters and clearest cut among clusters. To ensure precise clinical and molecular characterisations of subgroups, the silhouette coefficient score of each sample was computed by the cluster package (v2.1.2)[8] and samples with silhouette width values less than 0 were labeled as not defined.

5. Defining the Differentially Expressed Genes (DEGs) and Pathway Analysis

The limma package was used to compare each subgroup to other subgroups. Genes with an absolute $\log_2$ fold change (FC) bigger than 1 and a false discovery rate (FDR) p value less than 0.05 adjusted by using Benjamini–Hochberg (BH) method (or q-value) were considered as differentially expressed genes (DEGs).

Pathway enrichment analysis were generated through Metascape website (http://metascape.org) and Ingenuity Pathway Analysis (IPA)[9] software. Parameters of Metascape were set as "5 minimum overlapping genes, $p < 0.05$ and 1.5 minimum enrichment factor"[10]. GSEA (gene set enrichment analysis) was performed using GSEA software (the Broad Institute platform, v4.0.3)[11] with the default settings and 1,000 gene set permutations. Single-sample GSEA (ssGSEA) scores of pathways were calculated by using ssGSEA in the GSVA (v1.36.2) package[12]. "MYCN", "ADRN (adrenergic)" and "MES (mesenchymal)" signatures were collected from the previous reports[13, 14]. The MHC (major histocompatibility comple) score was calculated as the mean expression of HLA-A, HLA-B, HLA-C, TAP1, TAP2, NLRC5, PSMB9, PSMB8, and B2M[15]. The CYT (cytolytic activity) score was calculated as the geometric mean of GZMA and PRF1[16].

6. Weighted Gene Co-expression Network Analysis (WGCNA) and Protein Protein Interaction (PPI) Analysis

WGCNA[17] of the top 50% most variant genes (n = 5,792) was used to discover sets of genes with similar correlation patterns among subgroups. To choose the parameter (soft threshold) of the power adjacency function, we used the scale-free topology (SFT) criterion. We selected the power value as 5 because the model-fit saturation was above 0.8. The identification of modules was performed using cutreeDynamic function with the signed hybrid method, deepSplit = 2 and minClusterSize = 100. Next, automatic merging was performed using mergeCloseModules function with a cutHeight of ≤ 0.25, which means a Pearson correlation between module eigengenes of ≥ 0.75.

PPI in each module was demonstrated by STRING (v11.0)[18] with high confidence score > 0.700 interaction. Visualisation of network was performed by gephi[19].

7. Clinical Characterisation of Subtypes

The univariate and multivariate Cox proportional hazards model assessed the hazard ratio of each parameter through the survminer (v0.4.9)[20]. We performed log-rank test to compare Kaplan-Meier survival curves between each subgroup by survival (v3.2-10)[21]. Prediction error curves of each prognostic model were generated from pec (v2019.11.03)[22].

8. Submap Analysis

To compare the subgroup across independent neuroblastoma cohorts and melanoma datasets (GSE78220)[23, 24], we applied an unsupervised subclass mapping (SubMap) analysis from GenePattern module (https://www.genepattern.org/modules) to evaluate similarity of subgroups[25]. Gene used were the intersection of genes between two correspondent datasets. Module parameters were setting as default and a Bonferroni adjusted p value less than 0.05 was considered as the significant cutoff.

9. Single cell RNA-seq (scRNA-seq) analysis

A scRNA-seq dataset of 3 *MYCN*-non amplified neuroblastoma sample was downloaded from GEO (GSE137804)[26]. The processed gene expression matrix was imported into R and analyzed by Seurat Packages[27]. We first removed low-quality cells with detected genes less than 500 and those with more than 10% genes from the mitochondrial genome. The filtered gene expression matrix was normalised by SCTransform function. Variable features across 3 sample were found

using SelectIntegrationFeatures function and then identified anchors using the FindIntegrationAnchors function, follwed by the IntegrateData function to integrate these 3 sample. The integrated data were visualised using Uniform Manifold Approximation and Projection (UMAP) with RunUMAP funtion. The annotation of each cell type in the integrated dataset was performed using established signatures from a previous report[27].

10. CIBERSORTx Analysis

A signature matrix of each cell type in *MYCN*-nonamplified neuroblastoma samples was constructed by a deconvolutional tool CIBERSORTx [28]. Considering the memory limitation of CIBERSORTx, we randomly divided the train cohort into 3 partitions using the caret[6] package. Cell fractions in each partition from the train cohort as well as those from the test cohort were imputed using Cell Fraction analysis module with B-mode and other default parameters over 1,000 permutations[29].

11. Analysis of Clinically Actionable Genes and Drug Response

To investigate subgroup-specific druggable targets, we performed an integrative analysis to assess the associations between molecular features and the response to anticancer drugs in *MYCN* non-amplified neuroblastomas.

We downloaded gene dependency score CERES to estimate gene-dependency levels (https://depmap.org/ceres/). Meyers and colleagues developed CERES, a computational method to estimate gene dependency levels from CRISPR-Cas9 essentiality screens while accounting for the copy-number-specific effect[30]. DEGs in each subgroup with an average gene dependency score (CERES) in 20 neuroblastoma cell lines in the DepMap database less than -0.6 were considered as potential drug targets.

To assess drug response in cancer cell lines, we downloaded the drug sensitivity area under the dose–response curve (AUROC) and gene expression profiles for cancer cell lines from the GDSC (http://www.cancerrxgene.org/downloads)[31]. We calculated the Spearman's rank correlation between gene expression and the AUCs from the GDSC[32] and used Spearman's rank correlation coefficient $|r_s| > 0.3$ and p value < 0.01 for statistical significance.

12. Identification of Independent Predictors

To identifiy independent predictors for subgrouping, we applied a multi-cohort analysis pipeline via MetaIntegrator[4] and validated with the machine learning classifier, support vector machine (SVM).

In brief, we applied a multi-cohort analysis pipeline to find potential predictors via MetaIntegrator. Six modules with 1,322 differentially expressed genes (DEGs) (FDR < 5% and effect size > 1.3-fold) were identified among 3 subgroups. Of the DEGs in subgroup 1, 62 were up-regulated and 20 were down-regulated; in subgroup 2, 311 were up-regulated and 53 were down-regulated; in subgroup 3, 799 were up-regulated and 77 were down-regulated. We then performed a greedy forward search[33] to these 1,322 DEGs and found a set of 43 DEGs presenting the optimal predicted values. We calculated the predicted scores of each subgroup by subtracting the geometric mean expression of down-regulated signatures from the geometric mean expression of up-regulated signatures within the same subgroup. To facilitate the comparisons across different platforms, z-scored transformation was applied to each sample independently[34].

The original scope of MetaIntegrator was to find the best threshold of calculated scores to classify different conditions based on prior information. This tool failed to predict novel datasets. MetaIntegrator can only use binary classification, which means we should make at least 3 comparisons to classify subgrous. Therefore, to solve this issue, we built 2 machine learning classifiers (support vector machine, SVM; and XGBoost) based on the 43 DEGs identified above. First, we performed grid and random search to identify the hyperparameters that optimise the classifier performance. For SVM, we used 2 hyperparameters, the C penalty term (the cost value in SVM) and the kernel bandwidth, γ, and the model was trained with 200 cost values and 200 kernel bandwidth values. For XGBoost, we applied random search of 10,000 hyperparameters configurations including 1) maximum tree depth; 2) learning rate; 3) minimum loss reduction required to introduce a split; 4) fraction of samples for each tree; 5) fraction of columns for each tree; 6) minimum child weight; 7) maximum delta step; 8) L2 regularization term and 9) tree methods (exact or approximate). All these analysis were processed in the caret[6] and e1071[35] packages. We then used the test cohort to verify and determine the accuracy rates of the two methods: the the accuracy rate was 87.8%, and 87.2% and AUROC was 93.14% and 92.12%, respectively. Therefore, we selected SVM as the final model. The parameters were a cost value of 46.78894 and a gamma value of 1.043065.

References

1. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307-315 (2004).

2. Carvalho BS, Irizarry RA. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363-2367 (2010).

3. Ritchie ME*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

4. Haynes WA*, et al.* Empowering multi-cohort gene expression analysis to increase reproducibility. *Pac Symp Biocomput* **22**, 144-153 (2017).

5. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).

6. Kuhn M. caret: Classification and Regression Training. R package version 6.0-86 (2020).

7. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572-1573 (2010).

8. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2 (2021).

9. Krämer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **30**, 523-530 (2014).

10. Zhou Y*, et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).

11. Subramanian A*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

12. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

13. Valentijn LJ*, et al.* Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. *Proc Natl Acad Sci U S A* **109**, 19190-19195 (2012).

14. van Groningen T*, et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* **49**, 1261-1266 (2017).

15. Lauss M*, et al.* Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nat Commun* **8**, 1738 (2017).

16. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61 (2015).

17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).

18. Szklarczyk D*, et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-d613 (2019).

19. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*(2009).

20. Kassambara A, Kosinski M. survminer: Drawing Survival Curves using 'ggplot2'. R package version 0.4.9(2019).

21. Therneau T. A Package for Survival Analysis in S. R package version 3.2-10(2015).

22. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. *J Stat Softw* **50**, 1-23 (2012).

23. Roh W*, et al.* Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* **9**, (2017).

24. Hugo W*, et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35-44 (2016).

25. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).

26. Dong R*, et al.* Single-Cell Characterization of Malignant Phenotypes and Developmental Trajectories of Adrenal Neuroblastoma. *Cancer Cell* **38**, 716-733.e716 (2020).

27. Stuart T*, et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e1821 (2019).

28. Newman AM*, et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019).

29. Le T, Aronow RA, Kirshtein A, Shahriyari L. A review of digital cytometry methods: estimating the relative abundance of cell types in a bulk of cells. *Brief Bioinform*, (2020).

30. Meyers RM*, et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784 (2017).

31. Kim ES*, et al.* The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* **1**, 44-53 (2011).

32. Xiang Y*, et al.* Comprehensive Characterization of Alternative Polyadenylation in Human Cancer. *J Natl Cancer Inst* **110**, 379-389 (2018).

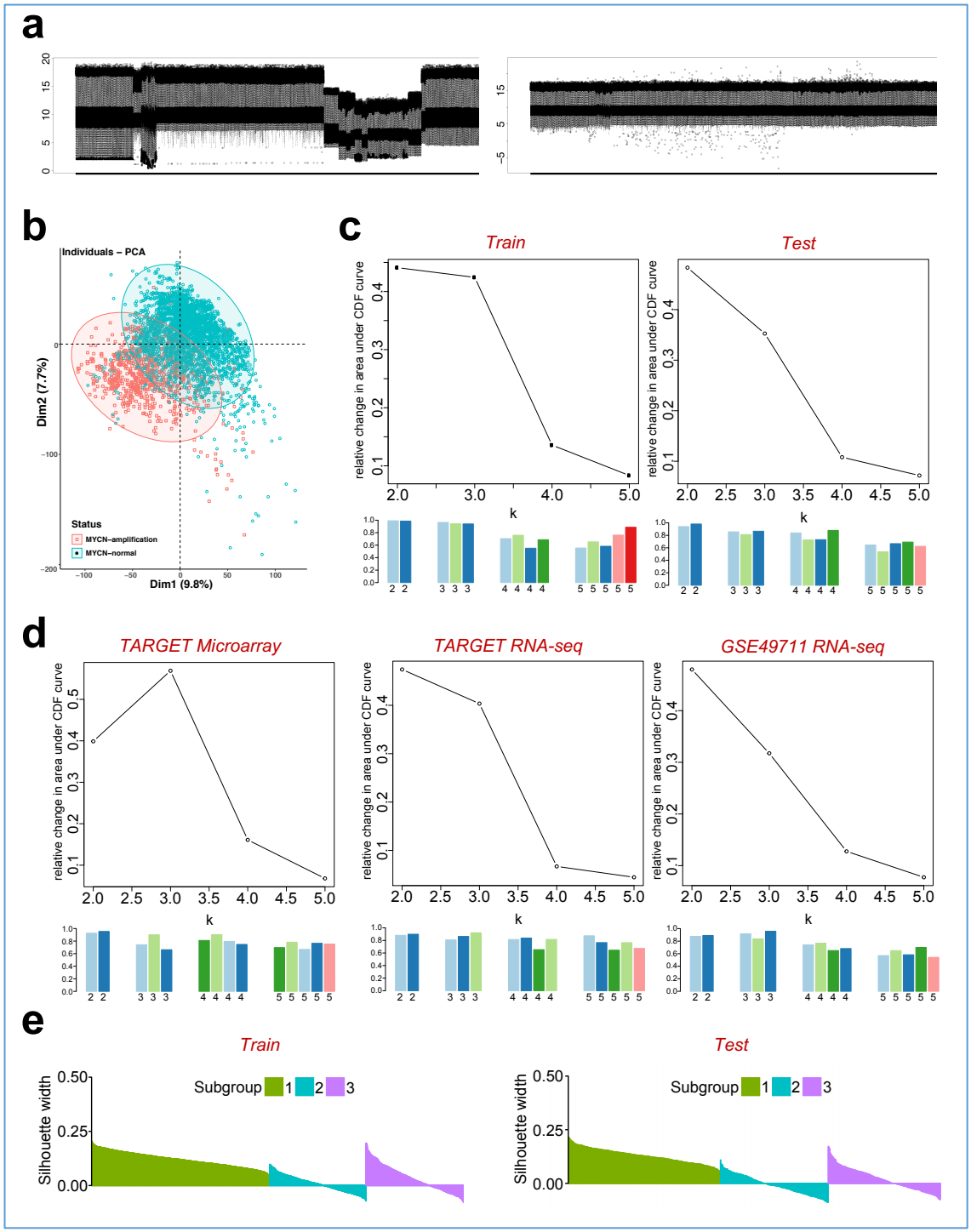
33. Sweeney TE, Shidham A, Wong HR, Khatri P. A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. *Sci Transl Med* **7**, 287ra271 (2015).

34. Su Z*, et al.* An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. *Genome Biol* **15**, 523 (2014).

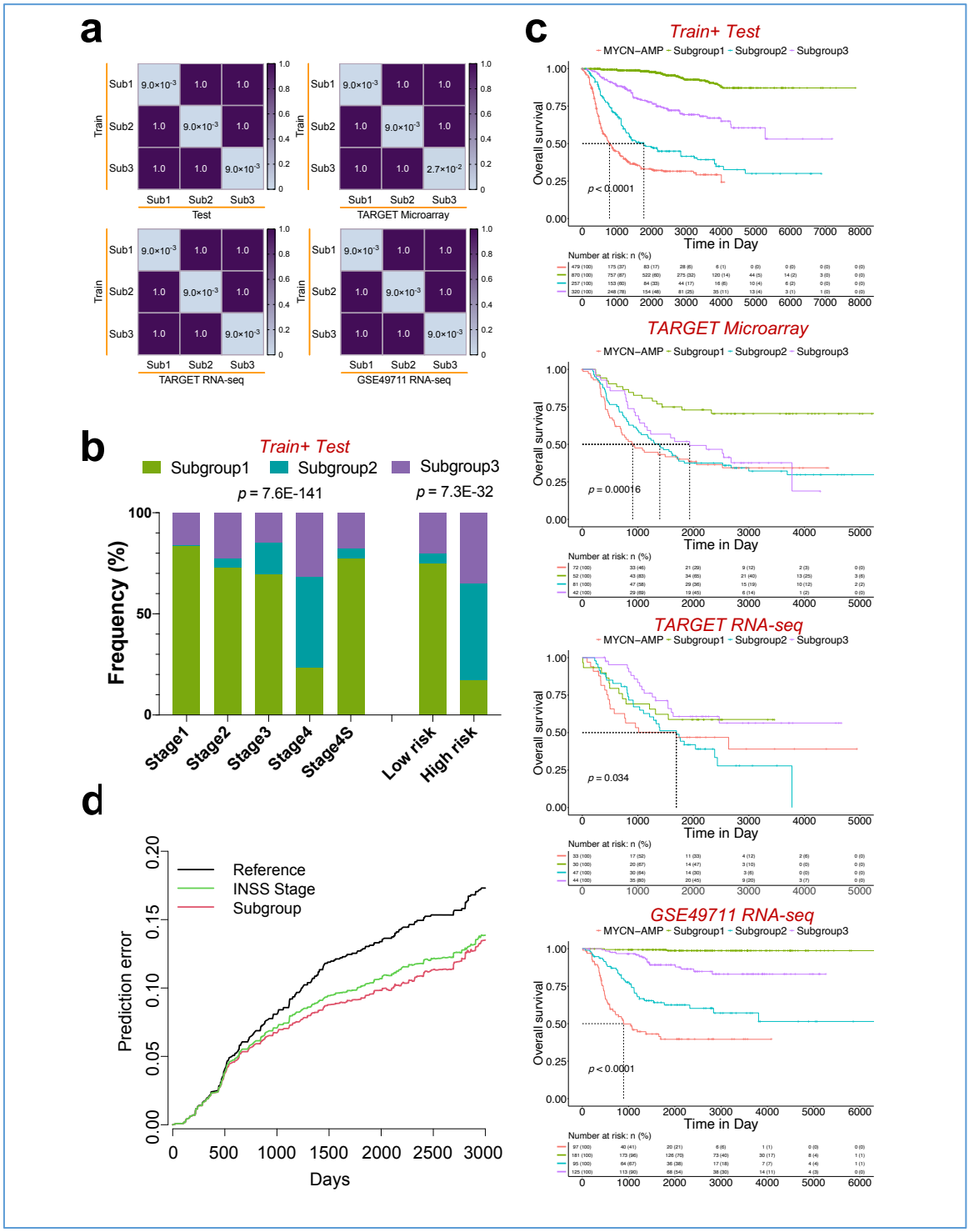
35. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-6(2021).

Supplementary Figures

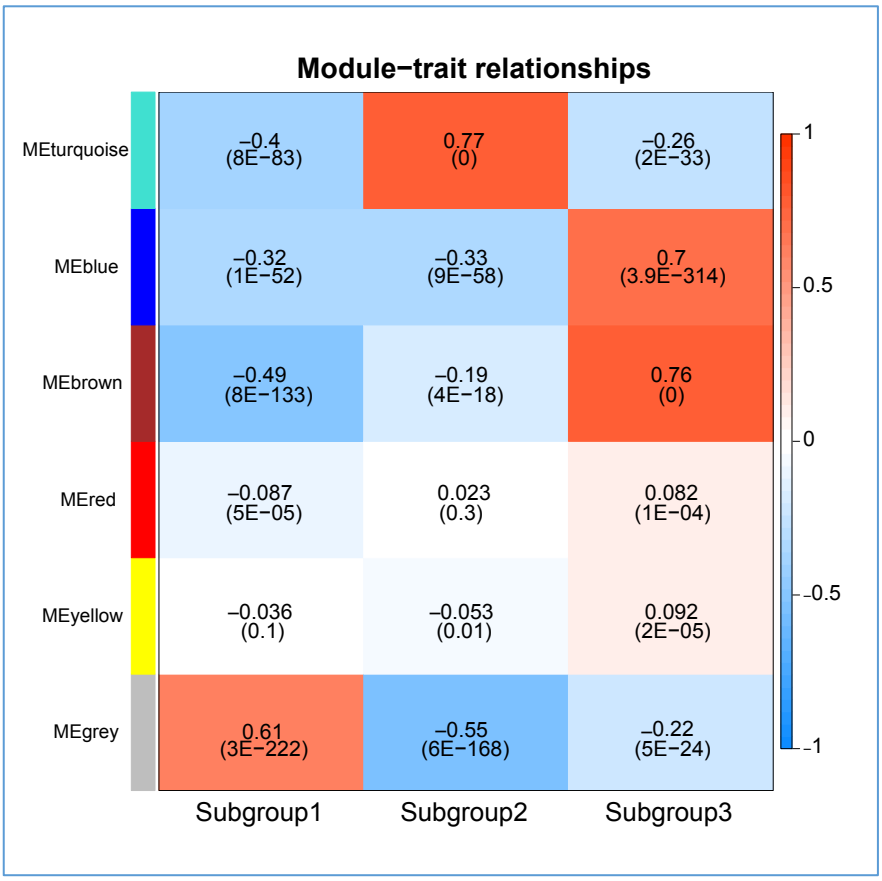
Supplementary Figure 1. Characterisation of molecular subtypes in *MYCN* non-amplified neuroblastomas. (a) Boxplot of merging neuroblastoma tumour sample from 18 datasets before and after removing batch effects. (b) Principal component analysis (PCA) of neuroblastoma patients after removing batch effect showed patients clustered according to *MYCN*-amplified status while there were overlapping areas between the two groups. (c, d) Relative area changes on the cumulative distribution function and cluster-consensus value of the train, test, TARGET microarray, TARGET RNA-seq and GSE49711 RNA-seq cohorts. (e) Silhouette plots of 3 subgroups in the train or test cohort. Silhouette score > 0 was considered as core samples.



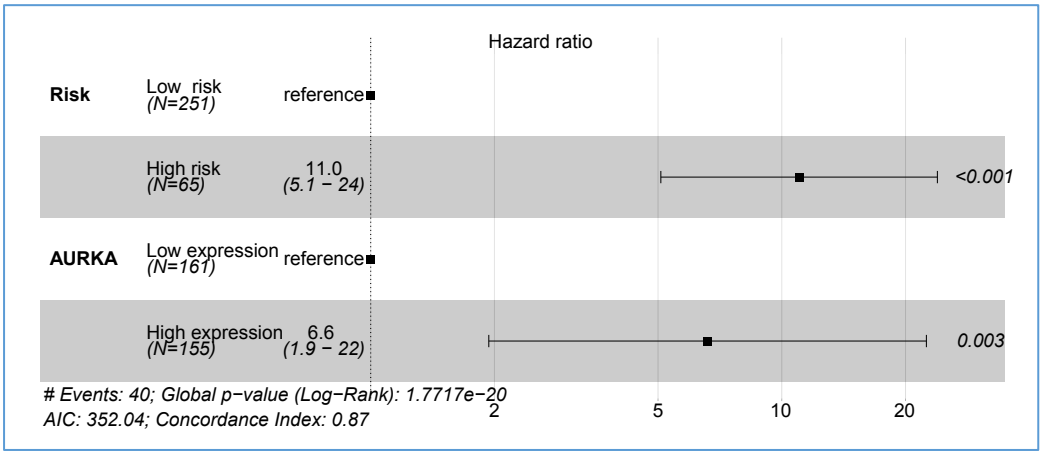
Supplementary Figure 2. Clinical characterisation of subtypes within *MYCN* non-amplified neuroblastomas identifies key distinguishing features. (a) Submap analysis of 3 subgroups in 4 cohorts. Bonferroni adjusted p-values were indicated. (b) Graphs showing the frequency (%) of each molecular subtype in different International Neuroblastoma Staging System (INSS) stages or risk status in train plus test cohort. P values are indicated. (c) Kaplan-Meier plots showing the overall survival in each molecular subtype or *MYCN*-amplification (*MYCN*-AMP) in train plus test cohort, TARGET microarray, TARGET RNA-seq or GSE49711 RNA-seq. Numbers below are n (%). P values are indicated. (d) Prediction error curves (indicating mean squared error in predicting survival status) are calculated for the subgroup (red) and (INSS) stages (green).



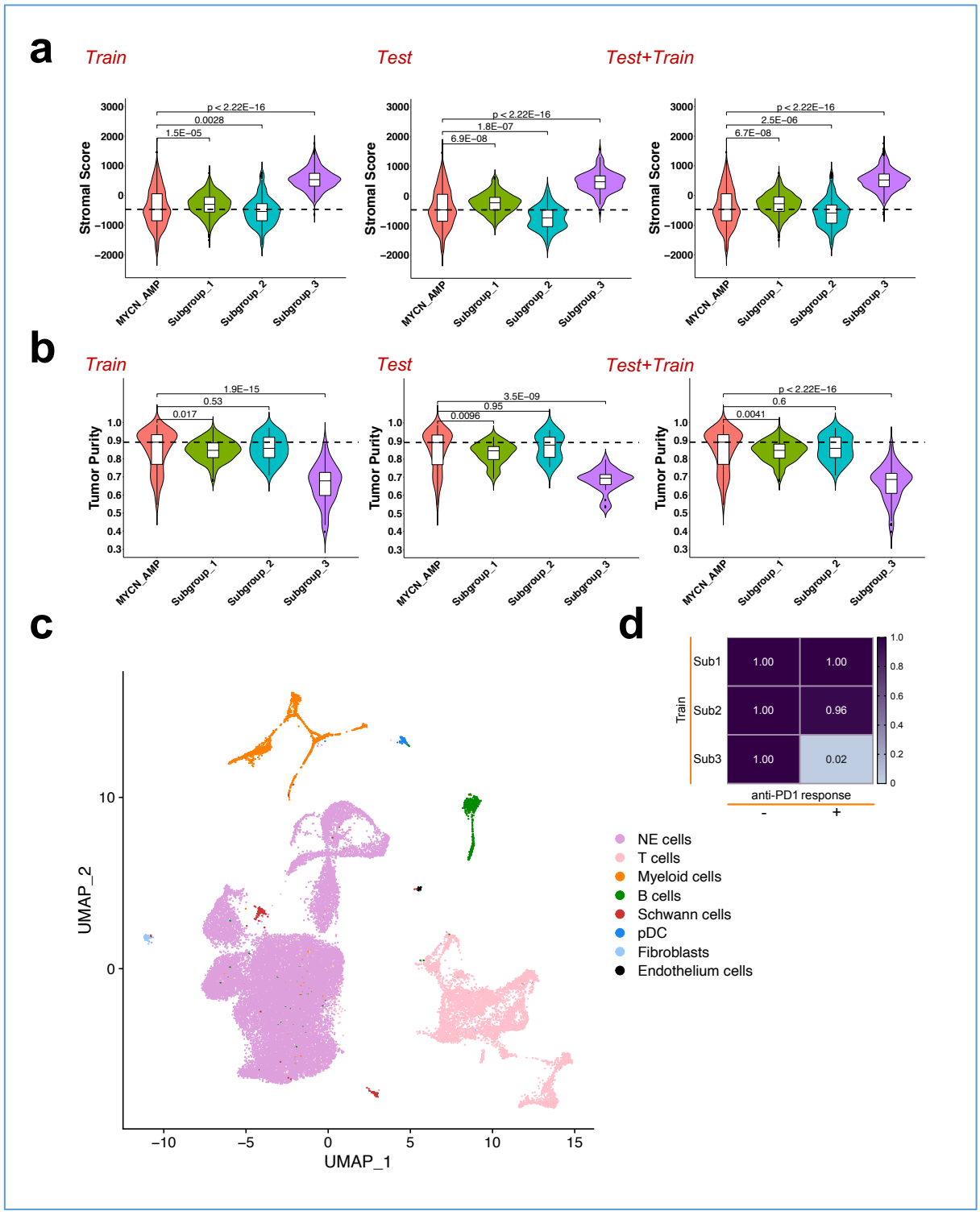
Supplementary Figure 3. Defining molecular features of 3 subtypes in *MYCN* non-amplified neuroblastomas. The correlation coefficients of WGCNA (weighted gene co-expression network analysis) modules and subgroups (red indicates positive correlated and blue negative correlated).



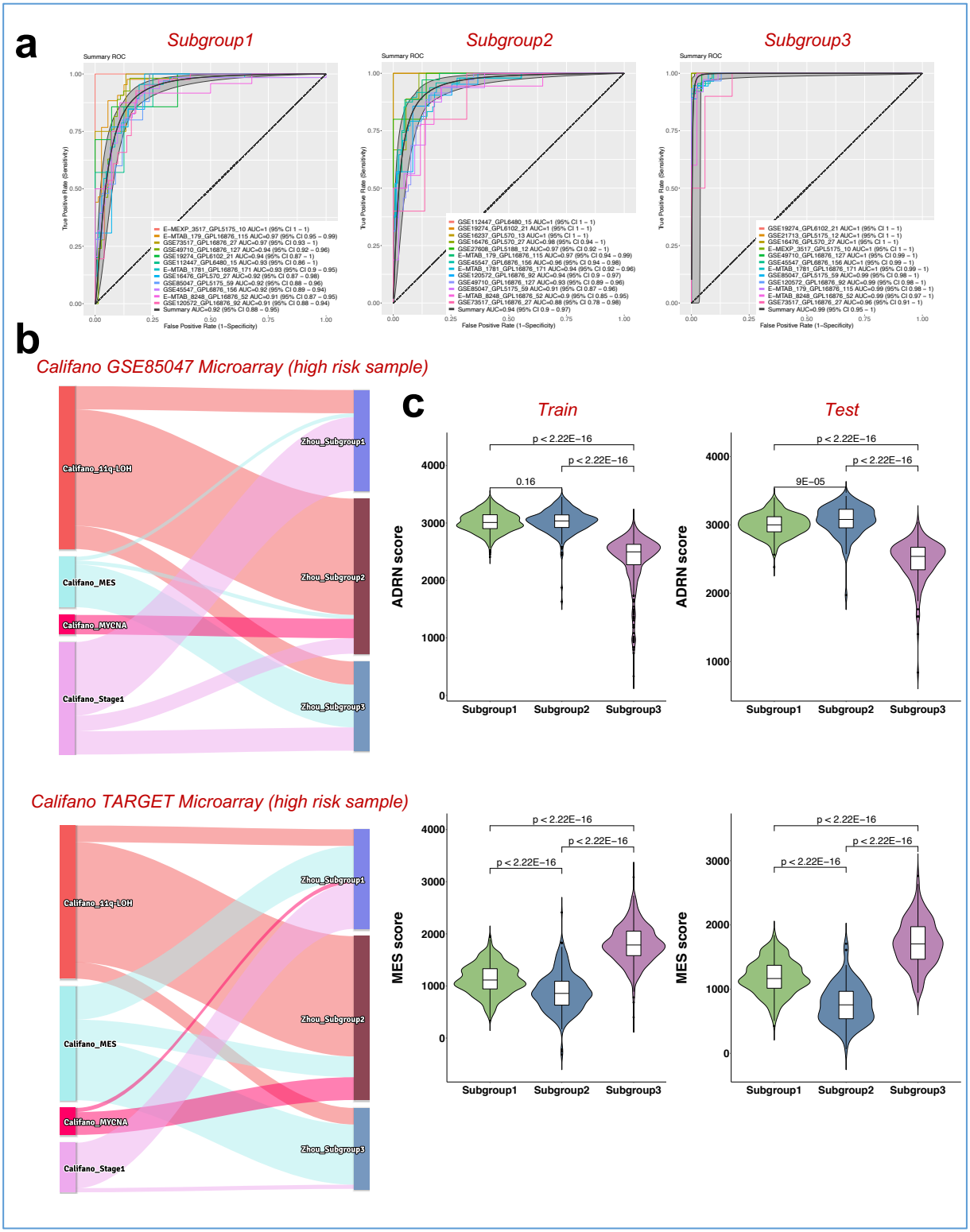
Supplementary Figure 4. Subgroup 2 shows a "*MYCN*" signature, potentially induced by Aurora Kinase A (*AURKA*) overexpression. Multivariate analysis of *AURKA* expression level and risk status in in *MYCN* non-amplified neuroblastomas. HR (hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown.



Supplementary Figure 5. Subgroup 3 is accompanied by an "inflamed" gene signature. (a,b) Violin plots showing stromal scores and tumour purity in different subgroups and *MYCN*-AMP in the train, test or train plus test cohort. (c) UMAP projection of *MYCN*-non amplified patients cells. The colors demonstrated the distinct cell types according to the established marker genes. (d) Submap analysis showing differential anti-PD1 immunotherapeutic response in 3 subgroups (GSE78220). Bonferroni adjusted p values indicated.



Supplementary Figure 6. Identification of independent predictors to subgroup patients within *MYCN* non-amplified neuroblastomas and evaluation of different patient stratification strategies. (a) AUROC (area under the receiver-operating characteristic curve) analysis in 3 subgroups from the test cohort. The average AUROC scores of each subgroup were 0.92, 0.94 and 0.99, respectively. (b) Prediction differences in GSE85047 or TARGET RNA-seq using subgrouping method from this report (named *Zhou*) or Califano and colleagues (*Califano*). (c) Violin plots showing ADRN and MES scores in different subgroups and *MYCN*-AMP in train and test cohort.



Supplementary Tables

Table S1. List of datasets collected for meta-analysis.

Table S2. List of top 50% variable genes for consensus clustering.

Table S3. Univariate and multivariate regression analysis in *MYCN* non-amplified neuroblastomas (n = 1,120).

Table S4. DEGs (differentially expressed genes) in subgroups.

Table S5. GSEA (gene set enrichment analysis) in subgroups.

Table S6. WGCNA (weighted gene co-expression network analysis) in subgroups.

Table S7. List of genes in PPI (protein–protein interaction) network analysis.

Table S8. List of 46 immune-related gene sets

Table S9. Analysis of clinically actionable genes and drug response.