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

Xiaoxiao Hu
Shanghai Jiaotong University
Yilu Zhou
University of Southampton
Charlotte Hill
University of Southampton
Kai Chen
Shanghai Jiaotong University
Yeming Wu
Xinhua Hospital, School of Medicine, Shanghai Jiaotong University
Zhongrong Li
Wenzhou Medical University
Cheng Cheng
Xinhua Hospital, School of Medicine, Shanghai Jiaotong University
Xiaowei Liu
Shanghai Jiaotong University
Peiwen Duan
Shanghai Jiaotong University
Rob Ewing
University of Southampton https://orcid.org/0000-0001-6510-4001
Yaoyao Gu
Shanghai Jiaotong University
Zhixiang Wu
Xinhua Hospital, School of Medicine, Shanghai Jiaotong University https://orcid.org/0000-0002-3947-2035
Yihua Wang (✉ yihua.wang@soton.ac.uk)
University of Southampton https://orcid.org/0000-0001-5561-0648

Article

Keywords: MYCN non-amplified neuroblastoma, tumour expression, prognosis, therapy

Posted Date: May 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-478917/v1

Abstract

Despite the extensive study of *MYCN*-amplified neuroblastomas, there is a significant unmet clinical need in *MYCN* non-amplified neuroblastomas. In particular, the extent of heterogeneity within the *MYCN* non-amplified population is unknown. Here, we investigate whether transcriptional subtyping of *MYCN* non-amplified neuroblastomas can identify molecular subtypes with discrete prognosis and therapeutic vulnerabilities. Using tumour expression data and ConsensusClusterPlus, we demonstrate that *MYCN* non-amplified neuroblastomas are heterogeneous and can be classified into 3 subgroups based on their transcriptional signatures. Within these groups, subgroup 2 has the worst prognosis and this group shows a "*MYCN*" signature that is potentially induced by the overexpression of Aurora Kinase A (AURKA); whilst subgroup 3 is characterised by an "inflamed" gene signature. The clinical implications of this subtype classification are significant, as each subtype demonstrates unique prognosis and vulnerability to investigational therapies. We propose that matching baseline tumour subtype to therapy may enhance precision prognosis and therapy stratification for patients with *MYCN* non-amplified neuroblastomas.

Xiaoxiao Hu and Yilu Zhou contributed equally to this work. Correspondence should be addressed to YW (e-mail: yihua.wang@soton.ac.uk) or ZW (e-mail: wuzhixiang@xinhuamed.com.cn)

Introduction

Neuroblastoma is the most common extra-cranial solid tumour in children, representing 6–10% of all childhood cancers[1]. It is an embryonic tumour arising from precursor cells in the sympathetic nervous system and adrenal medulla[2], with a median age of diagnosis of 18 months[3]. It can also present in the neck, chest, abdomen or pelvis[4]. Neuroblastoma is a highly heterogeneous disease, with clinical behavior ranging from spontaneous regression to drug resistance and metastasis ultimately resulting in death[5]. The prognosis of the disease is poor with a 5-year overall survival of approximately 20%, despite more aggressive therapies[6]. As a result, risk stratification and personalised treatment approaches in neuroblastomas are urgently needed.

The International Neuroblastoma Risk Group Staging System (INRGSS) defines the high-risk group to include patients with *MYCN* amplified tumours and patients > 18 months old with metastatic tumours[7]. N-MYC is a key regulator of transcription, which activates genes that affect cancer development. It is widely involved in various physiological and pathological processes of neuroblastoma including cell growth[8], apoptosis[9], differentiation[10], angiogenesis[11], tumour invasion and metastasis[12].

MYCN amplification was identified as the first independent prognostic factor indicating adverse clinical outcomes in neuroblastomas[13, 14], which is observed in approximately 20% of cases[15] and accounts for about 40% of high-risk neuroblastomas[16]. Despite the extensive study of *MYCN*-amplified

neuroblastomas, there is a significant unmet clinical need in *MYCN* non-amplified neuroblastomas. In particular, the extent of heterogeneity within *MYCN* non-amplified population is unknown.

Here, we investigated whether transcriptional subtyping of *MYCN* non-amplified neuroblastomas can identify molecular subtypes with discrete prognosis and therapeutic vulnerabilities. Our analysis suggested that *MYCN* non-amplified neuroblastomas were heterogeneous and could be classified into 3 subgroups based on their transcriptional profiling. Within them, subgroup 2 had the worst prognosis and this group had a "*MYCN*" signature that was potentially induced by the overexpression of Aurora Kinase A (AURKA); whilst subgroup 3 was accompanied by an "inflamed" gene signature. We propose that matching baseline tumour subtype to therapy may enhance precision prognosis and therapy stratification for patients with *MYCN* non-amplified neuroblastomas.

Results

Characterisation of molecular subtypes in *MYCN* non-amplified neuroblastomas.

A total of 3,799 samples from 18 datasets were identified in GEO (Gene Expression Omnibus) and ArrayExpress, in which 714 cases are with *MYCN* gene amplification (*MYCN*-AMP) and 3,085 cases *MYCN* non-amplified (*MYCN*-normal) (Fig. 1a; Supplementary Table 1). Following removal of batch effects (Supplementary Fig. 1a), two clear clusters corresponding to *MYCN*-AMP and *MYCN*-normal neuroblastomas, respectively, were visualised using principal component analysis (PCA) (Supplementary Fig. 1b). Samples in the *MYCN*-normal group (n = 3,085) were further randomly divided into a train and a test group with a 7:3 ratio, containing 2,160 and 925 cases, respectively.

In an unbiased attempt to identify subtypes within *MYCN* non-amplified neuroblastomas, we applied consensus clustering to both train and test groups based on 5,792 variable genes (top 50% median absolute deviation; Supplementary Table 2). As determined by the relative area under the cumulative distribution function and cluster-consensus scores, the optimal number of distinct clusters was 3 (Fig. 1b; Supplementary Fig. 1c), which was further confirmed in another 3 independent datasets (TARGET microarray, TARGET RNA-seq and GSE49711) (Supplementary Fig. 1d). To ensure precise clinical and molecular characterisations of subgroups, samples with silhouette width values less than 0 were labeled as not defined[17] and were excluded from the further analysis (Supplementary Fig. 1e; grey dots in Fig. 1c). In total, within the train group, subgroup 1 (green), 2 (blue) and 3 (purple) accounts for 60%, 18% and 22%, respectively (Fig. 1c). Cross-cohort analysis using an unsupervised method subMap (https://www.genepattern.org/modules) confirmed the robustness of this classification (Supplementary Fig. 2a; false discovery rate, FDR < 0.05).

Further clinical characterisation of these subtypes identified key distinguishing features. Patients within subgroup 2 were frequently observed in the advanced neuroblastomas according to the International Neuroblastoma Staging System (INSS) and in those defined as "high risk"[7] (Fig. 2a and b; Supplementary Fig. 2b). We then analysed their overall survival together with *MYCN*-AMP cases. Patients with *MYCN*

amplified had the worst prognosis (Fig. 2c and d; Supplementary Fig. 2c). Importantly, there was a high degree of variability for overall survival among *MYCN* non-amplified cases, in which subgroup 2 was associated with a poor prognosis, followed by subgroup 3; while patients within subgroup 1 had the most favorable outcomes. These observations were consistent in both train and test cohorts, and further 3 independent datasets (TARGET microarray, TARGET RNA-seq, and GSE49711) (Fig. 2c and d; Supplementary Fig. 2c). In addition, the molecular subtype classification was a strong independent predictor of mortality including in multivariate analysis with the risk classification that uses commonly measured clinical variables to predict mortality in neuroblastomas[7]. Using subgroup 1 as a reference, the hazard ratio (HR) and 95% confidence interval (CI) for subgroup 2 and 3 were 24.4 (5 ~ 120) and 9.7 (2 ~ 46), respectively (Fig. 2e). Similar results were obtained using univariate or multivariate cox regression analysis with age and INSS stages in *MYCN* non-amplified neuroblastomas (Supplementary Table 3). Impressively, the molecular subtype classification alone outperformed the risk classification in general (Fig. 2f) as well as INSS stages (Supplementary Fig. 2d).

Overall, subgroup 2 and subgroup 3 (to a lesser extent) were associated with poor survival in *MYCN* non-amplified neuroblastomas, suggesting fundamentally different mechanisms leading to an advanced disease.

Defining molecular features of 3 subtypes in *MYCN* non-amplified neuroblastomas.

Using the same 5,792 variable genes described above (Supplementary Table 2), we observed clear distinctions among these 3 subtypes in *MYCN* non-amplified neuroblastomas (Fig. 3a; Supplementary Tables 4). Intriguingly, subgroup 2 showed a similar signature to *MYCN*-AMP cases (Fig. 3a). This was in consistence with the Gene Set Enrichment Analysis (GSEA), showing HALLMARK_MYC_TARGETS_V1 and V2 significantly enriched in subgroup 2 (Fig. 3b; Supplementary Table 5; FDR (false discovery rate) < 0.001 and FDR = 0.009, respectively). In contrast, subgroup 3 exhibited an "inflamed" phenotype, with high expression of genes related to IL2_STAT5_SIGNALING, IL6_JAK_STAT3_SIGNALING, INTERFERON_ALPHA_RESPONSE and INTERFERON _GAMMA_RESPONSE (Fig. 3b; Supplementary Table 5; all FDR values less than 0.05). None of these pathways were enriched in subgroup 1 (Fig. 3b).

The above analysis was extended using weighted gene co-expression network analysis (WGCNA)[18]. Six molecular modules were identified (Supplementary Fig. 3; Supplementary Table 6) and were further used to construct a protein-protein network consisting of 1,699 gene and 4,197 edges (Fig. 3c; confidence score > 0.7). Molecular module MEturquoise, which was significantly correlated with subgroup 2 (Fig. 3d), was enriched for "Mitotic roles of pole-like kinase", "NER (nucleotide excision repair) pathway", "Cell cycle control of chromosomal replication" and "eNOS (endothelial nitric oxide synthase) signaling". In subgroup 3, there were 2 molecular modules, MEblue and MEbrown highly involved (Fig. 3d; Supplementary Table 6). Molecular module MEblue was enriched for signaling, such as "CDK5 (cyclin-dependent-like kinase 5)", "ILK (integrin-linked protein kinase)", "Neuregulin" and "HIF1α (hypoxia inducible factor 1α)'" whereas "PD-1, PD-L1 cancer immunotherapy", "IL-8 signaling", "Natural killer cell signaling pathway",

"Neuroinflammation", "Regulation of EMT (epithelial-mesenchymal transition)" and "Dendritic cell maturation" were significantly enriched in molecular module MEbrown.

Subgroup 2 shows a *"MYCN"* signature, potentially induced by Aurora Kinase A (AURKA) overexpression.

Our above analysis suggested that mechanism other than gene amplification induces N-MYC activity in subgroup 2. Indeed, the mRNA level of *MYCN* in subgroup 2 was significantly lower than cases within *MYCN*-AMP group (Fig. 4a; Supplementary Table 4; $p < 0.0001$). To evaluate N-MYC activity in neuroblastoma samples, a total of 87 genes upregulated by N-MYC were selected to classify its activity[19]. The *MYCN*-score for each sample was calculated using single-sample gene set enrichment analysis (ssGSEA) based on this 87-gene expression signature. *MYCN* scores in subgroup 2 were significantly higher than those in subgroups 1 and 3, and were comparable to those in *MYCN*-AMP group, although slightly lower (Fig. 4b). Moreover, the *MYCN* score was an independent predictor of mortality including in multivariate analysis with the risk classification (Fig. 4c; HR: 3.8; $p = 0.008$).

To investigate the potential mechanism that leads to higher *MYCN* scores in subgroup 2, correlation analysis coupled with protein protein interactions (PPI) network construction was performed (Fig. 4d; Supplementary Table 7). Among the candidate genes, *AURKA* (Aurora kinase A) was identified to interact with *MYCN*. AURKA, a serine/threonine kinase regulating the process of mitosis[20], was previously demonstrated to regulate N-MYC protein stability[21]. *AURKA* was expressed at significantly higher levels in subgroup 2 when compared to the other 2 subgroups and its levels were even slightly higher than those in the *MYCN*-AMP group (Fig. 4e). Classifying *MYCN* non-amplified neuroblastomas into high and low groups, we demonstrated that the *AURKA* mRNA levels alone could predict the overall survival (Fig. 4f; HR 5.7; $p < 0.0001$). In addition, the high level of *AURKA* was an independent predictor (HR 6.6, $p = 0.003$) of mortality including in multivariate analysis with the risk classification (Supplementary Fig. 4).

These findings were further investigated by immunohistochemistry (IHC) staining of N-MYC or AURKA in a custom neuroblastoma tissue microarray, which contains 94 *MYCN* non-amplified neuroblastomas. Within them, 22 samples were positive for N-MYC (Fig. 5a), and they had worse survival compared to those with N-MYC negative staining (n = 72) (Fig. 5b; $p = 0.03$). In parallel, patients with higher levels of AURKA had unfavorable survival outcomes (Fig. 5c and d; $p = 0.00014$). Moreover, a higher percentage of patients with high AURKA staining was observed in the N-MYC-positive group when compared to it in the N-MYC-negative group (Fig. 5e; 64% *vs.* 39%; $p = 0.041$).

Taken together, these results suggested that a "*MYCN*" signature in subgroup 2 is potentially induced by AURKA overexpression in *MYCN* non-amplified neuroblastomas.

Subgroup 3 is accompanied by an "inflamed" gene signature.

Considering immune-related pathways were enriched in subgroup 3, the activity of immune cells and pathways were further systematically explored. ssGSEA was performed to calculate enrichment scores of 46 immune gene sets summarised from two previous studies[22, 23], and subgroup 3 showed a

significantly higher activity of immune cells and pathways compared to the other 2 subtypes as well as *MYCN*-AMP group (Fig. 6a; Supplementary Table 8). Consistently, MHC-1 (major histocompatibility complex-1) or cytolytic activity (CYT) scores were highest in subgroup 3 (Fig. 6b and c). This was also true when using the ESTIMATE algorithm to evaluate the immune scores, stromal scores and tumour purity scores in neuroblastomas[24], showing highest immune and stromal scores, and lowest tumour purity scores in subgroup 3 (Fig. 6d; Supplementary Fig. 5a and b).

For a comprehensive assessment of immune cell infiltration, we used CIBERSORTx deconvolution[25] to quantify various immune populations based on a single cell RNA sequencing (scRNA-seq) dataset in *MYCN* non-amplified neuroblastoma[26] (Supplementary Fig. 5c). While similar immune cell types were present in each subtype, the absolute number of several immune cell populations were markedly increased in subgroup 3, including B cells, myeloid cells, T cells and pDC (plasmacytoid dendritic cells) (Fig. 6e). Finally, to investigate whether subgroup 3 would benefit more from immunotherapy than the other subgroups, we compared the expression matrix of 3 subgroups with published melanoma datasets including response information after treating with immunotherapies[27, 28].The subMap analysis highlighted that patients within subgroup 3 are predicted to respond to anti-PD1 immunotherapy (Fig. 6f; Supplementary Fig. 5d).

Taken together, these results demonstrated that subgroup 3 is accompanied by an "inflamed" gene signature, and is more likely to benefit from anti-PD1 therapies.

Integrative analysis of subgroup molecular features on 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 (see Supplementary Methods for details). These analyses generated 123 potential druggable targets (Supplementary Table 9). Of note, in subgroup 2, we identified 96 genes were that were overexpressed in subgroup 2 and showed high gene-dependency levels. The expression levels of these genes also negatively correlated (drug-sensitive) with the response to anticancer drugs. The top 15 genes in subgroup 2 were shown in Fig. 7a. These drugs target important pathways, including mitosis, metabolism, cell cycle and ERK/mitogen-activated protein kinase (MAPK) signalling (Fig. 7a; Supplementary Table 9). For example, *AURKA* was upregulated in subgroup 2, and *AURKA* expression negatively correlated (drug-sensitive) with the response to the cell cycle pathway inhibitor AZD7762 (correlation coefficient = − 0.57, p = 0.0054) and the NEDD8-activating enzyme (NAE) inhibitor Pevonedistat (correlation coefficient = − 0.60, p = 0.0031).

In subgroup 3, 8 genes were overexpressed and showed high gene-dependency levels (Fig. 7b). The expression levels of these genes also negatively correlated with the response to 18 anticancer drugs, which target pathways, such as WNT, protein stability and degradation, PI3K/mTOR signaling, DNA replication and EGFR signalling (Fig. 7b).

Together, these data suggest that each subgroup has unique therapeutic vulnerabilities, which emphasizes the importance of therapy stratification according to the molecular features.

Identification of independent predictors to subgroup patients within *MYCN* non-amplified neuroblastomas

To identify independent predictors for subgrouping, we applied a multi-cohort analysis pipeline via MetaIntegrator[29] (see Supplementary Methods). In total, 43 genes (Table 1) displayed the ability to predict different subgroups accurately (Supplementary Fig. 6a; area under the curve $(AUROC)_{Subgroup1}$ = 0.92; $AUROC_{Subgroup2}$ = 0.94 and $AUROC_{Subgroup3}$ = 0.99).

Table 1
Independent predictors to subgroup patients within *MYCN* non-amplified neuroblastomas.

Subgroup	Up	Down
Supgroup 1	*SCN2A, LONRF2, FRS3, CPEB4, SNAP25, PMP22, OLA1, C14orf132*	*TMEM109, LDHA, NDE1, PIM2, MRPL11, TNFRSF10B, COLEC12, TAF10, ELL, SIVA1*
Supgroup 2	*KIF4A, COX8A, UHRF1, ODC1, SNRPD1, CDCA4, SKP2, HMGB3, FANCI, CDK2;*	*PLA2G4C, NDEL1, ANXA2, PLSCR3 ;*
Supgroup 3	*IFITM2, ANXA11, STAT5A, FYCO1, LMNA, RARRES3, PLSCR4, SLC10A3, TMCO4, AGA*	*KBTBD7*

Furthermore, the machine learning classifier, support vector machine (SVM), was applied based on these 43 genes. SVM analysis showed good classifications across cohorts, with average AUROC of 0.931, 0.928, 0.759 and 0.954 in the test cohort, TARGET Microarray, TARGET RNA-seq and GSE49711, respectively (Fig. 8a). These independent predictors worked consistently between microarray and RNA-seq within GSE47792 (Fig. 8b).

Evaluation of different patient stratification strategies

Finally, we evaluated our subgrouping method (named Zhou_Subgroup) together with other reports. Recently, Westermann and colleagues reported 4 subgroups in neuroblastoma, including *MYCN*-amplified (MYCN), *MYCN* non-amplified high-risk (MNA-HR), *MYCN* non-amplified low-risk (MNA-LR) and mesenchymal (MES)[30]. With our method, patients within Westermann_MNA-HR can be further classified into 3 subtypes (Fig. 8c), showing different prognosis (Fig. 8d). This was also true for Westermann_ MNA-LR (Fig. 8c and d). A majority of cases in Westermann_ MES or MYCN belonged to subgroup 3 and 2, respectively (Fig. 8c).

Califano and colleagues classified high-risk neuroblastomas into 3 main subgroups ($MYCN^{Amp}$, MYCNA), 11q-LOH (loss of heterozygosity) and mesenchymal (MES)[31]. In comparisions, in the GSE85047 microarray, all cases of Califano_ MYCNA were classified in subgroup 2. Most cases in Califano_ MES or Stage1 belonged to subgroup 3 and 1 respectively (Supplementary Fig. 6b). Interestingly, most cases in

Califano_11q-LOH were classified in subgroup 2 (Supplementary Fig. 6b). Similar findings were observed in the TARGET microarray (Supplementary Fig. 6b).

van Groningen and colleagues reported that neuroblastoma is composed of two super-enhancer-associated differentiation states: an ‘ADRN’ subgroup showing up-regulated genes involved in adrenergic differentiation and an ‘MES’ subgroup with higher expressions of mesenchymal markers[32]. To quantify these characteristics, we calculated "ADRN" or "MES" scores of our subgroups based on a 369-gene "ADRN" signature or a 485-gene MES signature, respectively. We observed that subgroup 3 showed the highest "MES" scores and the lowest "ADRN" scores, consistent with our above findings; while subgroup 1 and 2 had the highest "ADRN" scores with the lowest "MES" scores in subgroup 2 (Supplementary Fig. 6c).

Together with other reports, our findings emphasised the extent of inner heterogeneity within *MYCN* non-amplified population and the importance of patient stratification.

Discussion

Neuroblastoma remains a challenge in the era of personalised therapy, largely due to inter- and intra-tumoural heterogeneity. Gene amplification in *MYCN* is the first genetic marker that indicates a highly invasive, advanced neuroblastoma, which has been observed in about 20% of primary and about 40% of high-risk neuroblastoma cases[33]. Despite the extensive study of *MYCN*-amplified neuroblastomas, there is a significant unmet clinical need in *MYCN* non-amplified neuroblastomas.

In this study, using tumour expression data and ConsensusClusterPlus, we demonstrate that *MYCN* non-amplified neuroblastomas are heterogeneous and can be further classified into 3 subgroups based on their transcriptional profiling. Within them, subgroup 2 has the worst prognosis and this group exhibits a "*MYCN*" signature that is potentially induced by the overexpression of AURKA. AURKA interacts with both N-MYC and SCF (Fbxw7) ubiquitin ligase, which ubiquitinates N-MYC for degradation. Consequently, overexpression of AURKA counteracts the degradation of N-MYC, leading to the growth of neuroblastoma cells[21, 34]. Interestingly, the majority of samples defined as "11q-LOH" subtype by Califano and colleagues[31] were classified into subgroup 2. 11q-LOH is one of the common genetic alterations of neuroblastoma, which is mainly found in *MYCN* non-amplified tumours[35] and has been included in the International Neuroblastoma Risk Group (INRG) risk classification system[7]. A previous study reported poor prognosis in patients within both the 11q-LOH group and the *MYCN* amplification group[36]. Our findings raise a potential of events within 11q-LOH leading to the activation of N-MYC signaling and the underlying mechanisms merit further investigation.

Subgroup 3 is accompanied by EMT and an "inflamed" phenotype, with high expression of genes related to IL2_STAT5 signaling, IL6 JAK STAT3 signaling, interferon-α activation, interferon-γ activation, and inflammatory response, consistent with the association between EMT and immune-related gene expression[37, 38]. Consistently, several groups reported a sub-population within neuroblastomas

characterized as a mesenchymal (MES) subtype[31, 32]. When evaluating their patient stratification strategies together with the subgrouping method in this report, we found that most cases classified as "MES" align to subgroup 3 as expected. The findings were further confirmed with using CIBERSORTx deconvolution[25] to quantify various immune populations based upon a *MYCN* non-amplified neuroblastoma scRNA-seq data[26], showing increased percentages of fibroblasts, B cells, myeloid cells, T cells and pDC (plasmacytoid dendritic cells).

The clinical implications of this subtype classification are significant, as each subtype demonstrates unique prognosis and vulnerability to investigational therapies. For example, patients in subgroup 1 show the most favorable prognosis with a long-term survival rate above 85%, despite some of them are clinically classified as INSS stage IV or high risk. As a result, without proper subtype classification, these patients tend be over-treated. With regard to therapy stratification, evidence showing significantly high MHC-I and CYT scores in subgroup 3 suggests that patients within this group may benefit from immunotherapy. Our analysis suggests that subgroup 3 is predicted to respond to anti-PD1 immunotherapy. The application of immunotherapy in neuroblastoma has started with treatments such as GD2 monoclonal antibody (dinutuximab) and Chimeric antigen receptor T cells (CAR-T) therapy[39, 40]. Further studies, including in *vitro*, in *vivo* and clinical validations, are required to investigate if patients within subgroup 3 can benefit from immunotherapy.

In addition, using integrative analysis based on subgroup molecular features on drug responses, several potential drugs and targets are suggested for patients within subgroup 2 or 3. For example, patients within subgroup 2 may benefit from AURKA inhibitors that can disrupt the interaction between AURKA and N-MYC. Indeed, AURKA inhibitors, MLN8054 and MLN8237(Alisertib), have been shown to be able to disrupt this interaction, leading to N-MYC degradation and subsequently cell death and differentiation in neuroblastoma cells[34, 41]. MLN8237 (Alisertib) is currently under phase 2 clinical evaluation in neuroblastoma (NCT01154816).

With the establishment of independent predictors, *MYCN* non-amplified neuroblastomas were easily classified into one of the 3 subtypes, permitting a realistic scenario in which prospective subtyping is performed in a cohort, wherein patients are assigned to different therapeutics (e.g., subgroup 3 to immunotherapy, subgroup 2 to AURKA inhibitors) on the basis of their subtype. If any one of these predictions demonstrated significant benefit, it would represent the first standard-of-care molecular biomarker selection for *MYCN* non-amplified neuroblastomas and a foundational step toward personalised therapy for this devastating disease.

Methods

Subtype identification

The study design is provided in Fig. 1a with a summary of datasets in Supplementary Table S1. Detailed description of the approach and further characterisation of the subtypes by principal component analysis

(PCA), ConsensusClusterPlus, single-sample Gene Set Enrichment Analysis (ssGSEA), weighted gene co-expression network analysis (WGCNA) and CIBERSORTx analysis is provided in the Supplementary Methods.

Analysis of hazard ratio and overall survival

The univariate and multivariate Cox proportional hazards model assessed the hazard ratio of each parameter through the survminer (v0.4.9). We performed log-rank test to compare Kaplan-Meier survival curves between each subgroup by survival (v3.2-10). Prediction error curves of each prognostic model were generated from pec (v2019.11.03)[42].

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. Detailed description of the approach is provided in Supplementary Methods.

Identification of independent predictors

To identify independent predictors for subgrouping, we applied a multi-cohort analysis pipeline via MetaIntegrator[29] and validated with the machine learning classifier, support vector machine (SVM) (see details in Supplementary Methods).

Tissue microarray (TMA) preparation and immunohistochemistry (IHC)

Separate a small part of the tissue specimen and shape it in a customized mold for chip production and fix overnight in 4% paraformaldehyde (PFA). Tissue blocks were embedded in paraffin in a prepared array. Then the sample was sliced (5 μm) and adhered to a poly-L-lysine coated glass slide for immunohistochemical staining, which was performed as previously described[43, 44], using specific antibody against N-MYC (1:600 dilutions; Cell Signaling Technology 51705) and Aurora kinase A (1:200 dilutions; Abcam ab52973). In a blind manner, with no knowledge of the clinicopathological characteristics of the tumour, the immunoreactivity in tissue sections was observed under three microscopes at random, and then evaluated by three pathologists. Differences in scoring were discussed until consensus was reached. The tissue sections were then scored under an optical microscope according to the degree of staining (0 ~ 3 points were negative staining, light yellow, light brown, dark brown) and the positive range (1 ~ 4 points were 0 ~ 25%, 26 ~ 50%, 51 ~ 75%, 76 ~ 100%). Finally, samples were divided into a high expression group and a low expression group based on the median of the final staining score. All procedures were in accordance with the ethical standards of the Clinical Committee of Xinhua Hospital, Shanghai Jiao Tong University School of Medicine (Approval No: XHEC-D-2016-037).

Code availability

Codes were implemented in R and have been deposited in GitHub: https://github.com/yz3n18/neuroblastoma.\

Data availability

All data supporting the findings of the current study are listed in Supplementary Materials.

Declarations

Conflict of interest

The authors declare that they have no relevant conflict of interest.

Acknowledgements

This project was supported by Medical Research Council (MR/S025480/1) [YWang], the Natural Science Foundation of China (No. 81874234), Shanghai "Rising Stars of Medical Talent" Youth Development Program–Outstanding Youth Medical Talents [ZW] and the Suzhou Clinical Medicine Innovation Team Introduction Project (SZYJTD201706) [YWu]. YZ was supported by an Institute for Life Sciences PhD Studentship. CH was supported by Gerald Kerkut Charitable Trust and University of Southampton Central VC Scholarship Scheme. We thank Dr Mark G. Jones for critical reading.

References

1. Stiller CA, Parkin DM. International variations in the incidence of neuroblastoma. *Int J Cancer* **52**, 538–543 (1992).
2. Tsubota S, Kadomatsu K. Origin and initiation mechanisms of neuroblastoma. *Cell Tissue Res* **372**, 211–221 (2018).
3. London WB, *et al.* Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children's Oncology Group. *J Clin Oncol* **23**, 6459–6465 (2005).
4. Song X, Huang C, Wang S, Yan L, Wang J, Li Y. Neck management in patients with olfactory neuroblastoma. *Oral Oncol* **101**, 104505 (2020).
5. Boeva V, *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat Genet* **49**, 1408–1413 (2017).
6. London WB, *et al.* Clinical and biologic features predictive of survival after relapse of neuroblastoma: a report from the International Neuroblastoma Risk Group project. *J Clin Oncol* **29**, 3286–3292 (2011).
7. Cohn SL, *et al.* The International Neuroblastoma Risk Group (INRG) classification system: an INRG Task Force report. *J Clin Oncol* **27**, 289–297 (2009).
8. Bell E, Lunec J, Tweddle DA. Cell cycle regulation targets of MYCN identified by gene expression microarrays. *Cell Cycle* **6**, 1249–1256 (2007).
9. Kang JH, Rychahou PG, Ishola TA, Qiao J, Evers BM, Chung DH. MYCN silencing induces differentiation and apoptosis in human neuroblastoma cells. *Biochem Biophys Res Commun* **351**, 192–197 (2006).

10. Wakamatsu Y, Watanabe Y, Nakamura H, Kondoh H. Regulation of the neural crest cell fate by N-myc: promotion of ventral migration and neuronal differentiation. *Development* **124**, 1953–1962 (1997).
11. Meitar D, Crawford SE, Rademaker AW, Cohn SL. Tumour angiogenesis correlates with metastatic disease, N-myc amplification, and poor outcome in human neuroblastoma. *J Clin Oncol* **14**, 405–414 (1996).
12. Goodman LA, *et al.* Modulation of N-myc expression alters the invasiveness of neuroblastoma. *Clin Exp Metastasis* **15**, 130–139 (1997).
13. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science* **224**, 1121–1124 (1984).
14. Brodeur GM, Seeger RC, Schwab M, Varmus HE, Bishop JM. Amplification of N-myc sequences in primary human neuroblastomas: correlation with advanced disease stage. *Prog Clin Biol Res* **175**, 105–113 (1985).
15. Look AT, *et al.* Clinical relevance of tumor cell ploidy and N-myc gene amplification in childhood neuroblastoma: a Pediatric Oncology Group study. *J Clin Oncol* **9**, 581–591 (1991).
16. Colon NC, Chung DH. Neuroblastoma. *Adv Pediatr* **58**, 297–311 (2011).
17. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2(2021).
18. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
19. 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).
20. Romain C, Paul P, Kim KW, Lee S, Qiao J, Chung DH. Targeting Aurora kinase-A downregulates cell proliferation and angiogenesis in neuroblastoma. *J Pediatr Surg* **49**, 159–165 (2014).
21. Otto T, *et al.* Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. *Cancer Cell* **15**, 67–78 (2009).
22. Jin W, Zhang Y, Liu Z, Che Z, Gao M, Peng H. Exploration of the molecular characteristics of the tumor-immune interaction and the development of an individualized immune prognostic signature for neuroblastoma. *J Cell Physiol* **236**, 294–308 (2021).
23. Li Y, *et al.* Pan-cancer characterization of immune-related lncRNAs identifies potential oncogenic biomarkers. *Nat Commun* **11**, 1000 (2020).
24. Yoshihara K, *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612 (2013).
25. Newman AM, *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453–457 (2015).
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. 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).
28. Hugo W, *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).
29. Haynes WA, *et al.* Empowering multi-cohort gene expression analysis to increase reproducibility. *Pac Symp Biocomput* **22**, 144–153 (2017).
30. Gartlgruber M, *et al.* Super enhancers define regulatory subtypes and cell identity in neuroblastoma. *Nature Cancer* **2**, 114–128 (2020).
31. Rajbhandari P, *et al.* Cross-Cohort Analysis Identifies a TEAD4-MYCN Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma. *Cancer Discov* **8**, 582–599 (2018).
32. van Groningen T, *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* **49**, 1261–1266 (2017).
33. Seeger RC, *et al.* Association of multiple copies of the N-myc oncogene with rapid progression of neuroblastomas. *N Engl J Med* **313**, 1111–1116 (1985).
34. Brockmann M, *et al.* Small molecule inhibitors of aurora-a induce proteasomal degradation of N-myc in childhood neuroblastoma. *Cancer Cell* **24**, 75–89 (2013).
35. Luttikhuis ME, *et al.* Neuroblastomas with chromosome 11q loss and single copy MYCN comprise a biologically distinct group of tumours with adverse prognosis. *Br J Cancer* **85**, 531–537 (2001).
36. Carén H, *et al.* High-risk neuroblastoma tumors with 11q-deletion display a poor prognostic, chromosome instability phenotype with later onset. *Proc Natl Acad Sci U S A* **107**, 4323–4328 (2010).
37. Terry S, *et al.* New insights into the role of EMT in tumor immune escape. *Mol Oncol* **11**, 824–846 (2017).
38. Dongre A, *et al.* Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas. *Cancer Res* **77**, 3982–3989 (2017).
39. Mujoo K, Cheresh DA, Yang HM, Reisfeld RA. Disialoganglioside GD2 on human neuroblastoma cells: target antigen for monoclonal antibody-mediated cytolysis and suppression of tumor growth. *Cancer Res* **47**, 1098–1104 (1987).
40. Richards RM, Sotillo E, Majzner RG. CAR T Cell Therapy for Neuroblastoma. *Front Immunol* **9**, 2380 (2018).
41. Yang Y, *et al.* Silencing of AURKA augments the antitumor efficacy of the AURKA inhibitor MLN8237 on neuroblastoma cells. *Cancer Cell Int* **20**, 9 (2020).
42. Mogensen UB, Ishwaran H, Gerds TA. Evaluating Random Forests for Survival Analysis using Prediction Error Curves. *J Stat Softw* **50**, 1–23 (2012).
43. Wang Y, *et al.* Annexin A2 could enhance multidrug resistance by regulating NF-κB signaling pathway in pediatric neuroblastoma. *J Exp Clin Cancer Res* **36**, 111 (2017).

44. Gu Y, *et al.* The deubiquitinating enzyme UCHL1 is a favorable prognostic marker in neuroblastoma as it promotes neuronal differentiation. *J Exp Clin Cancer Res* **37**, 258 (2018).

Figures

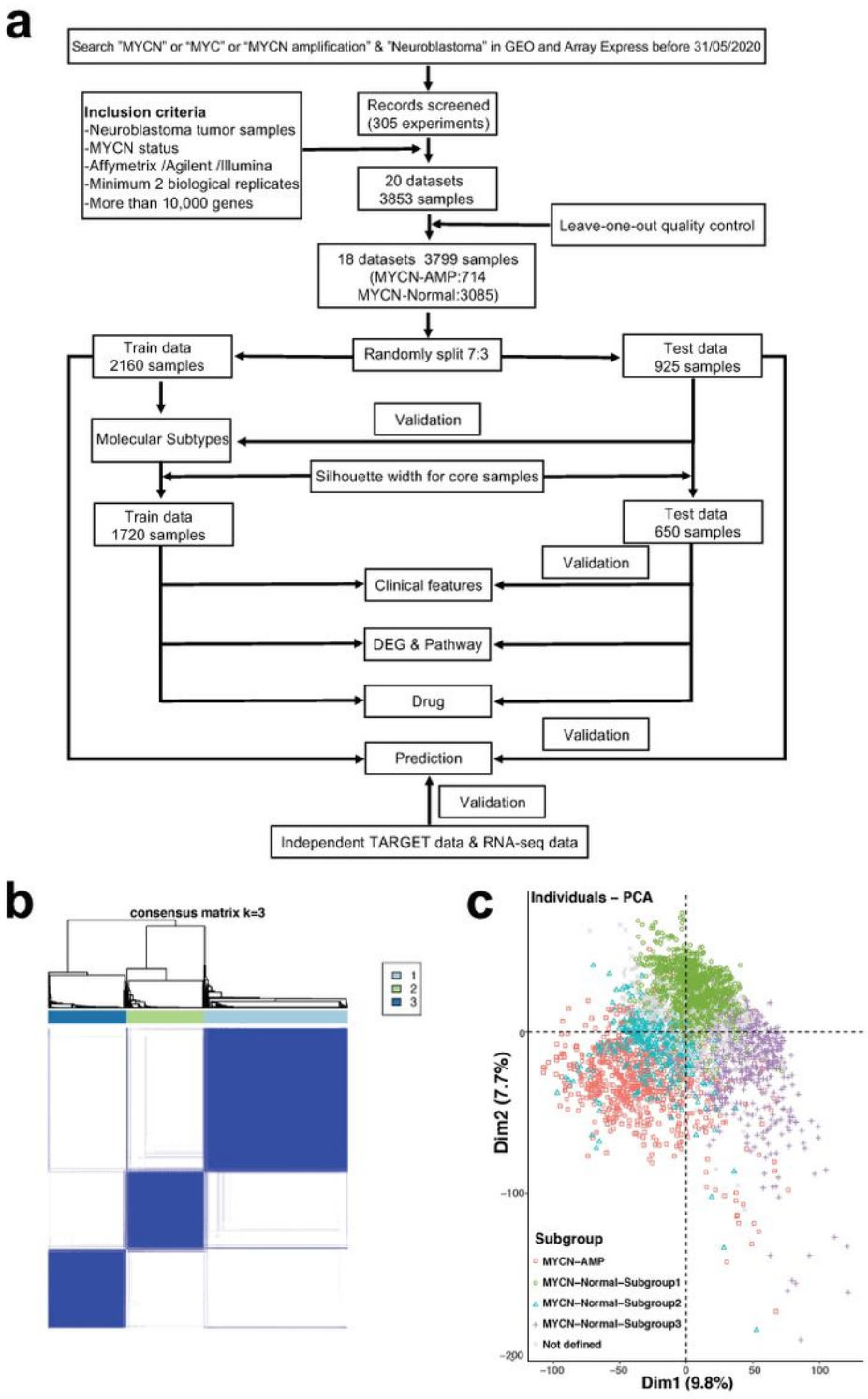

Figure 1

Characterisation of molecular subtypes in MYCN non-amplified neuroblastomas. (a) Workflow showing the study design (details provided in Supplementary Methods). (b) Consensus clustering of top 50% variable genes of train cohort. (c) Principal component analysis (PCA) showing neuroblastoma patients with subgroup annotations.

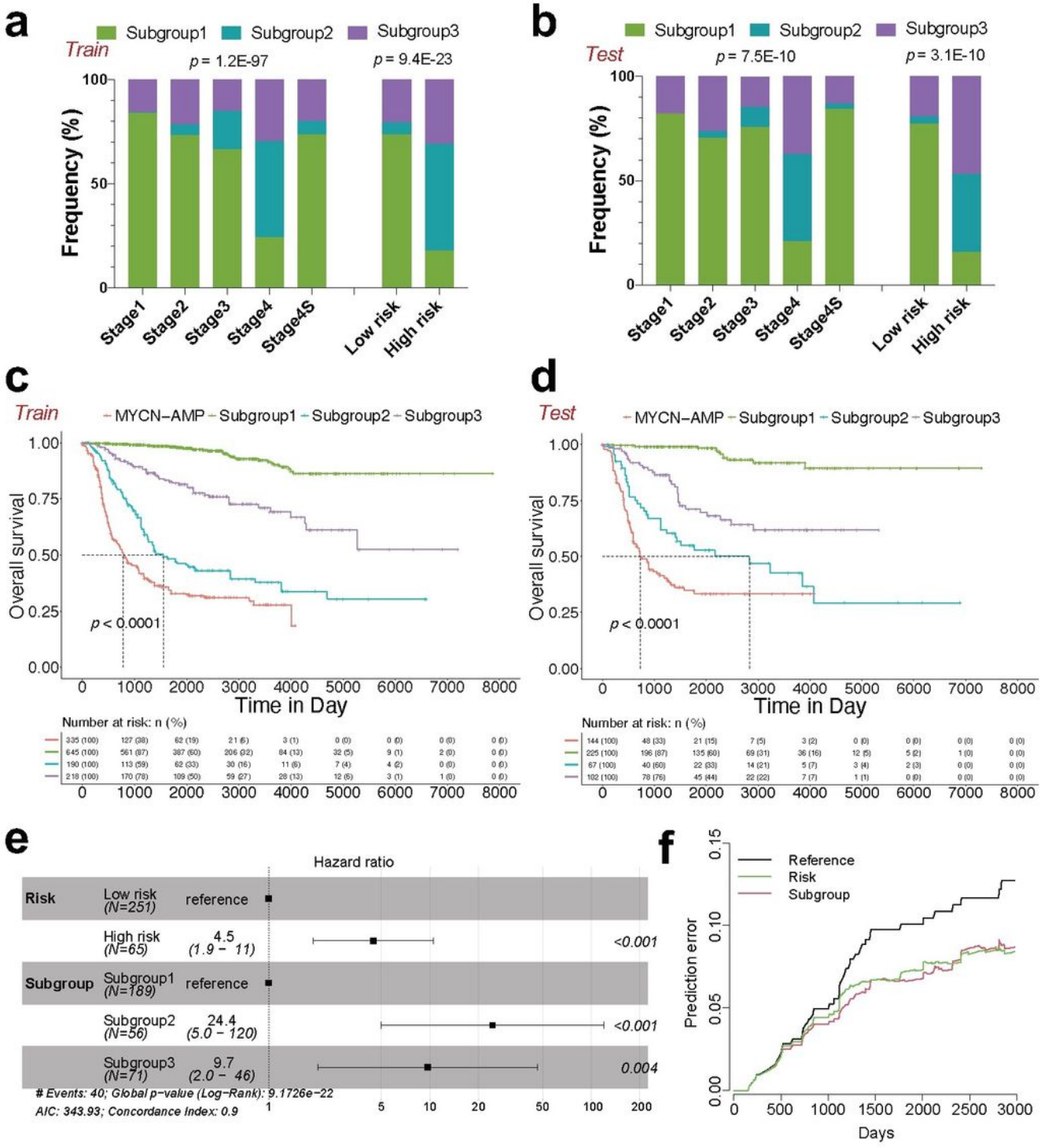

Figure 2

Clinical characterisation of subtypes within MYCN non-amplified neuroblastomas identifies key distinguishing features. Graphs showing the frequency (%) of each molecular subtype in different International Neuroblastoma Staging System (INSS) stages or risk status in either train (a) or test (b) cohort. P values are indicated. Kaplan-Meier plots showing the overall survival in each molecular subtype or MYCN-amplification (MYCN-AMP) in either train (c) or test (d) cohort. Numbers below are n (%). P values are indicated. (e) Multivariate analysis of subgroup classification with risk status in MYCN non-amplified neuroblastomas. HR (hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown. (f) Prediction error curves (indicating mean squared error in predicting survival status) are calculated for the subgroup (red) and risk status (green).

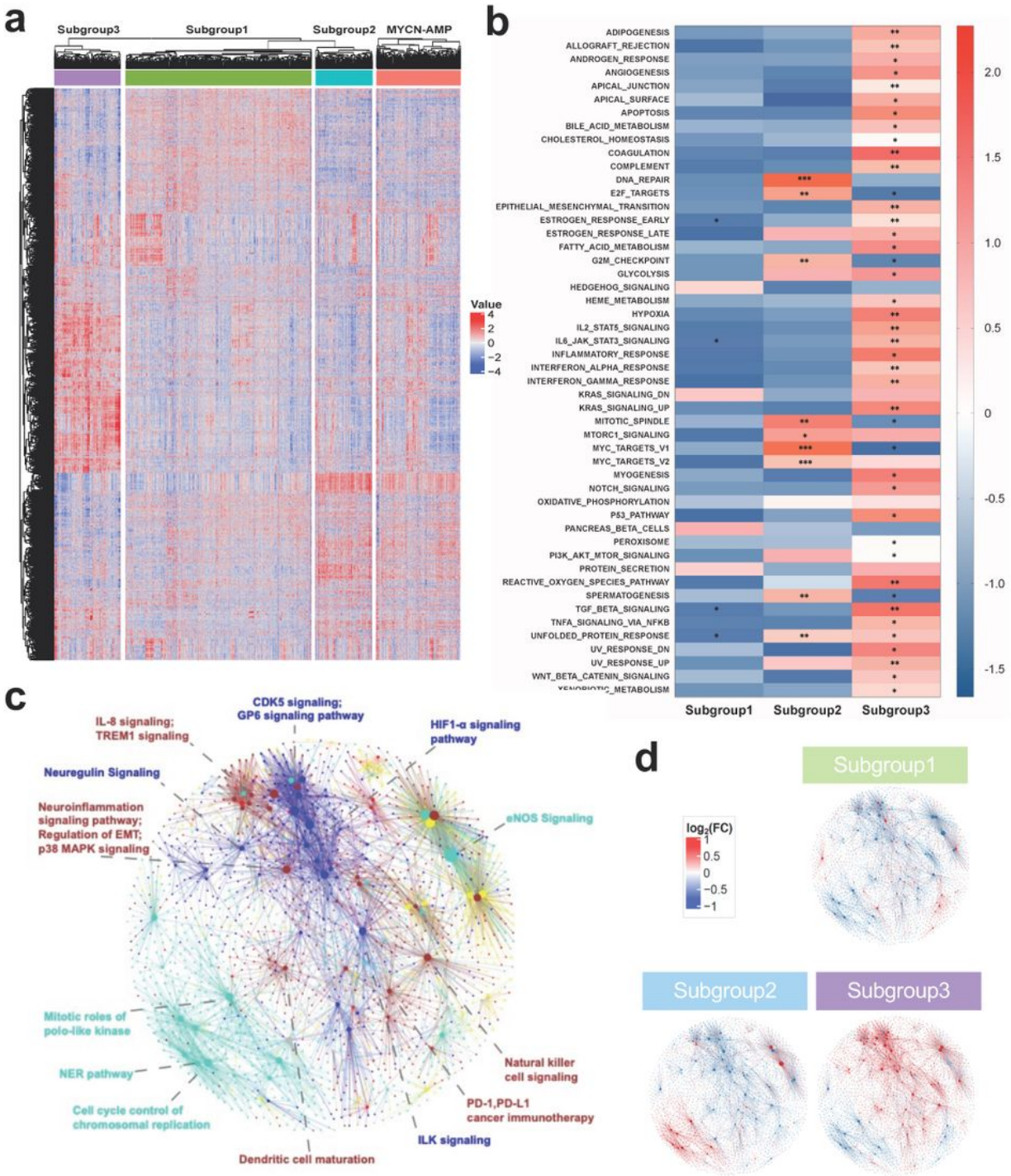

Figure 3

Defining molecular features of 3 subtypes in MYCN non-amplified neuroblastomas. (a) Heatmap showing differential expression of selected genes. Red indicates up-regulation and blue for down-regulation. Color bars show subgroup information. (b) Gene set enrichment analysis (GSEA) in 3 subtypes. *FDR (false discovery rate) < 0.25; **FDR < 0.05; ***FDR < 0.01. (c) Weighted gene co-expression network analysis (WGCNA) showing 6 molecular modules. Nodes are color coded according to the WGCNA modules.

Representative enriched pathway terms are indicated. (d) Overlay of the median-centered log2 fold change values per subgroup on the network.

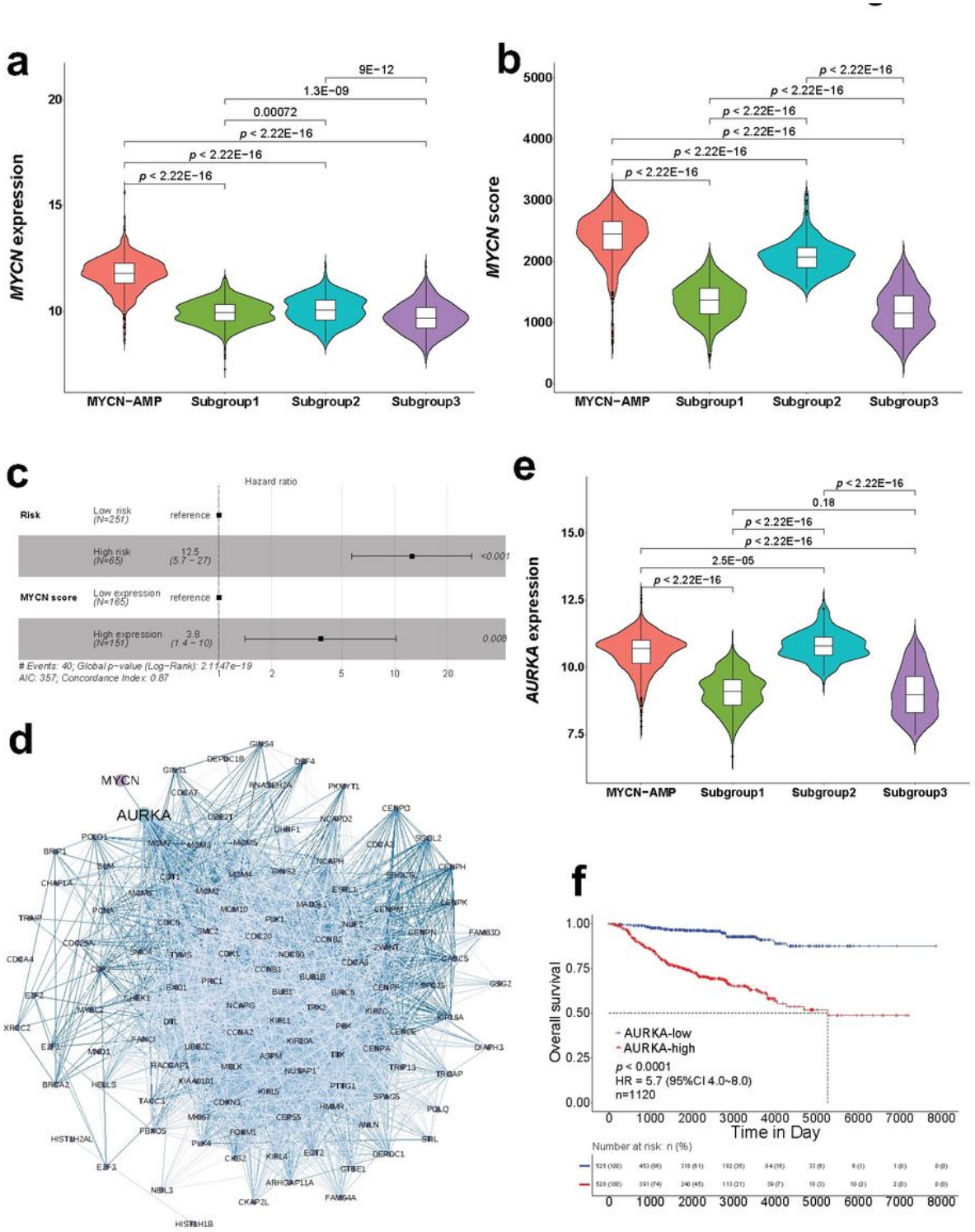

Figure 4

Subgroup 2 shows a "MYCN" signature, potentially induced by Aurora Kinase A overexpression. Violin plots showing MYCN mRNA levels (a) or MYCN scores (b) in neuroblastomas. P values are indicated. (c) Multivariate analysis of MYCN-score and risk status in in MYCN non-amplified neuroblastomas. HR

(hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown. (d) Protein protein interaction (PPI) network showing an interaction between AURKA and MYCN. (e) Violin plot showing AURKA mRNA levels in neuroblastomas. P values are indicated. (f) Kaplan-Meier plot showing the overall survival in samples with low vs. high AURKA expression. Numbers below are n (%). HR (hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown.

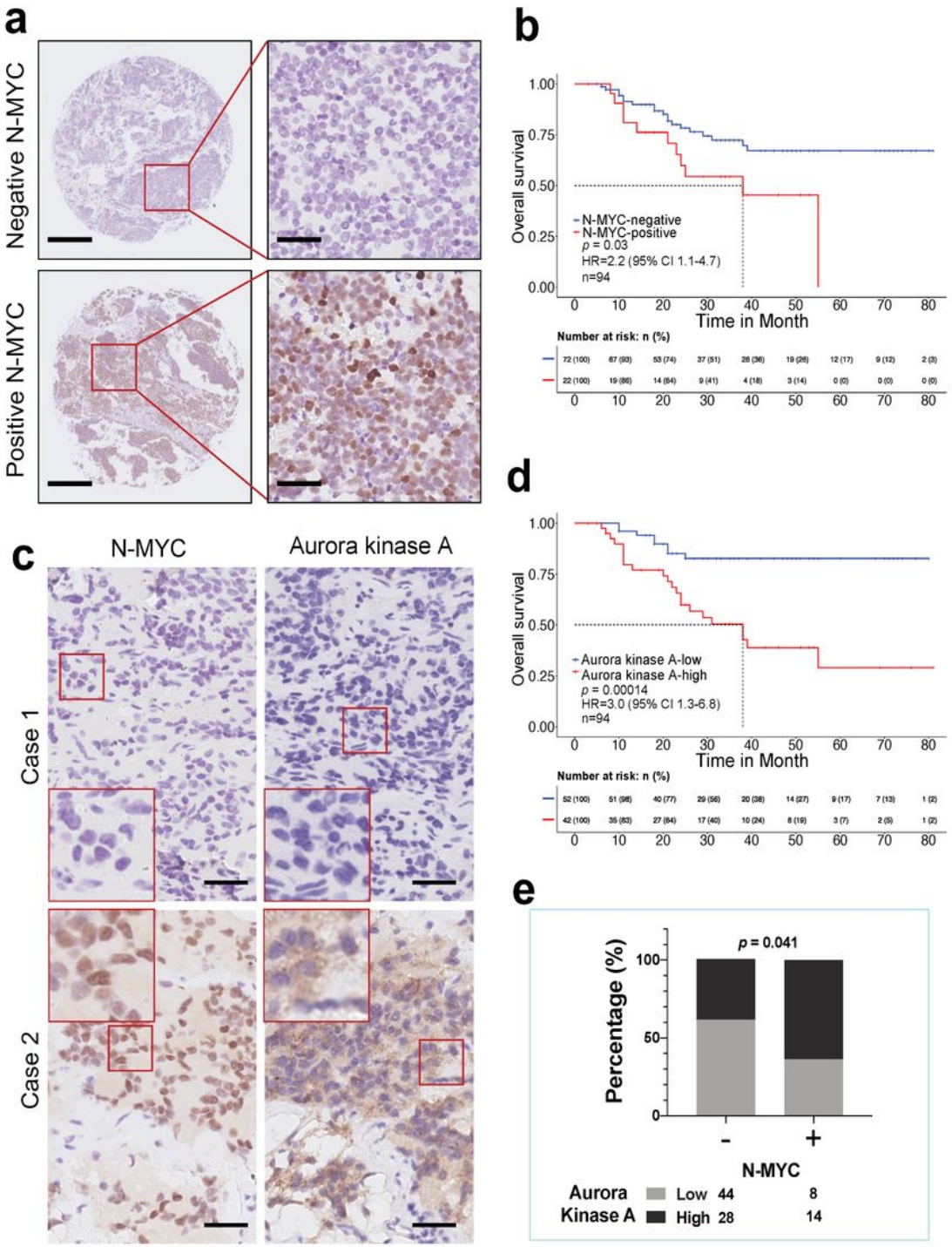

Figure 5

N-MYC expression correlates with Aurora kinase A status in MYCN non-amplified neuroblastomas, and is indicative of patient survival. (a) Representative N-MYC staining pattern (negative or positive N-MYC) in MYCN non-amplified neuroblastoma tissue microarray cores. Scale bar: 1mm (the left column) and 50μm (the right column). (b) Kaplan-Meier plot showing the overall survival in samples with negative vs. positive N-MYC expression. Numbers below are n (%). HR (hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown. (c) Adjacent tumour sections from representative cases showing N-MYC and Aurora Kinase A expression in MYCN non-amplified neuroblastoma. Scale bars: 50μm. (d) Kaplan-Meier plot showing the overall survival in samples with low vs. high Aurora kinase A expression. Numbers below are n (%). HR (hazard ratio), 95% CI (confidence interval), patient number (n) and p values are shown. (e) Graph showing percentage (%) and numbers of samples with low or high Aurora kinase A in the negative or positive N-MYC group. p = 0.041.

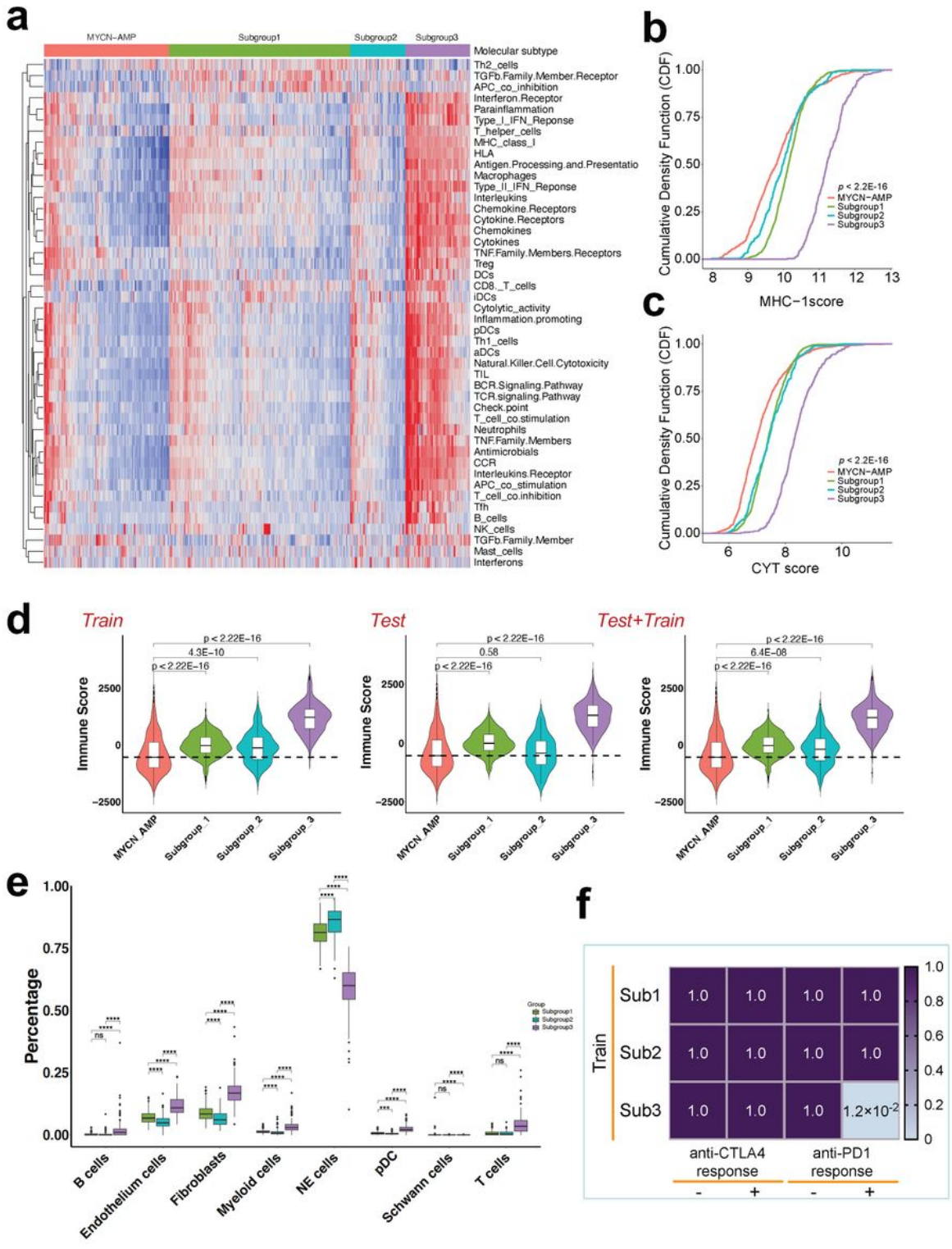

Figure 6

Subgroup 3 is accompanied by an "inflamed" gene signature. (a) Heatmap showing neuroblastoma associated immune pathways and immune cell signatures in subgroups and MYCN-AMP. Graphs showing the cumulative distribution of MHC-1 (b) or CYT (c) scores in different subgroups and MYCN-AMP. (d) Violin plots showing immune scores in different subgroups and MYCN-AMP in train, test or train plus test cohort. (e) Graph showing cell compositions of each subgroup using CIBERSORTx analysis. (f)

Graph showing differential putative immunotherapeutic response in different subgroups. Bonferroni adjusted p values indicated.

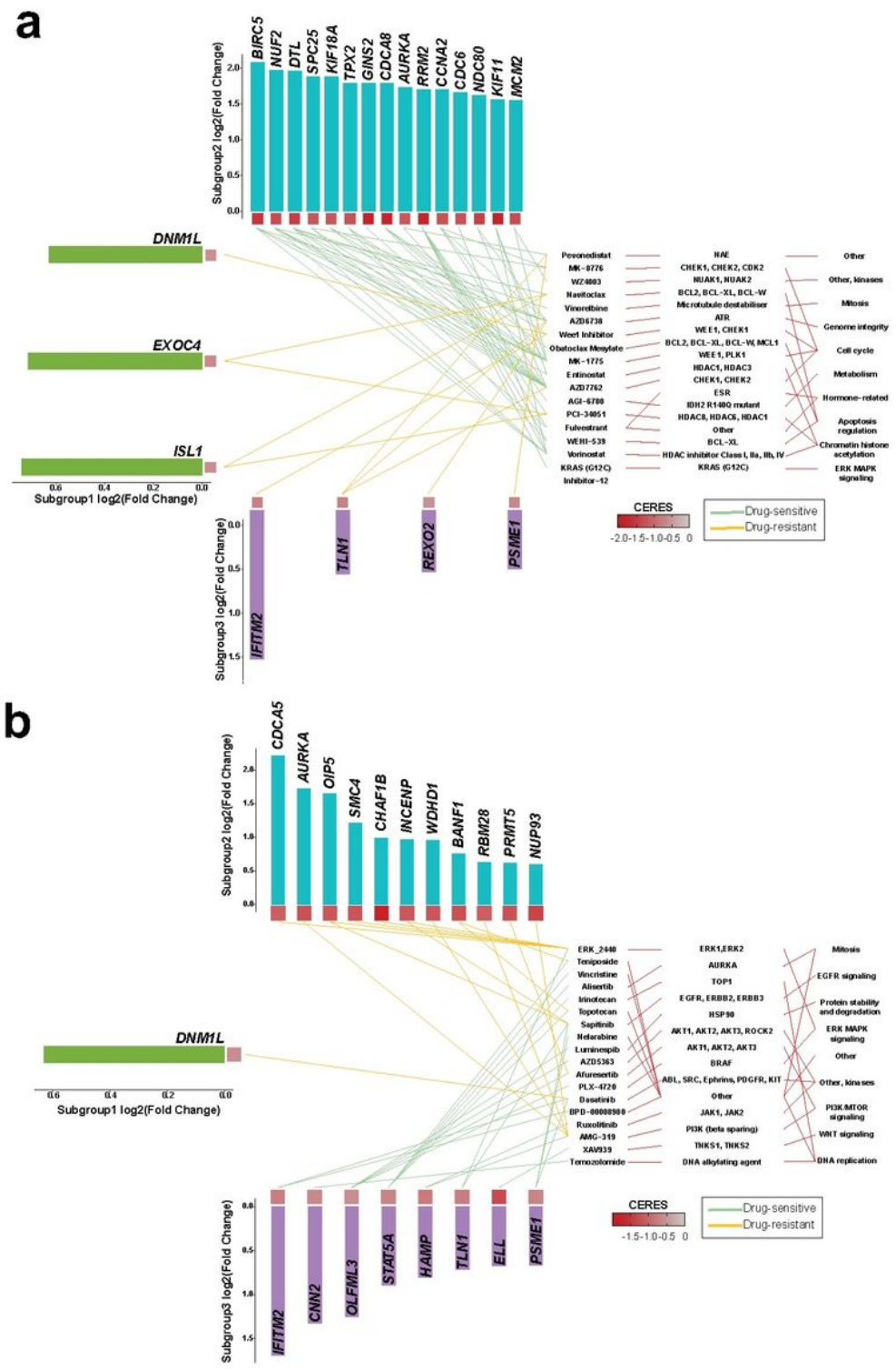

Figure 7

Integrative analysis of subgroup molecular features on drug response. Graphs showing interactions of up-regulated genes in each subgroup and drug response. (a) Upregulated genes (turquoise bars in upper panel) in subgroup 2 negatively correlate with the response to 18 drugs, while opposite in subgroup 1 and

3. (b) Upregulated genes (purple bars in lower panel) in subgroup 3 negatively correlate with the response to 18 drugs, while opposite in subgroup 1 and 2. The target genes and targeted pathways of drugs are listed in the panel to the right. The red line links the drug to its target and targeted pathway. Green lines indicate drug-sensitive and orange lines drug-resistant. CERES scores are also indicated.

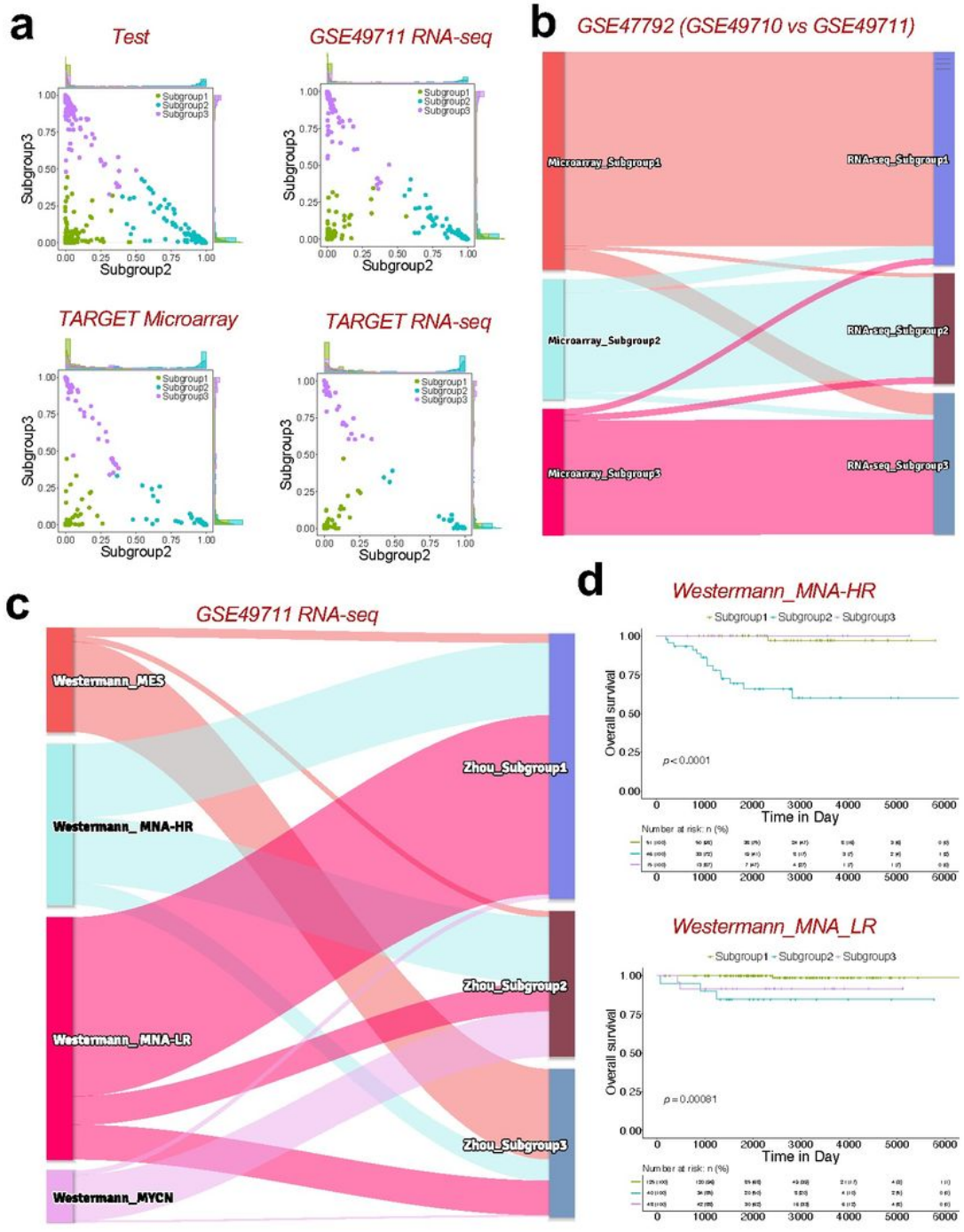

Figure 8

Identification and evaluation of independent predictors to subgroup patients within MYCN non-amplified neuroblastomas. (a) Predicted probability of each subgroup in 4 different cohorts. Each dot in the scatter plot corresponds to a sample (x-axis: predicted probability of subgroup 2, y-axis: predicted probability of subgroup 3). The histogram plot above the scatter plot displayed the distribution of subgroup 2 probabilities while the plot to the right of the scatter plot displayed the distribution of subgroup 3 probabilities. (b) Prediction differences in the superseries GSE47792 using data from either RNA-seq (GSE49711) or microarray (GSE49710). (c) Prediction differences in GSE49711 using subgrouping method from this report (named Zhou) or Westermann and colleagues (Westermann). (d) Kaplan-Meier plots showing the overall survival in Westermann_MNA-HR or Westermann_ MNA-LR patients using subgrouping method from this report. P values are shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx
- TableS2.xlsx
- TableS3.docx
- TableS4.xlsx
- TableS5.xlsx
- TableS6.xlsx
- TableS7.xlsx
- TableS8.xlsx
- TableS9.xlsx
- SupplementaryMaterialssubmittedLuwang.pdf