Clinically Relevant Biomarker Discovery in High-Risk Recurrent Neuroblastoma

Cancer Informatics
Volume 18: 1–16

DOI: 10.1177/1176935119832910

Peter Utnes[1], Cecilie Løkke[2], Trond Flægstad[1,2] and Christer Einvik[1,2]

[1]Department of Pediatrics, Division of Child and Adolescent Health, UNN – University Hospital of North-Norway, Tromsø, Norway. [2]Pediatric Research Group, Department of Clinical Medicine, Faculty of Health Science, The Arctic University of Norway – UiT, Tromsø, Norway.

ABSTRACT: Neuroblastoma is a pediatric cancer of the developing sympathetic nervous system. High-risk neuroblastoma patients typically undergo an initial remission in response to treatment, followed by recurrence of aggressive tumors that have become refractory to further treatment. The need for biomarkers that can select patients not responding well to therapy in an early phase is therefore needed. In this study, we used next generation sequencing technology to determine the expression profiles in high-risk neuroblastoma cell lines established before and after therapy. Using partial least squares-discriminant analysis (PLS-DA) with least absolute shrinkage and selection operator (LASSO) and leave-one-out cross-validation, we identified a panel of 55 messenger RNAs and 17 long non-coding RNAs (lncRNAs) which were significantly altered in the expression between cell lines isolated from primary and recurrent tumors. From a neuroblastoma patient cohort, we found 20 of the 55 protein-coding genes to be differentially expressed in patients with unfavorable compared with favorable outcome. We further found a twofold increase or decrease in hazard ratios in these genes when comparing patients with unfavorable and favorable outcome. Gene set enrichment analysis (GSEA) revealed that these genes were involved in proliferation, differentiation and regulated by Polycomb group (PcG) proteins. Of the 17 lncRNAs, 3 upregulated (*NEAT1, SH3BP5-AS1, NORAD*) and 3 downregulated lncRNAs (*DUBR, MEG3, DHRS4-AS1*) were also found to be differentially expressed in favorable compared with unfavorable outcome. Moreover, using expression profiles on both miRNAs and mRNAs in the same cohort of cell lines, we found 13 downregulated and 18 upregulated experimentally observed miRNA target genes targeted by *miR-21, -424* and *-30e, -29b, -138, -494, -181a, -34a, -29b*, respectively. The advantage of analyzing biomarkers in a clinically relevant neuroblastoma model system enables further studies on the effect of individual genes upon gene perturbation. In summary, this study identified several genes, which may aid in the prediction of response to therapy and tumor recurrence.

KEYWORDS: neuroblastoma, high risk, tumor recurrence, biomarker discovery, next generation sequencing, miRNA, lncRNA

RECEIVED: November 1, 2018. **ACCEPTED:** December 27, 2018.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Northern Norway Regional Health Authority (project number SFP1155-14).

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Christer Einvik, Pediatric Research Group, Department of Clinical Medicine, Faculty of Health Science, The Arctic University of Norway – UiT, NO-9037 Tromsø, Norway. Email: christer.einvik@uit.no

Introduction

Neuroblastoma is a childhood cancer with a wide range of clinical courses, spanning from spontaneous differentiation to rapid progression and fatal outcome. Neuroblastoma accounts for more than 7% of malignancies in patients younger than 15 years and around 15% of all pediatric oncology deaths.[1] It is the most common extracranial solid tumor in childhood and the most frequently diagnosed neoplasm during infancy. In most newborn patients with neuroblastoma, tumors will spontaneously regress without any need for treatment even with disseminated disease to liver, skin, or bone marrow. For this patient group, comprising low-risk tumors, the overall survival rate is more than 95%.[2]

Several prognostic factors are implicated in the risk stratification of neuroblastoma including age at diagnosis, localization of tumor, number of metastasis, tumor histology, grade of differentiation, mitosis-karyorrhexis index (MKI), *ALK* mutations, *MYCN* amplification, and other segmental chromosome alterations (SCA), such as losses of chromosomes 1p, 3p, 4p, 11q and gains of 1q, 2p, 17q.[3–8] In spite of these factors, patients in the high-risk group display an overall survival less than 50%.[6,9] Genome-wide association studies (GWAS) have identified several single-nucleotide polymorphisms (SNPs) associated with an increased risk of neuroblastoma. However, while these aberrations indicate an increased susceptibility to the disease, they are less useful for stratifying patients into risk groups after diagnosis.[10] Hence, risk stratification after diagnosis and better prediction of response to therapy and tumor recurrence is needed.

Several studies have implicated genetic markers in neuroblastoma through the use of high-throughput technologies.[11–15] In these studies, protein-coding genes have gained most attention. Large-scale efforts have shown that mRNAs only cover 2.94% of the human genome.[16] As such, efforts to study the remaining ~97% has culminated in more than 40 000 non-coding RNAs (ncRNAs, data from ENSEMBL database version 92). Therefore, risk prediction based on gene signatures may prove to be more accurate when including a significant larger repertoire of genes such as ncRNAs.

Although there are several subgroups of ncRNAs, considerable attention has been paid to micro RNAs (miRNAs) and long non-coding RNAs (lncRNAs). miRNAs are small regulatory RNAs (~18-24 nucleotides) which base pairs to target RNAs to induce post-transcriptional silencing in mammalian cells. Several previously described miRNAs have been linked to neuroblastoma.[17–24] lncRNAs (>200 nucleotides length) represent a large and heterogeneous group of ncRNAs and

Table 1. Overview of neuroblastoma cell lines used in this study.

CELL LINE	AGE (MONTHS)	DOUBLING TIME (HOURS)	TREATMENT						
SMS-KCN	11	109							
SMS-KCNR	11	72		CPM	DXR				
CHLA-15	18	21							
CHLA-20	24	28	CPPD	CPM	DXR	VM-26			
SK-N-BE(1)	24	96							
SK-N-BE(2)	29	72		CPM	DXR		VCR	RAD	
CHLA-122	24	72							
CHLA-136	36	44			DXR	ETOP		RAD	BMT
SMS-KAN	36	95							
SMS-KANR	36	69		CPM	DXR			RAD	
NBL-W	6	24							
NBL-WR	12	NA	CPPD	CPM/DTIC	DXR + DTIC	ETOP	VCR		

Abbreviations: BMT, bone marrow transplant; CDDP, cisplatin; CPM, cyclophosphamide; DTIC, dacarbazine; DXR, doxorubicin; ETOP, etoposide; RAD, radiation therapy; VM-26, teniposide; NA, not available.
All cell lines used in this study were *MYCN*-amplified (MNA) and stage IV, except CHLA-15/-20 (stage IV, non-MNA) and NBL-W/-WR (MNA, stage IVS). Data based on data sheets provided from the Children's Oncology Group. Data on NBL-W/-WR based on Foley et al[48] and Cohn et al.[49]

have been associated with diagnosis, classification, progression, and response to treatment in a number of diseases.[25,26] A variety of functions have been associated with lncRNAs[27] and numerous lncRNAs have been associated with neuroblastoma including *CASC15*,[28–30] *DALIR*,[31] *GAS5*,[32] *HAND2-AS1*,[33,34] *LINC00467*,[35] *MYCNUT*,[36] *MEG3*,[37] *NBAT1*,[38] *MIAT*,[39] *MYCNOS*,[40–42] *SNHG1*,[43] and *SNHG16*.[44]

Common methods to detect differentially expressed genes from transcriptomic studies have been univariate statistical methods including analysis of variance (ANOVA), linear models, and t-tests. However, these methods ignore the relationship between different features and may miss important biological information.[45] Dimension reduction analysis has emerged as a popular method to investigate the large number of variables reported from transcriptomic studies where several methods have been developed for this purpose.[46] In this article, we used next generation sequencing technology to determine the expression profiles of mRNAs and lncRNAs in neuroblastoma cell lines established before and after therapy followed by multilevel partial least squares-discriminant analysis (PLS-DA) introduced in the R package *mixOmics*. The advantage of using a multivariate paired data analysis improves the power to detect the underlying treatment effects and eliminates variation due to intrinsic variation between the subjects.[47] Using bioinformatic analysis and a large patient cohort, we identified mRNAs and lncRNAs that may be important in predicting response to therapy and tumor recurrence. In addition, we investigated the miRNA-mRNA regulatory axis based on our previous miRNA expression profiling study.[17]

Materials and Methods

Neuroblastoma cell lines

We used a unique panel of 12 neuroblastoma cell lines (Table 1) isolated from six neuroblastoma patients before (SK-N-BE(1), SMS-KAN, SMS-KCN, CHLA-15, CHLA-122, NBL-W) and after (SK-N-BE(2), SMS-KANR, SMS-KCNR, CHLA-20, CHLA-136, NBL-WR) treatment with combination chemotherapy regimens.

CHLA cells were grown in Iscove's modified Dulbecco's medium supplemented with 20% fetal bovine serum, 4 mM L-glutamine, and 1× ITS (5 µg/mL insulin, 5 µg/mL transferrin, 5 ng/mL selenous acid). The other cell lines were grown in RPMI-1640 supplemented with 10% fetal bovine serum and 2 mM L-glutamine (final concentrations).

All cell lines were cultured at 37°C in a humidified incubator with 95% air and 5% CO_2 atmosphere without antibiotics. The identity of all cell lines was authenticated by short tandem repeat (STR) analysis at the Center of Forensic Genetics, The Arctic University of Norway—UiT, Norway. Cells were tested and confirmed negative for mycoplasma contamination.

RNA isolation

Total RNA was extracted by adding 1 mL of TRIzol (Invitrogen, Carlsbad, USA) to the samples, vortexed for 1 min, and placed at room temperature for 5 min before adding 200 µL chloroform. The phases were mixed and phase separation was done by centrifuging at 12 000×g, 15 min at 4°C. To maximize yields and reducing the risk of phenol contamination, the lower

phenol phase was removed and the sample was re-spun at 12 000×g, 15 min at 4°C. The aqueous phase was moved to a new tube with glycogen (Ambion) and 1 volume of 0.2 M NH_4OAc isopropanol. The mixture was chilled at −20°C for 15 min and RNA was co-precipitated with glycogen by centrifuging at 12 000×g, 15 min, 4°C. The visible RNA pellet was washed twice in 1 mL 75 % ethanol and re-acquiring of the pellet was done by spinning down at 7500×g, 5 min at 4°C. The pellet was resuspended in Tris-ethylenediaminetetraacetic acid (EDTA) buffer (10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA). RNA concentration and DNA contamination were assessed using NanoDrop 2000 (Thermo Scientific).

RNA sequencing

Total RNA samples were shipped to NXT-DX (Ghent, Belgia) for RNA sequencing on the HiSeq 2000 platform. Briefly, RNA concentrations were measured using Agilent's Eukaryote Total RNA Pico Assay. After fragmentation, a quality control (QC) was done on Agilent 2100 HS DNA chip. Concentration was determined with smear analysis on the Agilent 2100 and checked for degradation. Samples were prepared using the NEBNext Ultra Directional RNA Library Protocol (New England Biolabs). The paired-end, cDNA libraries were consequently sequenced on the Illumina HiSeq 2000 sequencer. BaseCalling was done with Illumina Casava Pipeline 1.8.2. Initial quality assessment was based on data passing the Illumina Chastity filter. The second quality assessment was based on the reads using the FASTQC quality control. Reads were mapped with STAR Aligner 2.5[50–52] onto the human reference genome GRCh38. Duplicate removal was done with Picard's tool MarkDuplicates (http://broadinstitute.github.io/picard). Read counting was performed with featureCounts v1.5.0-p2[53] on ENSEMBL gene annotations. Normalization was carried out using the Relative Log Expression (RLE) method.[54]

RT-qPCR and validation of RNA sequencing results

RNA was reverse transcribed using High Capacity Reverse Transcription Kit (Applied Biosystems). Ten microliters of RNA was mixed with 10 µl of cDNA reverse transcription master mix following the standard protocol. Briefly, 1 µg RNA was reverse transcribed using the PTC-200 Thermal Cycler with the following thermal cycling profile: (1) pre-incubation at 25°C for 10 min, (2) reverse transcription at 37°C for 120 min, and (3) inactivation of the reverse transcriptase by heating to 85°C for 5 min. RT-qPCR was carried out with SYBR Green (Applied Biosystems) on the 7300 Real-Time PCR System (Applied Biosystem). For each reaction, 5 µL cDNA (25 ng) was mixed with 15 µL SYBR Master Mix containing 10 µL SYBR, 0.8 µL 5 µM stock primer solution, and 4.2 µL nuclease free water. The complete cycler profile was (1) pre-incubation at 50°C for 2 min, (2) heat denaturation at 90°C for 10 min, (3) 40 repetitions of 95°C for 15 s (denaturation step) followed by data collection at 60°C for 1 min (annealing step), and (4) to validate each primer pair, a dissociation stage for each plate was applied using 95°C, 15 s; 60°C, 1 min; 95°C, 15 s; 60°C, 15 s. Samples were normalized against *RNF111* and *GUSB*. Primers used are listed in Supplemental file 8. The ddCT method was applied in the calculation of relative expression.

Gene set enrichment analysis

For gene set enrichment analysis (GSEA), Fisher's exact test was used to compute significant enriched terms. With gene sets from enrichR, the use of a gene background model was omitted since we here wanted to compare our model system on a global scale. In the case of searching for enriched biological processes and pathways separating primary and recurrent cell lines, a background gene model was provided to adjust for differences between the two groups. The R packages clusterProfiler and ReactomePA was used in the GSEA. In addition, the simplify function provided in clusterProfiler was used to remove redundant terms. Gene sets were collected from several resources including Harmonizome,[55] enrichR,[56,57] Molecular Signature Database (MSigDB),[58,59] and ENCODE Gene Set Hub.

Principal components analysis

Principal component analysis (PCA) seeks to find the linear directions that maximize the variance within a dataset. PCA does not require the class label of each sample prior to analysis (unsupervised clustering). The linear directions are referred to as principal components (PCs) and are selected orthogonal to each other with decreasing degree of variation (ie, the first component holds the largest amount of variation). In each PC, each feature (eg, gene) contributes to the linear subspace according to its variation. This contribution is measured as the distance from the centroid (center of variation) to each individual feature and is referred to as the loading vector. A mathematical definition of a vector is that it gives both direction and the length (magnitude) of a particular feature. We initially scaled the data prior to PCA using Z-score normalization, but found many genes in our final model with small fold changes between cell lines isolated from primary and recurrent tumors. To circumvent this, we instead applied a log-transformation of the data prior to PCA to shrink the effect of genes with high variation while omitting the scaling procedure. The same procedure was applied to PLS-DA. The R package mixOmics was used to perform PCA and PLS-DA on repeated measurements (multilevel studies). The within-subject deviation matrix was therefore calculated before PCA and PLS-DA was carried out. The 3D plot of the first three components was made with the R package plot3D.

Partial least squares-discriminant analysis

PLS-DA is similar to PCA in that it also finds a linear subspace within the dataset. However, instead of searching for dimensions with maximum variance, it searches for dimensions with the maximum amount of covariance between two data matrices X (eg, transcriptomics) and Y (eg, proteomics). In the discriminant version of PLS (PLS-DA), the Y matrix is recoded with dummy variables to represent the class membership of each sample (supervised clustering). This approach was not originally designed in the use of PLS, but has often been applied in a manner that has proven successful in selecting variables that can discriminate variables in X with regard to Y.[60] In PCA, as opposed to PLS-DA, linear transformation is only applied with considerations to X. To facilitate the use of PLS-DA on repeated measurements, we took the advantage of using multilevel PLS-DA implemented with the mixOmics R package. After calculating the contribution of each latent factor, we applied least absolute shrinkage and selection operator (LASSO) to select the most important variables. LASSO evaluates the model by shrinking the estimated coefficients to zero. Only latent factors with perfect stability were selected with leave-one out cross-validation (LOOCV). LOOCV was preferred over M-fold cross-validation due to the limited number of samples in our cohort of cell lines.

PLS-DA on lncRNAs followed the same methodology as above. Here, only genes annotated as HGNC symbols or NCBI genes were included in the analysis. This step removes unreliable gene annotations from the ENSEMBL clone-based and RFAM database.

Datasets used in this study

Raw sequence read files were uploaded to the Sequence Read Archive (SRA accession: PRJNA491629). The neuroblastoma patient cohort SEQC (n = 498) was obtained from Gene Expression Omnibus (GEO) with accession GSE62564 and GSE49711.[61–64] SEQC consists of 498 patients where 272 tumors were classified as either unfavorable (n = 91) or favorable (n = 181) tumors. Normalized count data were obtained from GSE62564 and Refseq IDs was re-annotated with ENSEMBL gene identifiers and HUGO gene names. Sample characteristics were extracted from the series matrix file provided with the two accession identifiers.

Code availability

Scripts and additional code is available for download at https://github.com/utnesp/Neuroblastoma_Biomarker_2018

Statistical analysis

Statistical associations and differences were determined using Pearson correlation, Mann-Whitney, Fisher's exact test, and Cox Proportional Hazards Model Survival where appropriate. Survival data were presented using Kaplan-Meier survival analysis with the log-rank test. A *P*-value less than .05 was considered significant. For all box and whisker plots, the horizontal line in each box indicates the median of the data, whereas the top and bottom of the box represent the upper and lower quartiles, respectively. The whiskers extend to the most extreme point within 1.5 times the interquartile range (IQR) of the box. Data beyond 1.5 times the IQR are depicted as dots.

Results

Sequencing and validation of paired neuroblastoma cell lines

We have previously reported on the miRNnome of paired cell lines isolated from neuroblastoma patients before and after intensive treatments.[17] To broaden our understanding of genes contributing to recurrence in neuroblastoma, we included deep sequencing of mRNAs and lncRNAs in the same paired cell lines (Table 1).

Before conducting the sequencing, we performed a STR analysis to authenticate the origin of the cell lines. All cell line pairs were authenticated, except SMS-KCN/-KCNR (Supplemental file 1). Consequently, the KCN/KCNR pair was discarded from further analyses. In addition, cells were tested and confirmed negative for mycoplasma contamination. From our polyA-enriched, directional RNA sequencing, we obtained 16.7-21.8 million uniquely mapped paired-end reads onto the human reference genome following the ENSEMBL GRCh38 annotation. Raw sequence read files were uploaded to the Sequence Read Archive (SRA accession: PRJNA491629). A full report on the sequencing, raw and normalized counts, may be found in Supplemental files 2 to 4, respectively. Finally, we successfully validated a subset of the RNA sequencing data with RT-qPCR (Figure 1). Here, the Pearson correlation coefficient was 0.73-0.94 in 9 genes used to compare the two methods ($P < .05$). The expression of *MYCN* and *MYCNOS* was as expected absent in cell lines that were not *MYCN*-amplified (MNA, CHLA-15/-20).

Characterization of individual RNA classes

After expression values were calculated and filtered from the RNA sequencing results, a total of 9704 mRNAs and 439 noncoding RNAs with unique ENSEMBL gene IDs were identified. To characterize the individual RNA classes, we classified each gene according to its gene biotype and investigated the mean level of expression (log2 transformed) in our dataset (Figure 2). Our findings show that lncRNAs in general have lower expression than protein-coding genes. In brief overview, the top 10 expressed genes (in decreasing order) in each category were as follows: protein coding: *EEF2I, GAPDH, EEF1A1, ACTB, ACTG1, HSP90AB1, MAP1B, TUBB, HNRNPA2B1*, and *RPL3*. lincRNA: *MALAT1, NEAT1, MEG3, MIAT, NORAD, LINC00599, CASC15,*

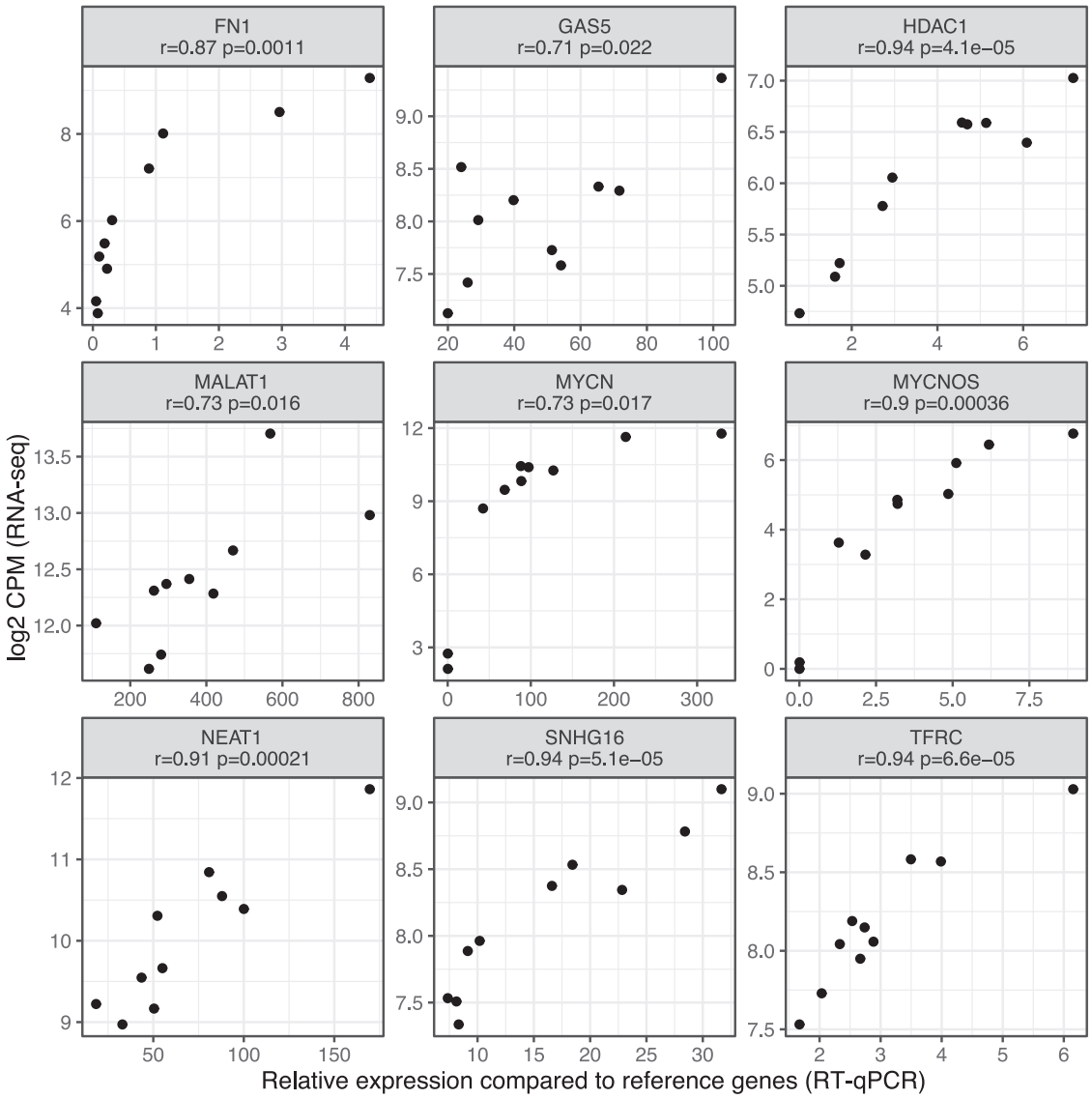

Figure 1. Validation of RNA sequencing results.
Scatterplot of RNA sequencing expression levels (Y-axis, CPM log2-transformed) compared with RT-qPCR expression levels relative to two reference genes, *RNF111* and *GUSB* (X-axis). Expression values in RT-qPCR were calculated using the ddCT method. Title above each plot signifies the Pearson correlation coefficient (r) and corresponding *P*-value. Data points used in RT-qPCR were mean expression of two biological replicates. CPM indicates counts per million.

EBP41L4A-AS1, *SNHG15*, and *LINC00667*. Antisense: *HAND2-AS1*, *GABPB1-AS1*, *KCNQ1OT1*, *FGD5-AS1*, *HCG18*, *PSMA3-AS1*, *SLC25A25-AS1*, *THAP9-AS1*, *SNHG7*, and *TTN-AS1*. Processed transcript: *LRRC75A-AS1*, *SNHG16*, *GAS5*, *SNHG1*, *OIP5-AS1*, *SNHG14*, *LINC00641*, *SNHG5*, *LINC001578*, and *PCBP1-AS1*. A complete overview of expressed genes in each category may be found at https://utnesp.github.io/Neuroblastoma_Biomarker_2018/

Gene expression profile of cell lines represents MYCN-amplified neuroblastoma

As an approach to assess the model system and the relevance of the obtained expression data, we applied gene set enrichment using the top 5% of highest expressed genes after sorting by mean expression across cell lines (n = 306). After sorting the enrichment results by *P*-values, we selected the top 10 terms in the 2 gene sets *ENCODE and ChEA Consensus TFs from ChIP-X* and *Cancer Cell Line Encyclopedia* obtained from enrichR.[56,57] Not surprisingly, knowing that 4/5 of the cell line pairs were MNA, we found *MYC* as the most significantly enriched transcription factor (Figure 3A). Moreover, we found that all top 10 terms enriched in *Cancer Cell Line Encyclopedia* gene sets were neuroblastoma cell lines (Figure 3B). We therefore concluded that the gene expression profile indeed is representative of MNA neuroblastoma cell lines.

55 mRNAs separates cell lines isolated from primary and recurrent tumors

To compare the gene expression profiles from neuroblastoma cell lines isolated before and after treatment, we employed PCA

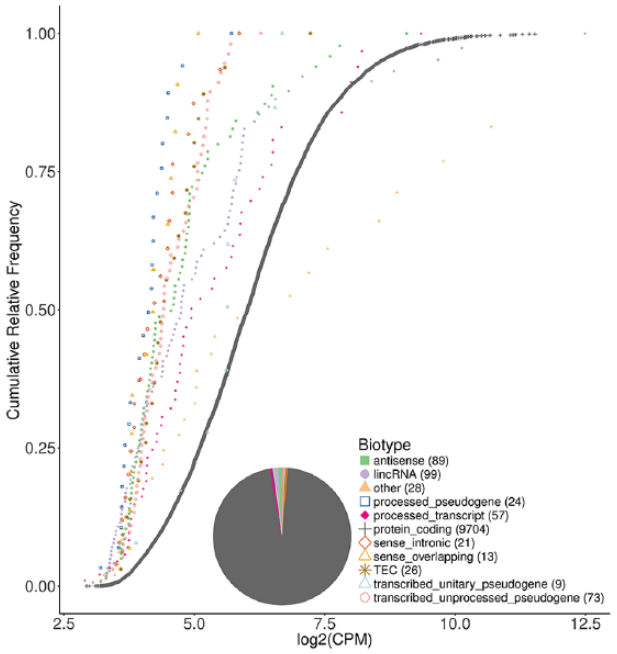

Figure 2. Characterization of individual RNA classes.
Each point in the plot corresponds to an individual gene. X- and Y-axis represent the average CPM across all cell lines (log2 transformed) and the cumulative relative frequency, respectively. Embedded pie chart shows the distribution of the different biotypes. Numbers in parenthesis correspond to number of genes within each RNA class. CPM indicates counts per million; lincRNA, long intergenic RNAs.

and the Discriminant Analysis version of Partial Least Squares (PLS-DA) using the R package mixOmics.[45]

Due to the elevated expression levels of mRNAs compared with lncRNAs, we divided the dataset into two subsets. The reason for this subgrouping is two sided. First, the loading vectors for each gene will vary according to its variance/covariance. In this situation, larger expression differences will have an impact on the magnitude of the vectors in the PC and the discriminant analysis. As such, lncRNAs will tend to have lower loading vector scores. Second, we get a better overview of the two different classes. In this section, only mRNAs (n = 9704) are included in the analysis.

As a first approach to explore the mRNA dataset, we employed PCA to assess the variation across the 5 paired samples. By utilizing the first three PCs, explaining 76% of the total variation, we were able to discriminate between cell lines isolated from primary and recurrent tumors (Figure 4A). In this context, all stage IV cell line pairs, except the stage IVS NBL-W/NBL-WR pair, separated well within the first PC. We also note that the first PC only described 33 % of the total variation, which we assume is due to the heterogeneity between samples. To reduce between sample variation and to highlight the treatment effects, we applied multilevel PLS-DA. First, we visually observed the separation of the paired cell lines in our PLS discriminant analysis. As can be seen from Figure 4B, we observed a good separation of the cell lines isolated from primary and recurrent tumors in the X-variate component 1.

To further select the genes best suitable to discriminate between the two groups in this component, we applied LASSO and leave-one-out cross-validation. Based on these criteria, we identified 55 mRNAs where 25 and 30 genes were downregulated and upregulated in recurrent cell lines, respectively (Figure 5).

20 clinically relevant mRNAs may predict neuroblastoma outcome

We next sought to investigate the 55 mRNAs identified in our cell lines in a clinical relevant setting based on favorable compared with unfavorable outcome according to the International Neuroblastoma Pathology Classification (INPC) system.[65,66] This outcome considers three measures: tumor subtype (differentiation status), MKI and age (older patients generally respond worse to treatment). Based on Mann-Whitney test, 20 mRNAs were clinically relevant when comparing unfavorable (n = 91) and favorable (n = 181) tumors in a clinical cohort of neuroblastoma patients (Figure 6).

We further applied the Cox proportional hazard (Cox PH) model to test for difference in survival rates. In accordance with the Mann-Whitney test, all differentially expressed genes in primary and favorable, or recurrent and unfavorable tumors, showed at least a twofold decrease or increase in hazard ratios, respectively (Figure 7). We also inspected the genes using a log-rank test as well as a visual confirmation using Kaplan-Meier plots (Supplemental file 6). The findings were consistent with the Cox PH analysis. We conclude that these 20 biomarkers may be clinically relevant in the prediction of neuroblastoma outcome.

Cell lines isolated from recurrent tumors reveal higher proliferation rates and negative regulation of cell differentiation

After obtaining loading vector scores from the PLS-DA analysis, we were interested in biological processes and signaling pathways describing the two groups of cell lines. We applied GSEA by including genes with PLS-DA loading vectors with an absolute magnitude larger than 0.02 within the X-variate component 1. This cut-off was selected to return meaningful and significantly enriched gene sets and resulted in an adequate number of 201 and 210 genes used in the enrichment analysis of cell lines isolated from primary and recurrent tumors, respectively. Based on this criterion, we identified signaling and organization within the synapse and the plasma membrane, as well as pathways in axon guidance, neuronal system, signaling by G protein-coupled receptors, *L1CAM* interactions, and protein-protein interactions at synapses as significant biological processes in primary cell lines (Figure 8). Recurrent cell lines on the other hand exhibited negative regulation of cell differentiation and development, and a positive regulation of cell proliferation.

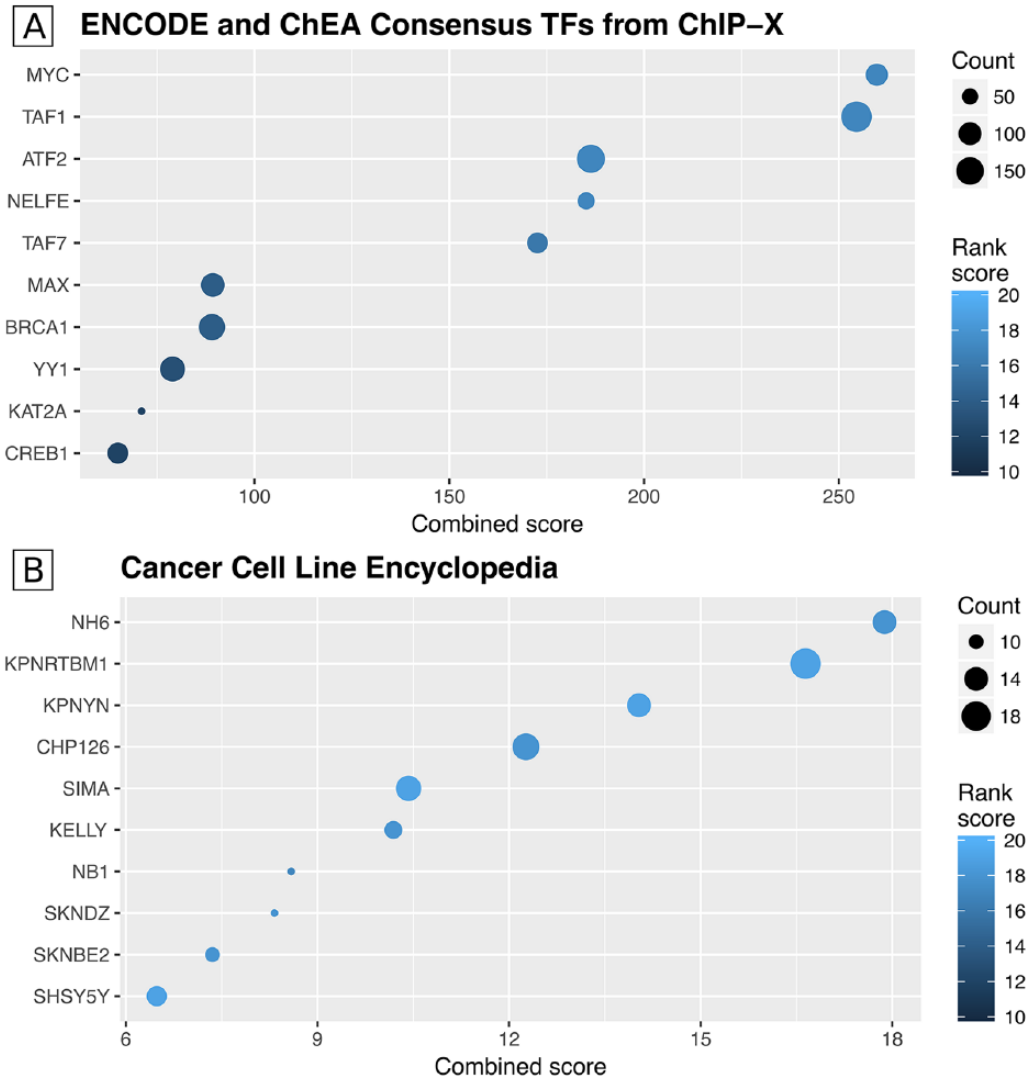

Figure 3. Model system represents MYCN-amplified neuroblastoma. (A) Transcription factors enriched in the panel of cell lines used in this study. Top 5% of expressed genes (n = 306) were enriched in gene sets with *MYC* as the most significant enriched term using gene sets from *ENCODE and ChEA Consensus TFs from ChIP-X*. (B) All top 10 enriched terms using gene sets from *Cancer Cell Line Encyclopedia* show gene sets in neuroblastoma cell lines.
Count: Number of genes within enriched term. Rank score: –log10 of *P*-value from the Fisher exact test using a gene input-based look-up table as background model. Combined score: Represent the product of *P*-value and the log of rank score.

In accordance with this, we also found the linear regression coefficient concerning the doubling time between cell lines isolated from primary and recurrent tumors to be 0.57 ($P = .05$, doubling times listed in Table 1). This determines almost a doubling in proliferation rates of recurrent compared with primary cell lines. Moreover, pathways found to be enriched in recurrent cell lines were G protein-coupled receptor signaling, extracellular matrix organization and degradation, insulin growth factor signaling, and post-translational protein phosphorylation.

Finally, we employed enrichment analysis using transcription factor gene sets from ENCODE and ChEA. This analysis identified *SUZ12, TRIM28, EZH2, RNF2, JARID2*, and *EED* as putative upstream master regulators of the 55 genes identified previously (Figure 9).

Long non-coding RNAs in prediction of neuroblastoma outcome

Similar to our analyses on mRNA expression, we applied the PLS-DA methodology to investigate lncRNAs separating cell lines isolated from primary tumors and their respective recurrent counterparts. We found 9 upregulated and 8 downregulated lncRNAs in the recurrent cell lines (Figure 10). We also inspected the expression of these genes in favorable (n = 181) compared with unfavorable (n = 91) neuroblastoma samples. Based on Mann-Whitney test, genes upregulated in recurrent cell lines and differentially expressed in a cohort of neuroblastoma patients were *NEAT1, SH3BP5-AS1*, and *NORAD*. Genes downregulated in recurrent cell lines and differentially expressed in the same cohort were *DUB3, MEG3*, and

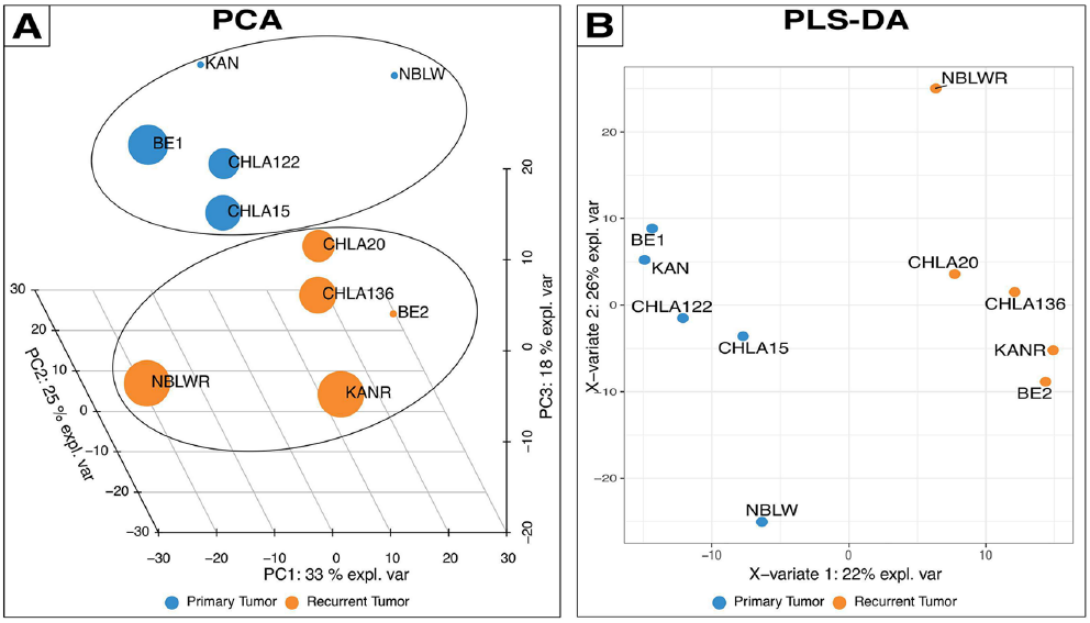

Figure 4. Separation of cell lines isolated from primary and recurrent tumors using principal component and PLS discriminant analysis. (A) Cell lines isolated from primary (upper circle, blue dots) and recurrent tumors (lower circle, orange dots) were found to separate using three principal components in the principal component analysis. All cell lines, except NBL-W and NBL-WR, separated within principal component (PC) 1. The explained variance of each PC is depicted at each axis. Dotsize represents the distance from the lower left coordinate. (B) Cell lines isolated from primary (blue dots) and recurrent tumors (orange dots) were found to separate well within the X-variate component 1 of the partial least squares-discriminant analysis (PLS-DA). The X-variate component 2 is only included here to visualize the separation in a two-dimensional space. PCA indicates principal component analysis. Axes in PCA and PLS-DA score plot refer to component scores.

DHRS4-AS1. In summary, 6 lncRNAs may be able to predict neuroblastoma outcome in a clinically relevant setting.

Association between miRNAs and predicted miRNA target gene levels

We have previously reported on differentially expressed miRNAs in the same panel of neuroblastoma cell lines as used in this study.[17] 34 downregulated (*mir-100-5p, -10b-5p, -136-5p, -136-3p, -138-1-5p/-138-2-5p, -154-3p, -16-2-3p, -181a-1-3p, -199a-1-3p/-199a-2-3p/-199b-3p, -29b-1-3p/-29b-2-3p, -30e-5p, -323a-3p, -323b-3p, -329-3p, -337-3p, -33a-5p, -34a-5p, -369-3p, -376a-3p, -376b-3p, -376c-3p, -379-5p, -380-3p, -381-3p, -409-3p, -410-3p, -4454-5p, -487a-3p, -487b-3p, -494-3p, -495-3p, -543-3p, -654-3p, -7-2-3p*) and 8 upregulated (*mir-181a-1-5p/-181a-2-5p, -181b-1-5p/-181b-2-5p, -193b-3p, -21-5p, -30a-5p, -375, -424-5p, -99b-5p*) miRNAs in primary compared with recurrent cell lines were predicted to target 482 experimentally observed mRNAs (Table S3 from our previous study[17]). The miRNA expression data used in our previous study are submitted here as Supplemental file 7.

Using the expression data on mRNAs in this study, we found 324 of the 482 predicted mRNAs to be expressed in either primary or recurrent cell lines. We further defined a predicted miRNA target gene when its expression pattern was inversely expressed with that of the 34 downregulated and 8 upregulated miRNAs. Here, we selected miRNA target genes with loading vectors with an absolute magnitude larger than 0.01 from the PLS discriminant analysis. A lower cut-off within the discriminant analysis was selected since miRNAs are often thought of as fine tuners of mRNA expression. Based on this rule, we found two upregulated (*miR-21-5p, -424-5p*) and 7 downregulated microRNAs (mir-*30e-5p, -29b-3p, -138-5p, -494-3p, -181a-5p, -34a-5p, -29b-3p*) which were inversely expressed with 13 and 18 predicted miRNA targets in the recurrent cell lines, respectively (Table 2).

Discussion

We have previously reported on the miRNnome of cell lines isolated from neuroblastoma patients before and after intensive treatments.[17] As miRNAs only constitute a small portion of the human genome, we here expanded our sequencing to include mRNAs and lncRNAs from the same cell lines. The drug-resistance patterns expressed by these cell lines have previously been demonstrated to correlate with treatment intensities exposed to patients.[67,68] In this study, we predicted response to therapy and tumor recurrence in high-risk neuroblastoma using both mRNA, lncRNA, and miRNA expression data.

We identified 55 mRNAs able to discriminate primary and recurrent cell lines (Figure 5). Twenty of the 55 mRNAs were differentially expressed in patients with favorable compared with unfavorable outcome (Figure 6). These genes were also associated with overall survival rate of patients with at least a twofold increase or decrease in hazard ratios (Figure 7, Supplemental file 6). The strongest predictor of tumor

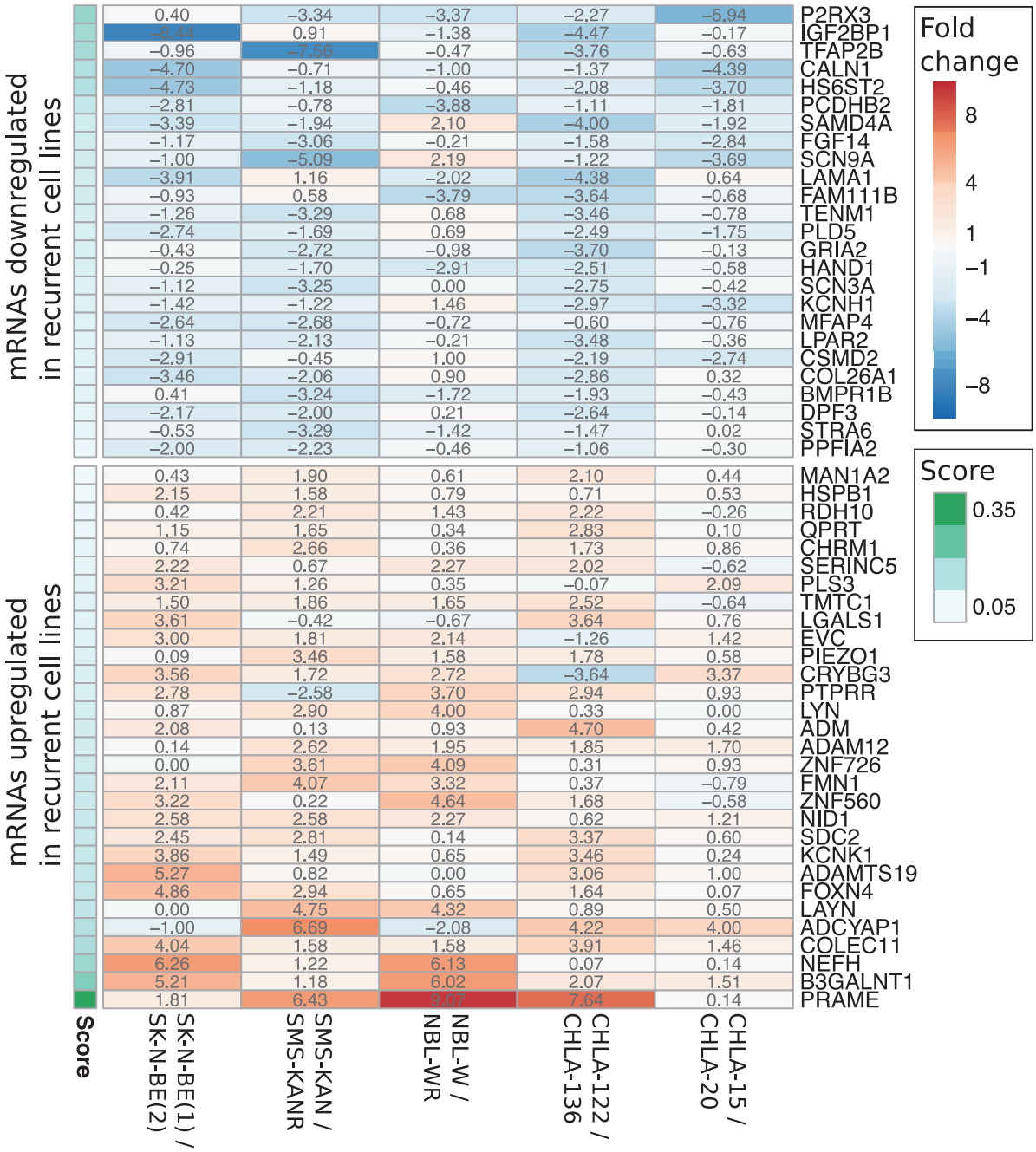

Figure 5. 55 mRNAs separate cell lines isolated from primary and recurrent tumors.
25 and 30 mRNAs were found to be downregulated and upregulated in recurrent cell lines, respectively. Each column represents individual cell lines isolated at diagnosis (primary) and after treatment (recurrent). Scale bar and numbers within each cell represent fold change values between recurrent and primary cell lines. Genes are plotted in order of the magnitude of the loading vector (score) from the X-variate component 1 in PLS-DA.

recurrence and unfavorable outcome was Preferentially Expressed Antigen In Melanoma (*PRAME*). *PRAME* has previously been universally associated with higher tumor stage and age of patients at diagnosis in neuroblastoma.[69] Interestingly, natural killer cells have been shown to facilitate *PRAME*-specific cytotoxic T cell response in neuroblastoma.[70] These findings encourage further investigations of *PRAME* as a potential target for immunotherapeutic strategies in neuroblastoma.

Positive regulation of cell proliferation and negative regulation of cell differentiation were significantly enriched terms in our panel of recurrent cell lines (Figure 8). Moreover, the expression of transcription factor activation protein 2 beta (*TFAP2B*) was downregulated in recurrent cell lines (Figure 5) and in unfavorable tumors (Figure 6). In line with this observation, a recent study showed that forced expression of *TFAP2B* promotes and is required for neuronal differentiation in neuroblastoma cells.[71]

We identified members of the Polycomb group (PcG) proteins, *SUZ12, EZH2, RNF2, JARID2*, and *EED*, as putative upstream master regulators of the genes separating cell lines from primary and recurrent cell lines (Figure 9). Specifically,

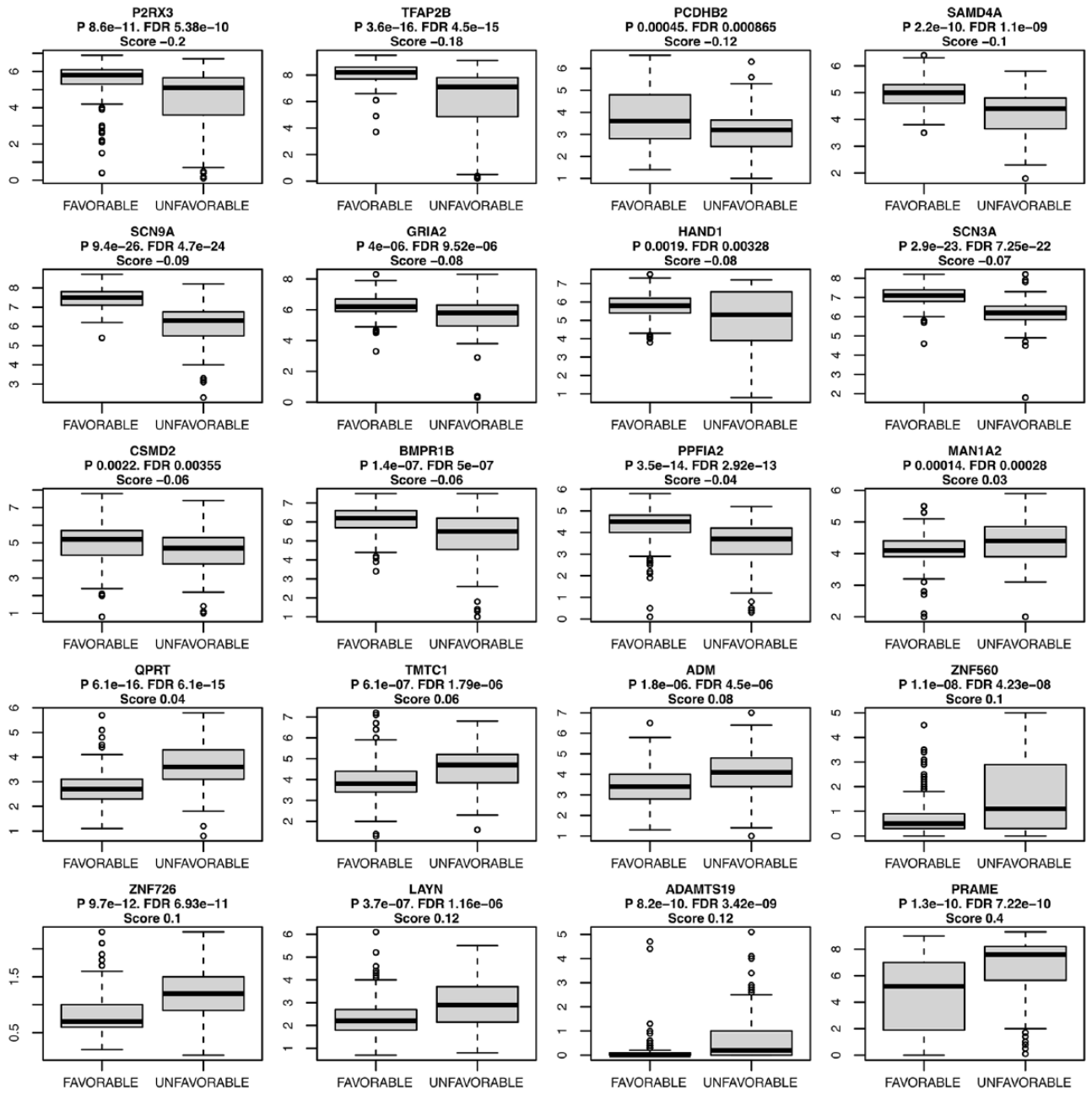

Figure 6. Boxplot of 20 genes found to be differentially expressed in unfavorable compared with favorable neuroblastoma tumors.
Boxplots of genes were plotted in order of the magnitude of the loading vector (score) from the X-variate component 1 in PLS-DA. Differential expression was analyzed using a Mann-Whitney test. *P*-values were corrected for multiple testing using the Benjamini-Hochberg procedure. Y-axes indicate normalized expression (log2 RPM) within favorable (n = 91) and favorable (n = 181) neuroblastoma patients derived from the RNA-seq dataset SEQC (GSE62564).

we show high *EZH2* expression in all neuroblastoma cell lines used in this study as well as upregulated expression in stage 4 recurrent cell lines (Supplemental file 4). *EZH2* is a member of the polycomb repressive complex 2 (PRC2) together with two additional members: embryonic ectoderm development (*EED*) and suppressor of zeste 12 (*SUZ12*).[72] *EZH2* acts as a histone methyltransferase and is generally associated with H3K27 trimethylation and gene silencing. In line with the upregulation of *EZH2* expression in recurrent cell lines, high expression of *EZH2* is correlated with an unfavorable prognosis in neuroblastoma and knockdown of *EZH2* significantly induced neuroblastoma cell differentiation.[73] Other recent studies have shown *EZH2* to directly bind to the promoter regions of certain genes and act as a transcriptional co-activator independent of its histone methyltransferase activity.[74–76]

As stated in the previous section, we found high expression of *EZH2* in our cell lines representative of MNA neuroblastoma (Supplemental file 4, Figure 3). In line with this observation, N-Myc, an established prognostic factor in neuroblastoma,[77] has been shown to bind the *EZH2* promoter to enhance its expression.[78] *EZH2* also interacts with the Myc box domain 3, a segment of *MYC* known to be essential for its transforming capacity.[79] Moreover, from a collection of 341 cancer cell lines from 26 tumor types investigated, MNA neuroblastoma cells

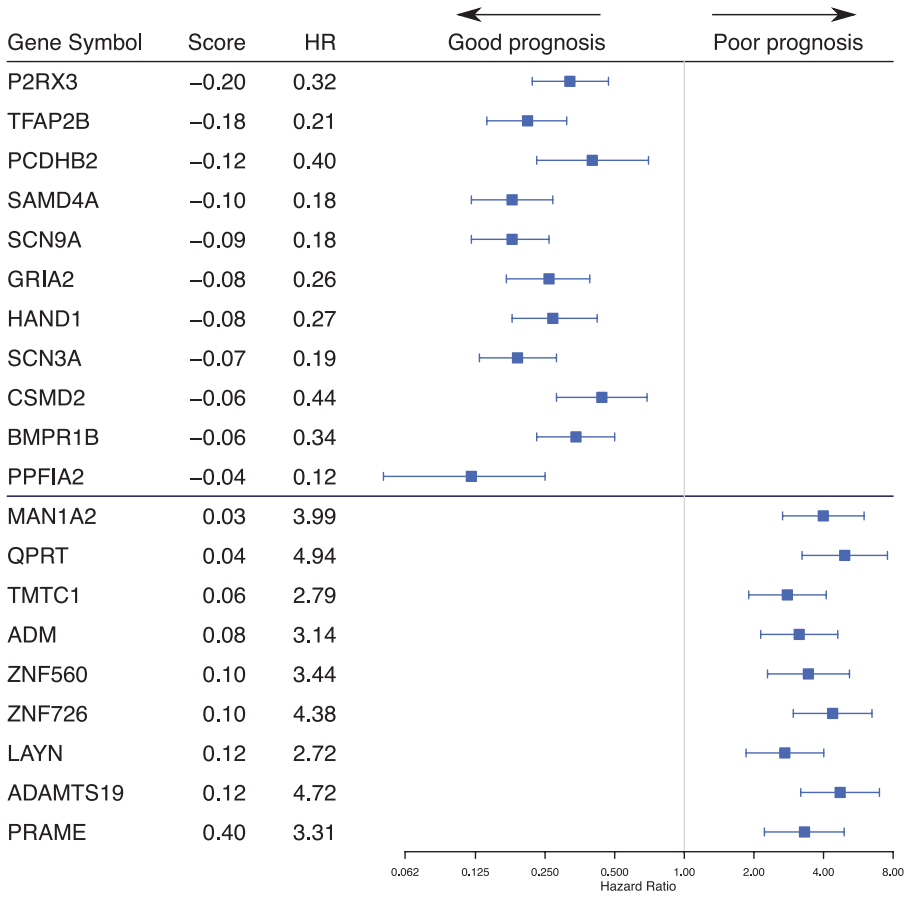

Figure 7. Forestplot with hazard ratio of 20 genes found to be differentially expressed in unfavorable compared with favorable neuroblastoma tumors. An increase in hazard ratios are associated with poor prognosis (ie, a hazard ratio of 0.5 or 2 signifies a twofold decrease or increase in death rate). All differentially expressed genes in favorable and primary, or unfavorable and recurrent tumors, showed at least a twofold decrease or increase in hazard ratios, respectively. Hazard ratios were based on the binary classification of overall survival of patients in the SEQC cohort. Accompanying Kaplan-Meier plots are provided in Supplemental file 6. Scale bars represent 95% confidence interval.
Score: Length of loading vector within the X-variate component 1 in PLS-DA. HR indicates hazard ratio.

revealed the highest dependency on the PRC2 complex components *EZH2, EED*, and *SUZ12*.[78] In the same study, suppression of *EZH2* expression in the MNA neuroblastoma cell lines SK-N-BE(2), Kelly and LAN-1 lead to reduced cell growth. Interestingly, several PcG proteins have been shown to be involved in stem cell maintenance and DNA damage response.[80]

TRIM28, another enriched transcription factor, interacts with *EZH2* to activate, rather than repress genes.[81] Intriguingly, we found a downregulation of *TRIM28* in recurrent cell lines and an inverse expression between *TRIM28* and *EZH2* (Supplemental file 4). *TRIM28* depletion has been shown to increase cell proliferation in breast and lung cancer.[82] Taken together, investigating the therapeutic potential of PcG proteins in high-risk neuroblastoma may hold promise in pharmacologic difficult targets such as *MYC* transcription factors.

Risk prediction based on gene signatures may prove to be more accurate when including a significant larger repertoire of genes such as ncRNAs. In this setting, we included and found a total of 17 lncRNAs able to discriminate primary and recurrent cell lines. Of these, 3 upregulated (*NEAT1, SH3BP5-AS1, NORAD*) and 3 downregulated lncRNAs (*DUBR, MEG3, DHRS4-AS1*) were also found to be differentially expressed in favorable compared with unfavorable outcome. *NEAT1* has been attributed as an oncogene promoting migration, invasion, paraspeckle formation, and cancer progression.[83–85] *NORAD* has been shown to assemble a topoisomerase complex critical for genomic stability, regulation of *PUM2* and as a marker for poor prognosis in some cancers.[86–91] In neuroblastoma, *MEG3* influences the proliferation and apoptosis via the HIF-1α and p53 pathways.[37]

We have previously mapped the expression levels of miRNAs contributing to tumor recurrence in the same cohort of cell lines.[17] In this study, we were able to elucidate the miRNA-mRNA regulatory axis by integrating both miRNA and mRNA expression data. To this extent, we found 31 of the previously 482 identified mRNA targets to display an inverse expression with their respective miRNAs (Table 2). *Mir-21*, one of the miRNAs previously identified, was shown to be upregulated after intensive treatment and tumor recurrence in all cell line pairs, except SK-N-BE(1) and SK-N-BE(2).[17] In this study, we extend our knowledge by showing that *mir-21* is a predicted target and inversely expressed with *TIMP3* (TIMP Metallopeptidase Inhibitor 3) and the p53 target gene p21/*CDKN1A* (Cyclin-Dependent Kinase Inhibitor 1A,

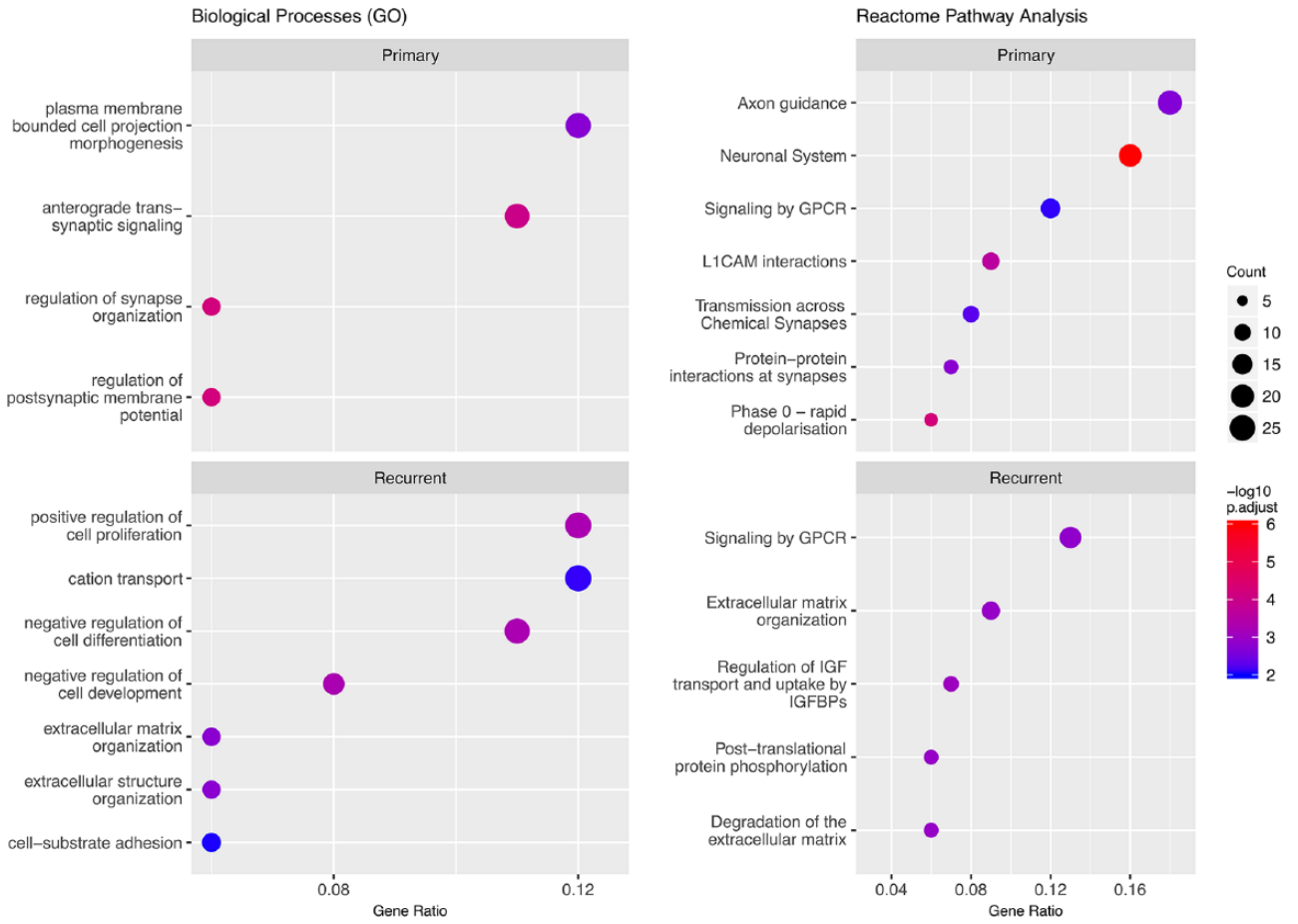

Figure 8. Gene set enrichment analysis representative of primary and recurrent cell lines.
X-axis represents number of enriched genes compared with total number of genes in each term. Dotsize and color represent number of enriched genes and –log10 FDR adjusted *P*-value in each term, respectively. Fisher's exact test was used to calculate enriched terms in gene sets within biological processes (left) and Reactome pathway (right). Enrichment was calculated individually for primary (upper panel, 201 genes) and recurrent (lower panel, 210 genes) cell lines based on genes with an absolute loading weight score higher than 0.02 in the first component of PLS-DA.

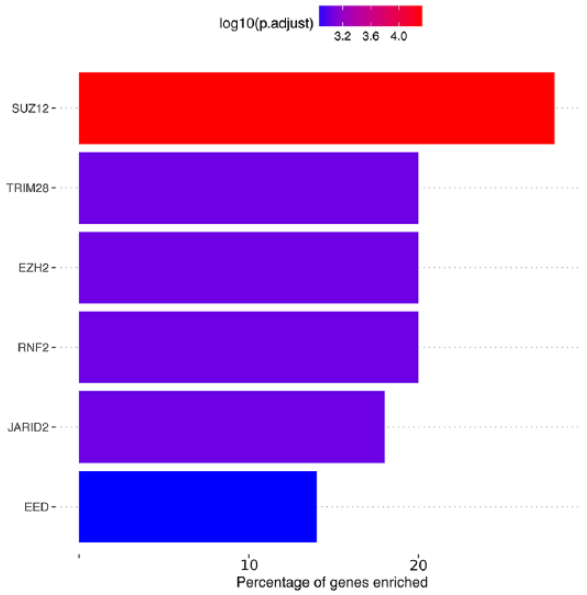

Figure 9. Transcription factor gene sets enriched in the 55 genes separating primary and recurrent tumors.
SUZ12, TRIM28, EZH2, RNF2, JARID2, and *EED* were identified as putative upstream master regulators of the 55 genes identified using transcription gene sets from ENCODE and ChEA.

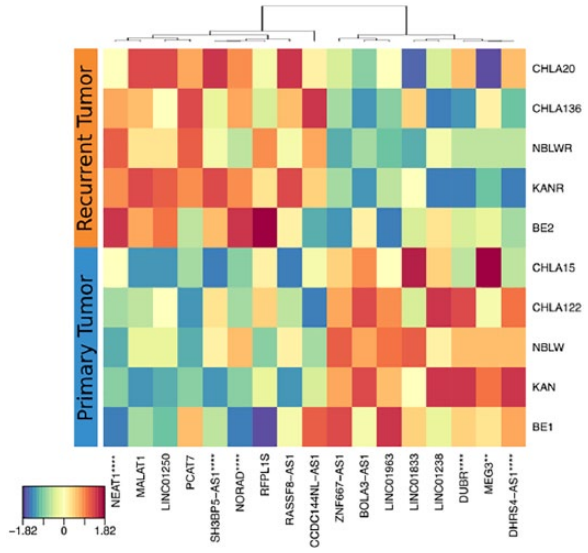

Figure 10. Clustered image map of long non-coding RNAs separating primary and recurrent tumors.
Color key represents Z-score. Genes denoted with asterisk were also found to be differentially expressed (adjusted *P*-values) in favorable compared with unfavorable patients. Differential expression was analyzed using a Mann-Whitney test. *P*-values were corrected for multiple testing using the Benjamini-Hochberg procedure. ****$P < .0001$, **$P < .01$.

Table 2. mRNAs targeted and inversely expressed with miRNAs in primary and recurrent cell lines.

TARGET	MIRNA	MIMAT ID	MIRNA FAMILY	PREDICTION SOURCE	MIRNA CHANGE	MRNA SCORE
TIMP3	21-5p	0000076	21-5p	Ingenuity	UP	−0.041
YIF1B	424-5p	0001341	16-5p	TarBase	UP	−0.021
CDKN1A	21-5p	0000076	21-5p	miRecords	UP	−0.019
WIPF1	424-5p	0001341	16-5p	miRecords	UP	−0.017
HMGA1	424-5p	0001341	16-5p	TargetScan	UP	−0.016
UBE2S	424-5p	0001341	16-5p	TarBase	UP	−0.015
FGFR1	424-5p	0001341	16-5p	Ingenuity	UP	−0.014
HDHD2	424-5p	0001341	16-5p	miRecords	UP	−0.013
NAPG	424-5p	0001341	16-5p	T&T	UP	−0.013
CARD8	424-5p	0001341	16-5p	miRecords	UP	−0.011
PPIF	424-5p	0001341	16-5p	T&T	UP	−0.011
BCL2	424-5p	0001341	16-5p	Ingenuity	UP	−0.01
RAB30	424-5p	0001341	16-5p	TarBase	UP	−0.01
SLC12A4	30e-5p	0000692	30c-5p	TarBase	DOWN	0.032
LAMC1	29b-3p	0000100	29b-3p	Ingenuity	DOWN	0.025
RHOC	138-5p	0000430	138-5p	TargetScan	DOWN	0.024
FAM3C	29b-3p	0000100	29b-3p	Ingenuity	DOWN	0.018
PPM1D	29b-3p	0000100	29b-3p	Ingenuity	DOWN	0.016
PTEN	29b-3p	0000100	29b-3p	Ingenuity	DOWN	0.016
PTEN	494-3p	0002816	494-3p	TargetScan	DOWN	0.016
SYPL1	30e-5p	0000692	30c-5p	T&T	DOWN	0.016
GALNT7	30e-5p	0000692	30c-5p	T&T	DOWN	0.014
KRAS	181a-5p	0000256	181a-5p	TargetScan	DOWN	0.014
VEGFA	34a-5p	0000255	34a-5p	miRecords	DOWN	0.014
TET1	29b-3p	0000100	29b-3p	TargetScan	DOWN	0.012
CCND1	34a-5p	0000255	34a-5p	Ingenuity	DOWN	0.01
CD47	34a-5p	0000255	34a-5p	Ingenuity	DOWN	0.01
CPNE8	30e-5p	0000692	30c-5p	T&T	DOWN	0.01
MBNL1	30e-5p	0000692	30c-5p	T&T	DOWN	0.01
TMED7	30e-5p	0000692	30c-5p	TarBase	DOWN	0.01
TMEM87A	30e-5p	0000692	30c-5p	T&T	DOWN	0.01

Targets were based on Table S3 from our previous study on miRNAs in neuroblastoma.[17] Only mRNA targets with absolute loading vector length (mRNA score) in the X-variate component 1 of PLS-DA higher than 0.01 were selected, yielding 13 downreugulated and 18 upregulated mRNAs. miRNA change up and a positive mRNA score represent miRNAs and mRNAs that were upregulated in recurrent cell lines, and vice versa. Target prediction sources represent Ingenuity Expert Findings (manually curated), miRecords (predicted and validated targets), TarBase (strong-evidence), and TargetScan (miRNA seed region-based match). T&T represents both TarBase and TargetScan. The list from Table S3 contains 482 targets, of which 324 and 31 were expressed and targeted. A complete list of targets may be found in Supplemental file 9.

Table 2). Interestingly, the p53 non-functional SK-N-BE(2)c cell line[92] expresses low levels of p21, as well as *TIMP3* (Supplemental file 4). As such, the role of *mir-21* may prove to be dependent on the status and expression levels of p53, *TIMP3*, and p21 in neuroblastoma cells.

Mir-424-5p is another miRNA upregulated in most recurrent cell lines (Table 2 from Roth et al,[17] Supplemental file 7). In this study, we report 11 inversely expressed, predicted targets of *mir-424-5p*, including *YIF1B, WIPF1, HMGA1, UBE2S, FGFR1, HDHD2, NAPG, CARD8, PPIF, BCL2*, and *RAB30*. Previous studies have shown *mir-424* acting both as tumor suppressor[93–95] and tumor promoter.[96,97] A very recent study has also linked *mir-424* to the regulation of *ALK*.[98] In our analysis, *ALK* was not part of the list of predicted targets and was therefore not included. With inverse expression between *mir-424* and *ALK* in 4 out of 5 cell line pairs used in this study (Supplemental files 4 and 7), our data support *ALK* as a target of *mir-424-5p* in neuroblastoma.

Conclusions

We found 20 mRNAs and 6 lncRNAs, which may be clinically relevant in the prediction of tumor recurrence and response to therapy in neuroblastoma patients. Moreover, we used miRNA expression data to investigate the miRNA-mRNA regulatory axis in primary and recurrent cell lines. The advantage of finding biomarkers in a clinically relevant neuroblastoma model system enables further studies on the effect of individual genes upon gene perturbation.

Acknowledgements

The NBL-W and NBL-WR cell lines were kindly provided by Professor Susan Cohn, University of Chicago, Chicago, USA. SK-N-BE(1), SK-N-BE(2), SMS-KAN, SMS-KANR, CHLA-15, CHLA-20, CHLA-122, and CHLA-136 were provided by Children's Oncology Group (COG) Cell Culture and Xenograft Repository, Texas Tech University Health Sciences Center, Lubbock, Texas, USA. The computations/simulations/[SIMILAR] were performed on resources provided by UNINETT Sigma2—the National Infrastructure for High Performance Computing and Data Storage in Norway. Special thanks to the High Performance Computing Group at UiT—The Arctic University of Norway for their support regarding this service.

Author Contributions

PU contributed to all of the work unless otherwise stated. PU also wrote the article. CL contributed to the passaging and maintenance of the cell lines used in this study, as well as support in experimental procedures. TF and CE contributed to the supervision and attaining of funds for this study, as well as major critical revisions of the manuscript. All authors reviewed and accepted the content of the final version of this article.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. *Lancet*. 2007;369:2106–2120.
2. De Bernardi B, Gerrard M, Boni L, et al. Excellent outcome with reduced treatment for infants with disseminated neuroblastoma without MYCN gene amplification. *J Clin Oncol*. 2009;27:1034–1040.
3. Jiang M, Stanke J, Lahti JM. The connections between neural crest development and neuroblastoma. In: Michael AD, ed. *Current Topics in Developmental Biology*. Vol 94. New York: Academic Press; 2011:77–127.
4. Cheung NK, Dyer MA. Neuroblastoma: developmental biology, cancer genomics and immunotherapy. *Nat Rev Cancer*. 2013;13:397–411.
5. Park JR, Bagatell R, London WB, et al. Children's Oncology Group's 2013 blueprint for research: neuroblastoma. *Pediatr Blood Cancer*. 2013;60:985–993.
6. Maris JM. Recent advances in neuroblastoma. *N Engl J Med*. 2010;362:2202–2211.
7. Schleiermacher G, Mosseri V, London WB, et al. Segmental chromosomal alterations have prognostic impact in neuroblastoma: a report from the INRG project. *Br J Cancer*. 2012;107:1418–1422.
8. Uryu K, Nishimura R, Kataoka K, et al. Identification of the genetic and clinical characteristics of neuroblastomas using genome-wide analysis. *Oncotarget*. 2017;8:107513–107529.
9. Matthay KK, Villablanca JG, Seeger RC, et al. Treatment of high-risk neuroblastoma with intensive chemotherapy, radiotherapy, autologous bone marrow transplantation, and 13-cis-retinoic acid. Children's Cancer Group. *N Engl J Med*. 1999;341:1165–1173.
10. Grimes T, Walker AR, Datta S, Datta S. Predicting survival times for neuroblastoma patients using RNA-seq expression profiles. *Biol Direct*. 2018;13:11.
11. Pugh TJ, Morozova O, Attiyeh EF, et al. The genetic landscape of high-risk neuroblastoma. *Nat Genet*. 2013;45:279–284.
12. Marachelian A, Villablanca JG, Liu CW, et al. Expression of five neuroblastoma genes in bone marrow or blood of patients with relapsed/refractory neuroblastoma provides a new biomarker for disease and prognosis. *Clin Cancer Res*. 2017;23:5374–5383.
13. Asgharzadeh S, Pique-Regi R, Sposto R, et al. Prognostic significance of gene expression profiles of metastatic neuroblastomas lacking MYCN gene amplification. *J Natl Cancer Inst*. 2006;98:1193–1203.
14. Schramm A, Schulte JH, Klein-Hitpass L, et al. Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. *Oncogene*. 2005;24:7902–7912.
15. Valentijn LJ, Koster J, Haneveld F, et al. Functional MYCN signature predicts outcome of neuroblastoma irrespective of MYCN amplification. *Proc Natl Acad Sci U S A*. 2012;109:19190–19195.
16. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
17. Roth SA, Knutsen E, Fiskaa T, et al. Next generation sequencing of microRNAs from isogenic neuroblastoma cell lines isolated before and after treatment. *Cancer Lett*. 2016;372:128–136.
18. Roth SA, Hald OH, Fuchs S, et al. MicroRNA-193b-3p represses neuroblastoma cell growth via downregulation of Cyclin D1, MCL-1 and MYCN. *Oncotarget*. 2018;9:18160–18179.
19. Haug BH, Hald OH, Utnes P, et al. Exosome-like extracellular vesicles from MYCN-amplified neuroblastoma cells contain oncogenic miRNAs. *Anticancer Res*. 2015;35:2521–2530.
20. Buechner J, Einvik C. N-myc and noncoding RNAs in neuroblastoma. *Mol Cancer Res*. 2012;10:1243–1253.
21. Buechner J, Tomte E, Haug BH, et al. Tumour-suppressor microRNAs let-7 and mir-101 target the proto-oncogene MYCN and inhibit cell proliferation in MYCN-amplified neuroblastoma. *Br J Cancer*. 2011;105:296–303.
22. Haug BH, Henriksen JR, Buechner J, et al. MYCN-regulated miRNA-92 inhibits secretion of the tumor suppressor DICKKOPF-3 (DKK3) in neuroblastoma. *Carcinogenesis*. 2011;32:1005–1012.
23. Buechner J, Henriksen JR, Haug BH, Tomte E, Flaegstad T, Einvik C. Inhibition of mir-21, which is up-regulated during MYCN knockdown-mediated differentiation, does not prevent differentiation of neuroblastoma cells. *Differentiation*. 2011;81:25–34.
24. Loven J, Zinin N, Wahlstrom T, et al. MYCN-regulated microRNAs repress estrogen receptor-alpha (ESR1) expression and neuronal differentiation in human neuroblastoma. *Proc Natl Acad Sci U S A*. 2010;107:1553–1558.
25. Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol*. 2011;21:354–361.

26. Huarte M. The emerging role of lncRNAs in cancer. *Nat Med*. 2015;21:1253–1261.
27. Marchese FP, Raimondi I, Huarte M. The multidimensional mechanisms of long noncoding RNA function. *Genome Biol*. 2017;18:206.
28. Mondal T, Juvvuna PK, Kirkeby A, et al. Sense-antisense lncRNA pair encoded by Locus 6p22.3 determines neuroblastoma susceptibility via the USP36-CHD7-SOX9 regulatory axis. *Cancer Cell*. 2018;33:417–434.e7.
29. Zhang J, Zhuo ZJ, Wang J, et al. CASC15 gene polymorphisms reduce neuroblastoma risk in Chinese children. *Oncotarget*. 2017;8:91343–91349.
30. Russell MR, Penikis A, Oldridge DA, et al. CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. *Cancer Res*. 2015;75:3155–3166.
31. Chalei V, Sansom SN, Kong L, et al. The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. *Elife*. 2014;3:e04530.
32. Mazar J, Rosado A, Shelley J, Marchica J, Westmoreland TJ. The long noncoding RNA GAS5 differentially regulates cell cycle arrest and apoptosis through activation of BRCA1 and p53 in human neuroblastoma. *Oncotarget*. 2017;8:6589–6607.
33. Voth H, Oberthuer A, Simon T, Kahlert Y, Berthold F, Fischer M. Identification of DEIN, a novel gene with high expression levels in stage IVS neuroblastoma. *Mol Cancer Res*. 2007;5:1276–1284.
34. Yang X, Wang CC, Lee WYW, Trovik J, Chung TKH, Kwong J. Long noncoding RNA HAND2-AS1 inhibits invasion and metastasis in endometrioid endometrial carcinoma through inactivating neuromedin U. *Cancer Lett*. 2018;413:23–34.
35. Atmadibrata B, Liu PY, Sokolowski N, et al. The novel long noncoding RNA linc00467 promotes cell survival but is down-regulated by N-Myc. *PLoS ONE*. 2014;9:e88112.
36. Liu PY, Erriquez D, Marshall GM, et al. Effects of a novel long noncoding RNA, lncUSMycN, on N-Myc expression and neuroblastoma progression. *J Natl Cancer Inst*. 2014;106:dju113.
37. Tang W, Dong K, Li K, Dong R, Zheng S. MEG3, HCN3 and linc01105 influence the proliferation and apoptosis of neuroblastoma cells via the HIF-1alpha and p53 pathways. *Sci Rep*. 2016;6:36268.
38. Pandey GK, Mitra S, Subhash S, et al. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell*. 2014;26:722–737.
39. Ishizuka A, Hasegawa Y, Ishida K, Yanaka K, Nakagawa S. Formation of nuclear bodies by the lncRNA Gomafu-associating proteins Celf3 and SF1. *Genes Cells*. 2014;19:704–721.
40. O'Brien EM, Selfe JL, Martins AS, Walters ZS, Shipley JM. The long noncoding RNA MYCNOS-01 regulates MYCN protein levels and affects growth of MYCN-amplified rhabdomyosarcoma and neuroblastoma cells. *BMC Cancer*. 2018;18:217.
41. Zhao X, Li D, Pu J, et al. CTCF cooperates with noncoding RNA MYCNOS to promote neuroblastoma progression through facilitating MYCN expression. *Oncogene*. 2016;35:3565–3576.
42. Vadie N, Saayman S, Lenox A, et al. MYCNOS functions as an antisense RNA regulating MYCN. *RNA Biol*. 2015;12:893–899.
43. Chen Y, Lian YJ, Ma YQ, Wu CJ, Zheng YK, Xie NC. LncRNA SNHG1 promotes α-synuclein aggregation and toxicity by targeting miR-15b-5p to activate SIAH1 in human neuroblastoma SH-SY5Y cells. *Neurotoxicology*. 2018;68:212–221.
44. Yu M, Ohira M, Li Y, et al. High expression of ncRAN, a novel non-coding RNA mapped to chromosome 17q25.1, is associated with poor prognosis in neuroblastoma. *Int J Oncol*. 2009;34:931–938.
45. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13:e1005752.
46. Meng C, Zeleznik OA, Thallinger GG, Kuster B, Gholami AM, Culhane AC. Dimension reduction techniques for the integrative analysis of multi-omics data. *Brief Bioinform*. 2016;17:628–641.
47. Westerhuis JA, van Velzen EJ, Hoefsloot HC, Smilde AK. Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics*. 2010;6:119–128.
48. Foley J, Cohn SL, Salwen HR, et al. Differential expression of N-myc in phenotypically distinct subclones of a human neuroblastoma cell line. *Cancer Res*. 1991;51:6338–6345.
49. Cohn SL, Herst CV, Maurer HS, Rosen ST. N-myc amplification in an infant with stage IVS neuroblastoma. *J Clin Oncol*. 1987;5:1441–1444.
50. Dobin A, Gingeras TR. Optimizing RNA-Seq mapping with STAR. *Methods Mol Biol*. 2016;1415:245–262.
51. Dobin A, Gingeras TR. Mapping RNA-seq reads with STAR. *Curr Protoc Bioinformatics*. 2015;51:11.14.1–19.
52. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
53. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–930.
54. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11:R106.
55. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)*. 2016;2016:baw100.
56. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–W97.
57. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
58. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–15550.
59. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1:417–425.
60. Barker M, Rayens W. Partial least squares for discrimination. *J Chemometr*. 2003;17:166–173.
61. Wang C, Gong B, Bushel PR, et al. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat Biotechnol*. 2014;32:926–932.
62. Consortium SM-I. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol*. 2014;32:903–914.
63. Munro SA, Lund SP, Pine PS, et al. Assessing technical performance in differential gene expression experiments with external spike-in RNA control ratio mixtures. *Nat Commun*. 2014;5:5125.
64. Su Z, Fang H, Hong H, et al. An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era. *Genome Biol*. 2014;15:523.
65. Shimada H, Ambros IM, Dehner LP, et al. The international neuroblastoma pathology classification (the Shimada system). *Cancer*. 1999;86:364–372.
66. Shimada H, Umehara S, Monobe Y, et al. International neuroblastoma pathology classification for prognostic evaluation of patients with peripheral neuroblastic tumors: a report from the Children's Cancer Group. *Cancer*. 2001;92:2451–2461.
67. Keshelava N, Seeger RC, Reynolds CP. Drug resistance in human neuroblastoma cell lines correlates with clinical therapy. *Eur J Cancer*. 1997;33:2002–2006.
68. Keshelava N, Zuo JJ, Chen P, et al. Loss of p53 function confers high-level multidrug resistance in neuroblastoma cell lines. *Cancer Res*. 2001;61:6185–6193.
69. Oberthuer A, Hero B, Spitz R, Berthold F, Fischer M. The tumor-associated antigen PRAME is universally expressed in high-stage neuroblastoma and associated with poor outcome. *Clin Cancer Res*. 2004;10:4307–4313.
70. Spel L, Boelens JJ, van der Steen DM, et al. Natural killer cells facilitate PRAME-specific T-cell reactivity against neuroblastoma. *Oncotarget*. 2015;6:35770–35781.
71. Ikram F, Ackermann S, Kahlert Y, et al. Transcription factor activating protein 2 beta (TFAP2B) mediates noradrenergic neuronal differentiation in neuroblastoma. *Mol Oncol*. 2016;10:344–359.
72. Sauvageau M, Sauvageau G. Polycomb group proteins: multi-faceted regulators of somatic stem cells and cancer. *Cell Stem Cell*. 2010;7:299–313.
73. Li Z, Takenobu H, Setyawati AN, et al. EZH2 regulates neuroblastoma cell differentiation via NTRK1 promoter epigenetic modifications. *Oncogene*. 2018;37:2714–2727.
74. Yan J, Ng SB, Tay JL, et al. EZH2 overexpression in natural killer/T-cell lymphoma confers growth advantage independently of histone methyltransferase activity. *Blood*. 2013;121:4512–4520.
75. Xu K, Wu ZJ, Groner AC, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science*. 2012;338:1465–1469.
76. Kim KH, Kim W, Howard TP, et al. SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nat Med*. 2015;21:1491–1496.
77. Zeid R, Lawlor MA, Poon E, et al. Enhancer invasion shapes MYCN-dependent transcriptional amplification in neuroblastoma. *Nat Genet*. 2018;50:515-523.
78. Chen L, Alexe G, Dharia NV, et al. CRISPR-Cas9 screen reveals a MYCN-amplified neuroblastoma dependency on EZH2. *J Clin Invest*. 2018;128:446–462.
79. Corvetta D, Chayka O, Gherardi S, et al. Physical interaction between MYCN oncogene and polycomb repressive complex 2 (PRC2) in neuroblastoma: functional and therapeutic implications. *J Biol Chem*. 2013;288:8332–8341.
80. Gieni RS, Ismail IH, Campbell S, Hendzel MJ. Polycomb group proteins in the DNA damage response: a link between radiation resistance and "stemness." *Cell Cycle*. 2011;10:883–894.
81. Li J, Xi Y, Li W, et al. TRIM28 interacts with EZH2 and SWI/SNF to activate genes that promote mammosphere formation. *Oncogene*. 2017;36:2991–3001.
82. Chen L, Chen DT, Kurtyka C, et al. Tripartite motif containing 28 (Trim28) can regulate cell proliferation by bridging HDAC1/E2F interactions. *J Biol Chem*. 2012;287:40106–40118.
83. Li P, Huang R, Huang T, Cheng S, Chen Y, Wang Z. Long non-coding RNA NEAT1 promotes proliferation, migration and invasion of human osteosarcoma cells. *Int J Med Sci*. 2018;15:1227–1234.

84. Liu X, Shang W, Zheng F. Long non-coding RNA NEAT1 promotes migration and invasion of oral squamous cell carcinoma cells by sponging microRNA-365. *Exp Ther Med*. 2018;16:2243–2250.
85. Chen Q_, Cai J, Wang Q_, et al. Long noncoding RNA NEAT1, regulated by the EGFR pathway, contributes to glioblastoma progression through the WNT/beta-catenin pathway by scaffolding EZH2. *Clin Cancer Res*. 2018;24:684–695.
86. Munschauer M, Nguyen CT, Sirokman K, et al. The NORAD lncRNA assembles a topoisomerase complex critical for genome stability. *Nature*. 2018;561:132–136.
87. Huo H, Tian J, Wang R, Li Y, Qu C, Wang N. Long non-coding RNA NORAD upregulate SIP1 expression to promote cell proliferation and invasion in cervical cancer. *Biomed Pharmacother*. 2018;106:1454–1460.
88. Li Q_, Li C, Chen J, et al. High expression of long noncoding RNA NORAD indicates a poor prognosis and promotes clinical progression and metastasis in bladder cancer. *Urol Oncol*. 2018;36:310.e15–310.e22.
89. Tichon A, Perry RB, Stojic L, Ulitsky I. SAM68 is required for regulation of Pumilio by the NORAD long noncoding RNA. *Genes Dev*. 2018;32:70–78.
90. Wu X, Lim ZF, Li Z, et al. NORAD expression is associated with adverse prognosis in esophageal squamous cell carcinoma. *Oncol Res Treat*. 2017;40:370–374.
91. Lee S, Kopp F, Chang TC, et al. Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins. *Cell*. 2016;164:69–80.
92. Gogolin S, Ehemann V, Becker G, et al. CDK4 inhibition restores G(1)-S arrest in MYCN-amplified neuroblastoma cells in the context of doxorubicin-induced DNA damage. *Cell Cycle*. 2013;12:1091–1104.
93. Liu Q_, Fu H, Sun F, et al. miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Res*. 2008;36:5391–5404.
94. Xu J, Li Y, Wang F, et al. Suppressed miR-424 expression via upregulation of target gene Chk1 contributes to the progression of cervical cancer. *Oncogene*. 2013;32:976–987.
95. Ruiz-Llorente L, Ardila-Gonzalez S, Fanjul LF, Martinez-Iglesias O, Aranda A. microRNAs 424 and 503 are mediators of the anti-proliferative and anti-invasive action of the thyroid hormone receptor beta. *Oncotarget*. 2014;5:2918–2933.
96. Wu K, Hu G, He X, et al. MicroRNA-424-5p suppresses the expression of SOCS6 in pancreatic cancer. *Pathol Oncol Res*. 2013;19:739–748.
97. Zhang Y, Li T, Guo P, et al. MiR-424-5p reversed epithelial-mesenchymal transition of anchorage-independent HCC cells by directly targeting ICAT and suppressed HCC progression. *Sci Rep*. 2014;4:6248.
98. De Mariano M, Stigliani S, Moretti S, et al. A genome-wide microRNA profiling indicates miR-424-5p and miR-503-5p as regulators of ALK expression in neuroblastoma. *Oncotarget*. 2017;8:56518–56532.