

Diagnosis of the Small Round Blue Cell Tumors Using Multiplex Polymerase Chain Reaction

Qing-Rong Chen,^{*†} Gordon Vansant,[‡]
Kahuku Oades,[‡] Maria Pickering,[‡] Jun S. Wei,^{*}
Young K. Song,^{*} Joseph Monforte,[‡] and
Javed Khan^{*}

From the Oncogenomics Section,^{*} Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute, Gaithersburg, Maryland; the Advanced Biomedical Computing Center,[‡] SAIC-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland; and Althea Technologies,[‡] San Diego, California

The small round blue cell tumors of childhood, which include neuroblastoma, rhabdomyosarcoma, non-Hodgkin's lymphoma, and the Ewing's family of tumors, are so called because of their similar appearance on routine histology. Using cDNA microarray gene expression profiles and artificial neural networks (ANNs), we previously identified 93 genes capable of diagnosing these cancers. Using a subset of these, together with some additional genes (total 39), we developed a multiplex polymerase chain reaction (PCR) assay to diagnose these cancer types. Blinded testing of 96 new samples (26 Ewing's family of tumors, 29 rhabdomyosarcomas, 24 neuroblastomas, and 17 lymphomas) using ANNs in a complete leave-one-out analysis demonstrated that all except one sample were accurately diagnosed as their respective category. Moreover, using an ANN-based gene minimization strategy in a separate analysis, we found that the top 31 genes could correctly diagnose all 96 tumors. Our results suggest that this molecular test based on a multiplex PCR reaction may assist the physician in the rapid confirmation of the diagnosis of these cancers. (*J Mol Diagn* 2007, 9:80–88; DOI: 10.2353/jmoldx.2007.060111)

The highly malignant small round blue cell tumors (SR-BCTs) occur in the pediatric, adolescent, and young adult populations in some cases. Accurate diagnosis of these cancers, which include neuroblastoma (NB), rhabdomyosarcoma (RMS), non-Hodgkin's lymphoma, and the Ewing's family of tumors (EWS), is essential because the treatment options, responses to therapy, and prognoses vary widely depending on the diagnosis. As their name implies, these cancers are difficult to distinguish by light microscopy, and currently no single test can precisely distinguish these cancers. To confirm the diagno-

sis, pathologists rely on several techniques, including immunohistochemistry,¹ cytogenetics, interphase fluorescence *in situ* hybridization,² and reverse transcription-polymerase chain reaction (RT-PCR).³ Immunohistochemistry for individual protein markers is used to establish the diagnosis in many instances, but it can only examine a single protein at a time. Molecular techniques, such as RT-PCR of tumor-specific translocations, are used for the diagnosis of EWS containing the *EWS-FLI1* and alveolar rhabdomyosarcoma containing the *PAX3-FKHR*, but molecular markers do not always provide a definitive diagnosis because of either technical difficulties or the presence of variant translocations. Using cDNA microarray gene expression profiling and artificial neural networks (ANNs), we previously identified 93 genes as gene expression signature that was capable of presenting these SRBCTs to specific diagnostic categories.⁴ In this study, we have developed a reliable multiplex RT-PCR assay for the rapid diagnosis of these cancers using genes known to be differentially expressed in these cancers.

Materials and Methods

Tumor Samples

The source and other information for 96 tumor samples used in this study are described in Table 1. All of the original histological diagnoses were made at tertiary hospitals, which have reference diagnostic laboratories with extensive experience in the diagnosis of pediatric cancers. The EWSs ($n = 26$) had a spectrum of the expected translocations containing both primary tumor (labeled as EWS_T in Table 1) and xenografts derived from cell lines (EWS_X), and the RMSs ($n = 29$) were a mixture of

Supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research and in whole or in part by federal funds from the National Cancer Institute, National Institutes of Health, under contract N01-CO-12400.

Accepted for publication September 15, 2006.

The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

Address reprint requests to Javed Khan, M.D., Oncogenomics Section, Pediatric Oncology Branch, Advanced Technology Center, National Cancer Institute, 8717 Grovemont Circle, Gaithersburg, MD 20877, or Joseph Monforte, Ph.D., Althea Technologies, 11040 Roselle St., San Diego, CA 92121-1205. E-mail: khanjav@mail.nih.gov and jmonforte@altheatech.com.

alveolar RMS containing the *PAX3-FKHR* translocation (RMS_A), embryonal RMS (RMS_E), and botryoid subtype (RMS_B). The NBs ($n = 24$) contained both *MYCN* amplified and single-copy samples. The lymphoma ($n = 17$) contained several different types as listed in Table 1. RNA extraction was described previously.⁴

Primer Design

The genes and primers used in the multiplex RT-PCR reaction are described in Table 2. Each reverse primer is chimeric with the 5' end containing a 19-nucleotide universal priming sequence and the 3' end containing the gene-specific sequence (typically around 20 nucleotides). Each forward primer is chimeric with the 5' end containing a second, 18-nucleotide universal forward priming sequence and the 3' end containing the gene-specific sequence. Each of the primer pairs was designed to yield PCR products 4 to 7 bp apart, ranging from 137 to 300 bp. Primer design and multiplex optimization was performed using GeXP Express Profiler, Primer Design module (Beckman, Fullerton, CA). Primers were also designed to amplify from a kanamycin RNA transcript that is spiked into each reaction as an external control. Included in the PCR reaction are the two universal primers that are homologous to the 5' ends of the chimeric primers with the forward universal primer carrying the D4 dye label. The universal primers are included in the 5× GeXP PCR buffer.

Multiplex RT-PCR

The method is summarized in Figure 1. The gene expression patterns of multiple genes were examined from each of the above samples to the GenomeLab GeXP Analysis System Multiplex RT-PCR assay (Beckman). For each reaction, 3 μl of RNA was mixed with 1.5 μl of 10× DNase Buffer (Ambion, Austin, TX), 0.5 μl of DNase (Ambion), and 10 μl of dH_2O and incubated at 37°C for 20 minutes. Then, 1 μl of 25 mmol/L ethylenediamine tetraacetic acid was added to each reaction and was incubated at 70°C for 5 minutes. The RNA samples were then diluted down to a concentration of 5 ng/ μl . The chimeric primers were divided into two gene sets (PCR set in Table 2), and the multiplex RT-PCR was done separately for each set. In brief, 25 ng of RNA from each sample was reverse transcribed with both sets of chimeric reverse primers in individual reactions. The reverse transcription reactions were performed according to GeXP Start kit protocol using the following kit reagents: 5× RT Master Mix buffer, 1 μl of Moloney murine leukemia virus reverse transcriptase, and KanR/RI reagent, an internal reaction integrity control. The concentration of each primer varied from 0.003 to 0.05 $\mu\text{mol/L}$ to adjust the final signals of each amplified gene. The reverse transcription reactions were incubated for 1 minute at 48°C, 5 minutes at 37°C, 60 minutes at 42°C, and then 5 minutes at 95°C. The 20- μl reactions were performed in a Thermo-Fast 96-well PCR Detection Plate (ABgene, Epsom, Surrey, UK). GeXP Multiplex PCR was then performed on each sample as follows. An aliquot of 10 μl of cDNA from each above reverse

transcription reaction was added to the wells of a new 96-well PCR plate and 10 μl of a PCR reaction mix containing the each set of chimeric forward primers at 2 $\mu\text{mol/L}$ each (multiplex forward primer mixtures), GeXP 5× PCR buffer Master Mix, which contains the D4 dye-labeled forward universal primer and unlabeled reverse universal primer, 7 mmol/L MgCl_2 (USB, Cleveland, OH), and 3.5 units of *Taq* Polymerase (ABgene). The reactions were first subjected to 95°C for 10 minutes followed by 35 PCR cycles. Each PCR cycle consisted of the following conditions: 94°C for 30 seconds, 55°C for 30 seconds, and 68°C for 1 minute. The PCR products for each set were then prepared for capillary electrophoresis by adding 1 μl of each reaction to its corresponding well in a Beckman 96-well CEQ electrophoresis plate (Beckman) containing 39 μl of CEQ Sample Loading Solution (Beckman) and 0.5 μl of CEQ DNA Size Standard 400 (Beckman) per each reaction. The samples were mixed and placed in a GeXP Genetic Analysis System for capillary electrophoresis and fragment size analysis. The fragment results were analyzed on the eXpress Analysis module of the GeXP Genetic Analysis System. This software associates each PCR product with its corresponding gene and reports its peak area.

Data Analysis

We have chosen the housekeeping gene *PPIA* as the control gene.⁵ The gene expression data obtained from multiplex RT-PCR were normalized by dividing the peak area result of each gene by the peak area result of *PPIA* and were then \log_2 -transformed. Because there are two multiplex assays for each experiment, we combined the normalized data from both assays. For ANNs, we used feed-forward resilient back-propagation multilayer perceptron artificial neural networks⁶ with three layers: an input layer of the normalized expression ratio data, a hidden layer with three nodes, and an output layer generating a committee vote that discriminates four classes (EWS, RMS, NB, and lymphoma; Figure 2A). For each diagnostic category, each ANN model gave an output between 0 (not this category) and 1 (this category). Average artificial neural network committee votes were used to classify samples. The sample is classified as a particular cancer if it receives the highest committee vote for that cancer. We performed a leave-one-out prediction strategy, where we left out each sample (of the 96 unique samples) one time during the training of artificial neural networks and tested it as an independent sample to predict the diagnosis.

Results

Development of Multiplex RT-PCR Assay and Gene Selection

We developed the assay that combined the multiplex RT-PCR and fluorescence capillary electrophoresis techniques as illustrated in Figure 1. Two types of primers were designed for multiplex PCR amplification: chimeric primers and universal primers (see primer design in Materials and Methods for details). After reverse transcription of RNA, the

Table 1. Tumor Sample Information and ANN Diagnosis Prediction

Samples	ANN diagnosis	Histological diagnosis	Source label	Source	ANN committee vote			Lymphoma
					EWS	RMS	NB	
EWS_T_1	EWS	EWS	9512P350SP3	CHTN	0.96	0.03	0.03	0.04
EWS_T_2	EWS	EWS	9601P007SP1	CHTN	0.91	0.10	0.02	0.04
EWS_T_3	EWS	EWS	9607P075	CHTN	0.96	0.02	0.03	0.04
EWS_T_4	EWS	EWS	9704P013	CHTN	0.96	0.03	0.03	0.03
EWS_T_5	EWS	EWS	9706P044	CHTN	0.96	0.03	0.04	0.03
EWS_T_6	EWS	EWS	9708P076	CHTN	0.96	0.02	0.06	0.03
EWS_T_7	EWS	EWS	9810P202	CHTN	0.95	0.03	0.02	0.07
EWS_T_8	EWS	EWS	9904P6008	CHTN	0.94	0.06	0.02	0.04
EWS_T_9	EWS	EWS	9910P6003	CHTN	0.95	0.02	0.06	0.05
EWS_T_10	EWS	EWS	740987	CHTN	0.95	0.03	0.03	0.05
EWS_T_11	EWS	EWS	750880	CHTN	0.93	0.03	0.06	0.02
EWS_T_12	EWS	EWS	753118	CHTN	0.94	0.03	0.04	0.04
EWS_T_13	EWS	EWS	453034	CHTN	0.97	0.03	0.03	0.03
EWS_T_14	EWS	EWS	718336	CHTN	0.96	0.02	0.04	0.03
EWS_T_15	EWS	EWS	520645	CHTN	0.94	0.05	0.04	0.02
EWS_T_16	EWS	EWS	492359	CHTN	0.95	0.07	0.02	0.03
EWS_T_17	EWS	EWS	682034	CHTN	0.95	0.02	0.05	0.04
EWS_T_18	EWS	EWS	714795	CHTN	0.96	0.03	0.03	0.03
EWS_T_19	EWS	EWS	98-10-A040A	CHTN	0.91	0.06	0.03	0.03
EWS_X_1	EWS	EWS xenograft	CB-AGPN	CHLA	0.90	0.12	0.05	0.03
EWS_X_2	EWS	EWS xenograft	CHP-100	CHLA	0.90	0.06	0.04	0.10
EWS_X_3	EWS	EWS xenograft	SK-N-MC	CHLA	0.95	0.03	0.03	0.05
EWS_X_4	EWS	EWS xenograft	TC-268	CHLA	0.93	0.02	0.07	0.09
EWS_X_5	EWS	EWS xenograft	TC-32	CHLA	0.96	0.02	0.11	0.05
EWS_X_6	EWS	EWS xenograft	TC-71	CHLA	0.93	0.04	0.03	0.05
EWS_X_7	EWS	EWS xenograft	SK-PN-DW	CHLA	0.87	0.03	0.08	0.04
RMS_A_1	RMS	ARMS	RH4	SJCRH	0.04	0.97	0.03	0.03
RMS_A_2	RMS	ARMS	RH30	SJCRH	0.01	0.93	0.15	0.07
RMS_E_3	RMS	ERMS	RD	SJCRH	0.05	0.95	0.03	0.03
RMS_A_4	RMS	ARMS	200002P2054	CHTN	0.03	0.97	0.04	0.03
RMS_A_5	RMS	ARMS	200002P2065	CHTN	0.03	0.97	0.03	0.03
RMS_E_6	RMS	ERMS	200003P2080	CHTN	0.03	0.97	0.03	0.04
RMS_A_7	RMS	ARMS	200003P4067	CHTN	0.03	0.97	0.03	0.03
RMS_A_8	RMS	ARMS	200004P2174	CHTN	0.03	0.97	0.03	0.03
RMS_E_9	RMS	ERMS	9911P1241	CHTN	0.04	0.96	0.03	0.03
RMS_A_10	RMS	ARMS	9709P144	CHTN	0.04	0.96	0.02	0.05
RMS_A_11	RMS	ARMS	200008P6027	CHTN	0.02	0.95	0.05	0.11
RMS_A_12	RMS	ARMS	200009P4233	CHTN	0.02	0.96	0.05	0.03
RMS_A_13	RMS	ARMS	200010P1258	CHTN	0.07	0.95	0.02	0.03
RMS_E_14	RMS	ERMS	200104P4055	CHTN	0.03	0.97	0.03	0.03
RMS_E_15	RMS	ERMS	9701P126	CHTN	0.06	0.94	0.04	0.03
RMS_B_16	RMS	BRMS	9704P209	CHTN	0.03	0.97	0.03	0.03
RMS_E_17	EWS	ERMS	200004P2015	CHTN	0.49	0.30	0.18	0.01
RMS_A_18	RMS	ARMS	200006P2010	CHTN	0.04	0.97	0.03	0.03
RMS_A_19	RMS	ARMS	200006P2010	CHTN	0.03	0.97	0.03	0.03
RMS_A_20	RMS	ARMS	200007P1049	CHTN	0.05	0.97	0.03	0.03
RMS_A_21	RMS	ARMS	200007P1049	CHTN	0.03	0.97	0.03	0.04
RMS_E_22	RMS	ERMS	9609P032	CHTN	0.03	0.96	0.03	0.05
RMS_A_23	RMS	ARMS	9807P117	CHTN	0.08	0.95	0.02	0.05
RMS_A_24	RMS	ARMS	9807P117	CHTN	0.02	0.96	0.04	0.05
RMS_A_25	RMS	ARMS	9807P332	CHTN	0.03	0.97	0.04	0.03
RMS_A_26	RMS	ARMS	9808P189	CHTN	0.04	0.95	0.04	0.03
RMS_E_27	RMS	ERMS	9705P060	CHTN	0.10	0.84	0.04	0.04
RMS_B_28	RMS	BRMS	9809P631	CHTN	0.03	0.96	0.03	0.03
RMS_A_29	RMS	ARMS	9903P605	CHTN	0.04	0.97	0.03	0.03
NB_1	NB	NB	99-12-P2020	CHTN	0.03	0.04	0.96	0.03
NB_2	NB	NB	2000-03-P1273	CHTN	0.03	0.03	0.96	0.06
NB_3	NB	NB	2000-03-P2226	CHTN	0.03	0.03	0.97	0.03
NB_4	NB	NB	2000-04-P2103X	CHTN	0.03	0.04	0.96	0.03
NB_5	NB	NB	2000-05-P4140	CHTN	0.04	0.03	0.97	0.04
NB_6	NB	NB	2000-08-P1148	CHTN	0.04	0.04	0.96	0.03
NB_7	NB	NB	2000-09-P4042	CHTN	0.03	0.03	0.97	0.03
NB_8	NB	NB	2000-10-P1300	CHTN	0.04	0.03	0.97	0.03
NB_9	NB	NB	2000-12-P4028	CHTN	0.04	0.04	0.97	0.03
NB_10	NB	NB	2001-02-P1214	CHTN	0.04	0.03	0.97	0.03
NB_11	NB	NB	2001-03-P8006	CHTN	0.05	0.02	0.96	0.06
NB_12	NB	NB	2001-05-P8013	CHTN	0.04	0.03	0.97	0.03
NB_13	NB	NB	2001-05-P8041	CHTN	0.04	0.03	0.97	0.03
NB_14	NB	NB	2001-06-P8007	CHTN	0.04	0.04	0.96	0.03

(Table continues)

Table 1. *Continued*

Samples	ANN diagnosis	Histological diagnosis	Source label	Source	ANN committee vote			
					EWS	RMS	NB	Lymphoma
NB_15	NB	NB	2001-10-P6139	CHTN	0.05	0.02	0.95	0.07
NB_16	NB	NB	2001-12-P4075	CHTN	0.05	0.03	0.97	0.03
NB_17	NB	NB	2002-07-P6055	CHTN	0.05	0.03	0.97	0.03
NB_s18	NB	NB	2002-07-P6098	CHTN	0.03	0.03	0.97	0.04
NB_19	NB	NB	2002-07-P6111	CHTN	0.02	0.04	0.96	0.05
NB_20	NB	NB	2002-07-P6120	CHTN	0.03	0.03	0.96	0.04
NB_21	NB	NB	96-04-P328	CHTN	0.04	0.04	0.96	0.03
NB_22	NB	NB	0000-07-P6112	CHTN	0.04	0.04	0.96	0.03
NB_23	NB	NB	0000-07-P9394	CHTN	0.04	0.03	0.96	0.03
NB_24	NB	NB	0000-07-P9404	CHTN	0.04	0.03	0.97	0.03
Lymph_1	Lymph	BL	9809P1009	CHTN	0.03	0.05	0.03	0.96
Lymph_2	Lymph	BL	9903P903	CHTN	0.03	0.07	0.04	0.95
Lymph_3	Lymph	BL	9711P411	CHTN	0.02	0.08	0.03	0.95
Lymph_4	Lymph	HD	9508P228	CHTN	0.05	0.05	0.03	0.95
Lymph_5	Lymph	LL	9808P272	CHTN	0.03	0.06	0.03	0.95
Lymph_6	Lymph	NHL	9508P413	CHTN	0.24	0.09	0.42	0.67
Lymph_7	Lymph	NHL	9509P834	CHTN	0.06	0.04	0.04	0.96
Lymph_8	Lymph	APLC	9603P340	CHTN	0.07	0.04	0.02	0.93
Lymph_9	Lymph	APLC	9612P204	CHTN	0.05	0.06	0.03	0.91
Lymph_10	Lymph	BL	9704P100	CHTN	0.04	0.04	0.04	0.96
Lymph_11	Lymph	BL	9802P183	CHTN	0.04	0.09	0.04	0.95
Lymph_12	Lymph	BL	200005P6002	CHTN	0.03	0.05	0.05	0.96
Lymph_13	Lymph	SNCC	9504P051	CHTN	0.04	0.11	0.03	0.95
Lymph_14	Lymph	LCL	9508P351	CHTN	0.17	0.08	0.02	0.87
Lymph_15	Lymph	NHHL	9801P612	CHTN	0.43	0.15	0.01	0.70
Lymph_16	Lymph	NHL	9808P320	CHTN	0.10	0.05	0.04	0.95
Lymph_17	Lymph	SNCC	9712P137	CHTN	0.03	0.07	0.14	0.95

Source label refers to the original name of the sample as designated by the source. Histological diagnosis is defined as cancer type.

CHTN, Cooperative Human Tissue Network; CHLA, Children's Hospital Los Angeles; SJCRH, St. Jude Children's Research Hospital.

Lymphoma categories: APLC, anaplastic large cell; ARMS, alveolar rhabdomyosarcoma; ERMS, embryonal rhabdomyosarcoma; BRMS, rhabdomyosarcoma of botryoid subtype, BL, Burkitt's; HD, Hodgkin's; LCL, large cell; LL, lymphoblastic; NHL, non-Hodgkin's unknown; NHHL, non-Hodgkin's histiocytic; SNCC, small noncleaved cell.

chimeric primers containing gene-specific sequence and universal primer sequence are used at the first stage of amplification. The second stage converts to use one pair of universal primers for the multiplex amplification (Figure 1A). The obtained mixture of amplicons was then analyzed using fluorescence capillary electrophoresis to identify the peak location (gene identity) and peak fluorescence intensity (gene expression level) (Figure 1B). We show here an example of comparative chromatograms of four different categories of tumors (one in each category) from the one multiplex assay (Figure 1C). For most of the genes (peak location), we observed the differential gene expression (peak area).

We have previously used cDNA microarray gene expression profiling and ANN to identify 93 genes capable of diagnosing the SRBCTs to specific diagnostic categories.⁴ A subset of these genes ($n = 34$), which are specific to each diagnostic category, were selected for multiplex RT-PCR as seen in Table 2. In addition, several genes ($n = 5$) differentially expressed in NB tumors and normal tissues from Son et al⁷ and an unpublished result were also included in the assay (Table 2).

Diagnostic Classification

We applied ANN models to diagnose and classify tumors in each of the four SRBCT categories and used gene-expression data from multiplex RT-PCR results contain-

ing 39 genes of all 96 samples (26 EWSs, 29 RMSs, 24 NBs, and 17 lymphomas) as inputs for ANN as shown in Figure 2A. We performed a leave-one-out prediction strategy, where we left out each sample (of the 96 samples) one time during the training of artificial neural networks and tested it as an independent sample to predict the diagnostic category of tumors. A sample is classified to a diagnostic category if it receives the highest vote for that category, and because the classifier has only four possible outputs, all samples will be classified to one of the four categories. We found that the artificial neural networks correctly predicted 95 of 96 samples except one RMS sample, which was misclassified to EWS (Figure 2B; Table 1). The sensitivity of the ANN models (leave-one-out strategy) for diagnostic classification was 96.6% for RMS and 100% for EWS, NB, and lymphoma; the specificity was 98.6% for EWS and 100% for the rest of categories; the positive predictive value was 96.3% for EWS and 100% for others; the negative predictive value was 98.5% for RMS and 100% for the rest of categories (Table 3).

Minimization of Genes Used for Diagnostic Assay

To identify the optimal set of genes, resulting in minimal classification errors that construct the final diagnostic

Table 2. Genes and Primers Used in the Multiplex RT-PCR

Gene symbol	RefSeq	Diagnosis	PCR set	Reverse primer w/o universal tags	Forward primer w/o universal tags	Size of product
FGFR4*	NM_002011	RMS	Set 1	GGATCAGGTAGGAAGCTGGCAAT	AGCAGGAGCTGACAGTAGCC	194
SGCA*	NM_000023	RMS	Set 2	GGAAGTGCTGGTCCAGAATGAGG	CGACATCCAGATGGTCCAC	219
PTPN13*	NM_080685	EWS	Set 1	GGATTGCCTTCAAAGGGTGTTTC	GGTTTGAATCCAGCAGTGGT	137
TNNT1*	NM_003283	RMS	Set 2	GGAGCTTCTGTTCTGCCTTGACC	GCGCTTCAGAACTGAGAAGG	210
FVT1*	NM_002035	EWS	Set 2	GGAAGCATGTACCCATCTGAGCC	ACCCACCAGACACAGACACA	215
TNNT2*	NM_000364	RMS	Set 2	GGAACCTCTCTCCATCGGGGAT	AGAGGATGCTGAAGCAGAGG	205
CAV1*	NM_001753	EWS	Set 2	GGAAGTTCGATCTCCTTGGTGTG	TCTCTACACCGTCCCATCC	154
PIM2*	NM_006875	BL	Set 2	GGAAGGGCGTAGGTCTATCAGGA	TATCACAGAGAAGGGCCAC	179
PHOX2B**	NM_003924	NB	Set 1	GGATGGATAGGGCATCTTTCAGG	GGATTCTCTTGGGGCTCTTC	295
TNFAIP6*	NM_007115	EWS	Set 1	GGATTTTCTGGCTGCCTTAGC	ACCACAGAGAAGACCGGTCT	154
COL3A1*	NM_000090	RMS	Set 2	GGAGCTTCTGTTGTGCCAAAT	AGGGAACAACCTTGATGGTGC	199
HLA-DMA*	NM_006120	BL	Set 1	GGAGTTCACCTGTCAGCATGGGTG	GCCAAAACCTTGATGGGAAAA	179
OLFM1*	NM_014279	NB	Set 1	GGAAAACCAGAGTTTCCCGAGGT	GCAATGCCGTAGTTTGGATT	229
GYG2*	NM_003918	EWS	Set 2	GGAAAATGCCACCACTAATGCC	TAGCCTGTGTGAGCTGGATG	159
PFN21*	NM_053024	NB	Set 2	GGAGCAGGGAAAGAAATGGACAA	AAGGGTTCTTTTGAATGGC	164
CRMP1*	NM_001313	NB	Set 1	GGAACCCAGGGTTTCTGACACAGG	GCAGTCCAAAACCCCTTACA	215
ELAVL4***	NM_021952	NB	Set 1	GGATGCAACACTCTGTCTCCAG	GAAGTGGGTGGTGCATCTTT	263
GAP43*	NM_002045	NB	Set 2	GGAGGCTTGTTAGGCTCCTCCT	AGGAGAAAACCTTCCGAGG	230
CHRNA3***	NM_000743	NB	Set 1	GGAGACGCCAAAATGTCCTAGT	CAAAGGGTGCCTTCTAGCTG	300
ANXA1*	NM_000700	EWS	Set 2	GGACCCCTTCTCCTTCTCCTG	TCTTGCTAAGGGTGACCGAT	137
CD99*	NM_002414	EWS	Set 1	GGAGTTCTAACCCCTGCTGCCAAC	GAAAATGCAGAACAGGGGA	174
FPGS***	NM_004957	NB	Set 1	GGAGTCTTCAGGCCATAGCTTCG	AGCATGGAGTACCAGGATGC	288
GATA2*	NM_032638	NB	Set 1	GGAGGAAGAGTCCGCTGCTGTAG	TGACGGAGAGCATGAAGATG	184
MYC*	NM_002467	EWS/BL	Set 2	GGATAACTACCTTGGGGGCCCTT	AGGAGAATGTCAAGAGGCCGA	169
SFRP1*	NM_003012	NB	Set 1	GGATATCGAGGTGGTAAACCCACAG	CGGCAGAGTCTTAGTGGAG	219
MAP1B*	NM_005909	NB	Set 2	GGAAATCTTCAAACCCGCTCCT	GGAAACTGAGCCAGTCAAG	240
BIN1*	NM_139343	RMS	Set 1	GGAGAGGTTCTGGTTGAGCTTGC	CATCAAAGCCCAGAAAGGTGT	200
FCGRT*	NM_004107	EWS	Set 1	GGATATCGAGGTGGAAAACAGCC	ACACTGGCATCTCTGAGCT	159
PRKAR2B*	NM_002736	BL	Set 1	GGAGCAAGCCTCTTGCAACCTAT	ATAAACCGATTACAAGGCG	169
ELF1*	NM_172373	BL	Set 1	GGATAGCCTTGTCTGGAGCAGT	AGCCTCATCACCAGAACAGC	210
CDH2*	NM_001792	NB	Set 2	GGACCATTAAGCCGAGTGAAGT	GAGGCAGAGACTTCCGAAAC	189
IGF2*	NM_000612	RMS	Set 1	GGAAACCCGAGAGATTTTCGGGAT	ATTGCTCTACCCACCCAAGA	205
HLA-DPB1*	NM_002121	BL	Set 1	GGATCACCAGGATCTGGAAGGTC	CACGTGACGGATTCTACCC	164
MYL4*	NM_002476	RMS	Set 2	GGAATCTGTCAAAGACACGCAG	AAGCCCAAGCCTGAAGAGAT	174
COL4A2*	NM_001846	RMS	Set 2	GGACTCTGGCACCTTTTGCTAGG	GATCACCAGGACAGAAGGGA	184
HLA-DQA1*	NM_002122	BL	Set 2	GGAGGTGTCTGGAAGCACAACCT	CTTCTGAAACACTGGGAGCC	194
MYST1***	NM_032188	NB	Set 1	GGAAGTTCCTCCGATGTGGATCTTG	CGCTGACCAAGACAGTGAAG	268
AF1Q*	NM_006818	NB	Set 1	GGAAATCATGAGCACCACCAACA	GGCTCCTTGAATACAGCAC	240
TLE2*	NM_003260	EWS	Set 1	GGACTCATACAGGTGGCCTTCT	GGAAGGACAACCTGCTCAAC	189
PPIA****	BC000689	Control	Set 1 & 2	GGAGCATCAACCACTCAGTCT	GGGTTTATGTGTGACGGTGG	235

Paper source of genes: *, from Khan et al.⁴; **, from Son et al.⁷; ***, from unpublished data; ****, from de Kok et al.⁵. The genes are in ANN rank order.

multiplex RT-PCR assay, we performed a gene minimization procedure using all 96 samples. The rank of the genes and the misclassification rate are shown in Table 2 and Figure 2B. We observed that the top 31 genes resulted in the least classification error, although classification error is very low starting from nine genes (Figure 3A). Testing of 96 samples using these 31 genes in a complete leave-one-out ANN analysis demonstrated that all samples were accurately diagnosed as their respective category (data not shown). Multidimensional scaling analysis using these 31 genes clearly separated the four cancer types (Figure 3B). In addition, hierarchical clustering using the 31 genes showed all except one RMS (RMS_E_17) sample clustering with their respective category (Figure 3C).

Discussion

The SRBCTs represent four of the most aggressive solid cancers in the pediatric population, and accurate diagnosis is critical to the management of these patients. For instance, patients with high-stage RMSs, which are tumors originating from striated muscle, require a combi-

nation of high-dose chemotherapy, surgery, and radiation treatment, whereas patients with non-Hodgkin's lymphoma require repeated lumbar punctures with installation of intrathecal chemotherapy because of the propensity for them to spread to the central nervous system, and they rarely require surgery or radiation therapy. In addition, the majority of patients with high-risk neuroblastoma are currently treated with autologous stem cell transplants, unlike those with RMS, EWS, or non-Hodgkin's lymphoma, where only in rare cases in the setting of recurrent disease do they require stem cell transplants and then usually as an experimental therapy.⁸ Consequently, accurate diagnosis of these cancers is critical for the administration of the correct therapy and for avoiding unnecessary procedures to the patients.

We have therefore developed a rapid and reliable diagnostic assay to distinguish SRBCTs according to their diagnostic categories. Recently, multiple RT-PCR analyses have also been used as a screening tool for detection of genetic rearrangements,⁹ for *in vitro* toxicology screening,¹⁰ and for validation of gene sets obtained from global screens.¹¹ We verified the multiple RT-PCR results using the real-time RT-PCR method

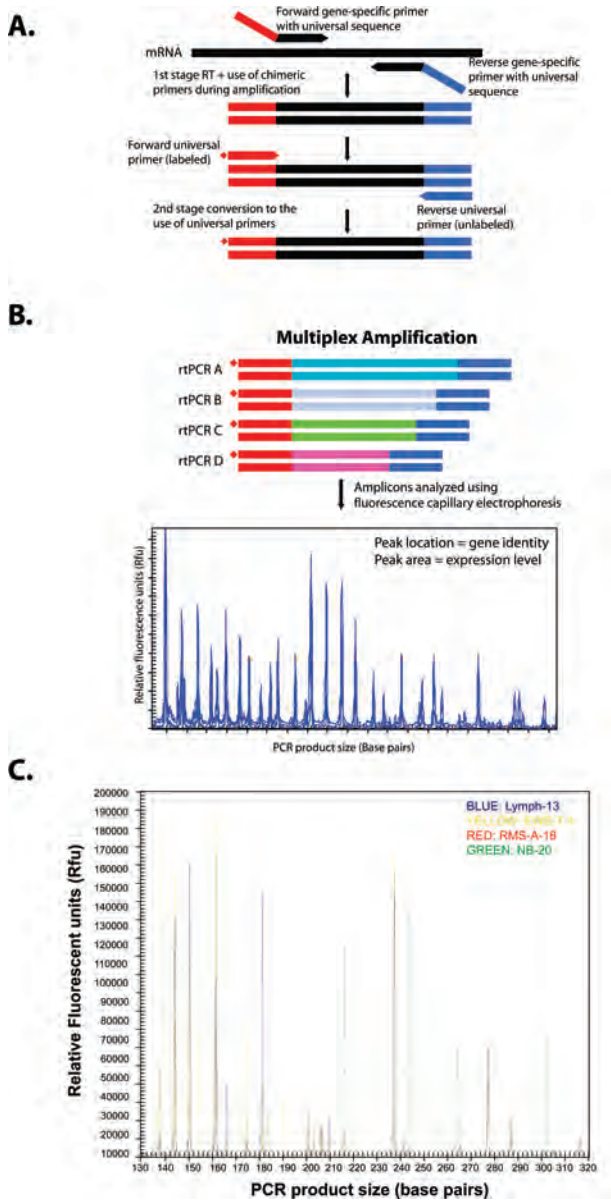


Figure 1. The schematic illustration of multiplex RT-PCR assay. **A:** The multiplex RT-PCR involves two stages: the first stage includes reverse transcription and amplification using chimeric primers, and the second stage converts to the use of a single pair of universal primers during amplification (see Materials and Methods for the primer design). **B:** The amplicons obtained from multiplex amplification were then analyzed using fluorescence capillary electrophoresis. The peak location represents the gene identity, and the peak area represents gene expression level. **C:** The comparative chromatograms of four different categories of tumor samples from one multiplex assay. Blue, Lymph-13; yellow, EWS-T-4; red, RMS-A-18; and green, NB-20.

in a previous study.¹⁰ As discussed above, the development of such a diagnostic assay is of particular importance for rapid clinical confirmation of the diagnosis of these SRBCT cancers because of the difficulty of these cancers to be distinguished by histology and the need for accurate diagnosis to guide therapy. Currently, no single test can precisely differentiate between these cancers. Immunohistochemistry for individual protein markers is used to establish the diagnosis in many instances, but occasionally it will have

diagnostic difficulties because the individual gene may also be expressed in other tumors. For example, *CD99* immunostaining is currently used to diagnose EWS; however, it alone cannot be used to discriminate EWS. Although *CD99* detects EWS with high sensitivity, it is also expressed in several RMS tumors. This issue is resolved using the multiplex RT-PCR method, because one is able to examine the expression levels of multiple genes within a single reaction. Amplifying the signals of many genes concurrently has allowed us to determine the unique gene expression signatures for each of the four tumor types and diminishes the complications of misclassification. Another advantage is that this procedure only requires nanograms of RNA, such as that isolated from needle biopsies. In addition, RT-PCR is a relatively simple procedure, is completed in a short amount of time, and is cost effective.

We have applied an artificial neural network-based method for predicting the diagnostic category of tumor samples using the expression profiles of 39 genes from multiplex RT-PCR assays. We started with two multiplex PCR assays with 40 unique genes (39 differential genes and one control gene) and later combined the data from two assays for classification analysis after normalization. This works well because blinded testing of 96 samples using artificial neural networks in a complete leave-one-out analysis demonstrated that all except one sample were accurately diagnosed as their respective category. This study is essentially equivalent to an independent study because we used genes derived from our previous SRBCT study⁴ and validated by multiplex RT-PCR on an independent set of tumors not used in the previous study. In addition, Figure 3B was done using a leave-one-out analysis, and there was no leakage of information into the testing samples.

When we tested whether we could further minimize the number of genes, we identified the top 31 genes that correctly diagnosed all 96 samples while using the leave-one-out strategy. This demonstrated that 31 genes are sufficient for the diagnostic purpose, and we were therefore able to reduce the number of genes required to diagnose these cancers. Thus, it will be possible to make a single multiplex-PCR assay to diagnose these cancers. Remarkably, the classification error is also very low starting from just 9 genes. However, a redundancy in the number of genes used to diagnose these cancers is important to avoid misdiagnosis in case a single gene fails to amplify.

It is important to note that although we can distinguish the broad categories of the SRBCTs, our method does not remove the necessity for detailed histological or other molecular analysis of these tumors, which gives important clues as to the degree of differentiation or presence of gene amplifications. There are also other similar small round cell tumors including Wilms, hepatoblastoma, desmoplastic small round cell tumor, and others. Our genes were chosen to distinguish only the four major categories including RMS, EWS, NB, and lymphoma, which will therefore identify the majority of the small round cell tumors of childhood. In addition, we recommend that our studies be performed in conjunction with detailed clinical

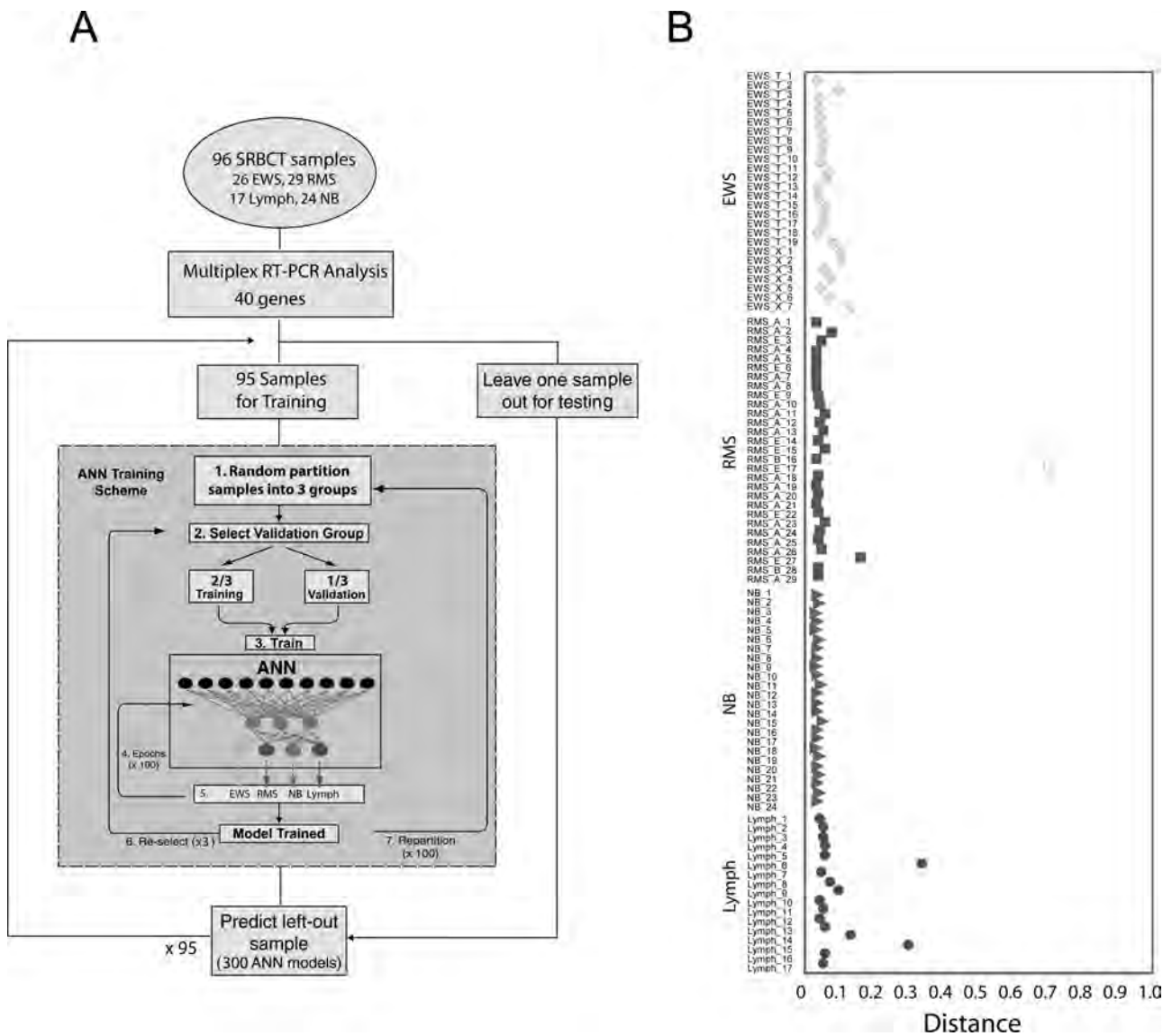


Figure 2. The artificial neural network. **A:** Workflow for a complete leave-one-out ANN analysis. Multiplex RT-PCR analysis using 40 genes was performed on tumors from 96 pediatric cancer patients (26 EWS, 29 RMSs, 17 lymphomas, and 24 NBs). One sample was left out as an independent test sample, and the ANNs were trained using the remaining 95 samples. ANN training scheme (gray box). 1, All samples were randomly partitioned into three groups. 2, One of the three groups (containing 32 samples) was selected as a validation set, whereas the remaining two groups (63 samples) were used to train the network. 3 and 4, The training weights were iteratively adjusted for 100 cycles (epochs). 5, The ANN output (0 to 1) for each of four classes (EWS, RMS, NB, and lymphoma) was calculated for each sample in the validation set. 6, A different validation set was selected from the same partitioning in 1, and the remaining two groups were used for training. Steps 2 through 6 were repeated until each of the three groups from 1 had been used as a validation set exactly one time. 7, The samples were randomly repartitioned into three new groups, and steps 2 through 6 were repeated. Sample partitioning was performed 100 times in total. Thus, steps 1 through 6 were repeated 100 times. Three hundred ANN models were thus trained and were used to predict the left-out test sample. This scheme was repeated for each left-out test sample. This scheme was repeated for each left-out test sample. **B:** Classification of the samples from a leave-one-out ANN analysis. A sample is classified to a cancer category according to its highest committee vote (average of all ANN outputs; Table 1). Plotted is the distance for each sample from its committee vote to the ideal vote for that category (for example, for EWS, it is EWS = 1, RMS = NB = Lymph = 0). The perfectly classified sample would be plotted with a distance of 0. The histological diagnosis of four different cancer categories was displayed in shape as diamond for EWS, square for RMS, triangle for NB, and circle for lymphoma. All samples were correctly classified except one RMS sample, which was misclassified as EWS.

Table 3. Performance of ANN Diagnosis (Leave-One-Out with 39 Genes)

Tumor type	Sensitivity (%)	Specificity (%)	Positive predictive value (%)	Negative predictive value (%)
EWS (<i>n</i> = 26)	100	98.6	96.3	100
RMS (<i>n</i> = 29)	96.6	100	100	98.5
NB (<i>n</i> = 24)	100	100	100	100
Lymphoma (<i>n</i> = 17)	100	100	100	100

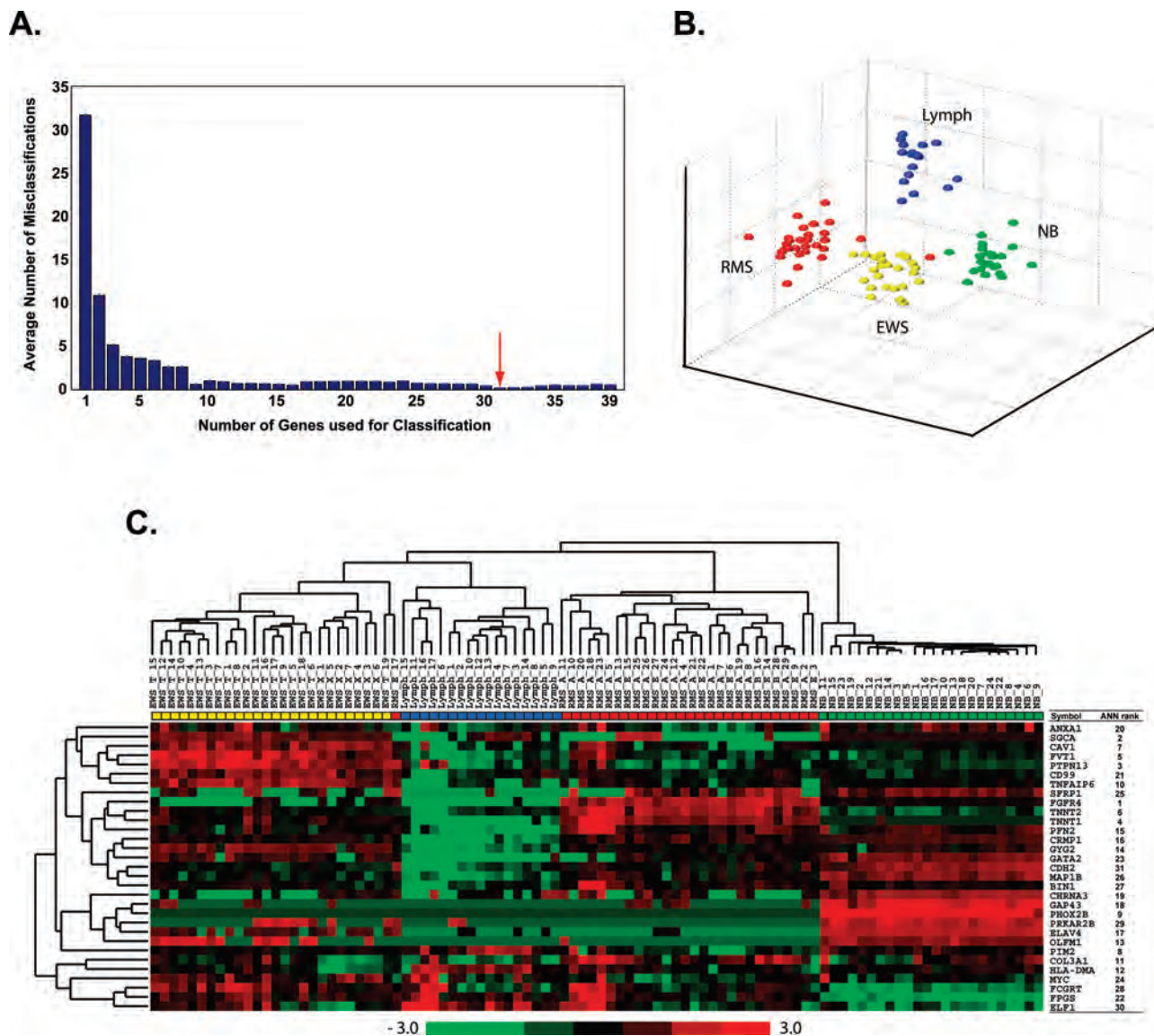


Figure 3. Hierarchical clustering and multidimensional scaling analysis. **A:** Gene minimization plot for ANN prediction. All of 39 genes were used for the analysis of 96 samples. ANNs were first trained using 96 samples, and 39 genes were ranked according to their importance to the ANN prediction. **Red arrow** marked the position of 31 genes. **B:** Multidimensional scaling analysis using 31 top-ranked genes. Three dimensions of the multidimensional scaling plot are shown. EWSs are depicted as yellow circles, RMS as red, NB as green, and lymphoma as blue. The samples clustered closely according to the four different cancer categories. **C:** Hierarchical clustering of all 96 samples and 31 top-ranked genes. Each row represents a gene, and each column, a separate sample. A pseudocolored representation of the ratio (log₂-transformed and z-scored across the samples) is shown. On the right are the gene symbols of 31 genes as well as the ANN gene rank.

investigation, including radiological and serological markers, immunohistopathology, and other molecular markers such as RT-PCR for fusion genes, which will distinguish the RMS, ESW, NB, and lymphoma from other small round cell tumors.

In conclusion, we have developed a simple and reliable diagnostic assay for the major SRBCTs including EWS, RMS, NB, and lymphoma. We believe our assay offers a powerful diagnostic tool for pathologists for a rapid diagnosis using a minimal amount of tissue. However, it will be valuable to include a broader range of other small round cell tumors in the future multiplex RT-PCR assays, and this will be incorporated as more microarray profiling data of these tumors become available.

Acknowledgments

We thank Drs. John Maris, Wendy London (Children's Oncology Group, Philadelphia, PA), and Steven Qualman (Cooperative Human Tissue Network, Columbus, OH) for the tumor samples and patient demographic information.

References

1. Triche TJ, Askin FB: Neuroblastoma and the differential diagnosis of small-, round-, blue-cell tumors. *Hum Pathol* 1983, 14:569-595
2. Taylor C, Patel K, Jones T, Kiely F, De Stavola BL, Sheer D: Diagnosis of Ewing's sarcoma and peripheral neuroectodermal tumour based on the detection of t(11;22) using fluorescence in situ hybridisation. *Br J Cancer* 1993, 67:128-133

3. McManus AP, Gusterson BA, Pinkerton CR, Shipley JM: The molecular pathology of small round-cell tumours: relevance to diagnosis, prognosis, and classification. *J Pathol* 1996, 178:116–121
4. Khan J, Wei JS, Ringner M, Saal LH, Ladanyi M, Westermann F, Berthold F, Schwab M, Antonescu CR, Peterson C, Meltzer PS: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 2001, 7:673–679
5. de Kok JB, Roelofs RW, Giesendorf BA, Pennings JL, Waas ET, Feuth T, Swinkels DW, Span PN: Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes. *Lab Invest* 2005, 85:154–159
6. Wei JS, Greer BT, Westermann F, Steinberg SM, Son CG, Chen QR, Whiteford CC, Bilke S, Krasnoselsky AL, Cenacchi N, Catchpoole D, Berthold F, Schwab M, Khan J: Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. *Cancer Res* 2004, 64:6883–6891
7. Son CG, Bilke S, Davis S, Greer BT, Wei JS, Whiteford CC, Chen QR, Cenacchi N, Khan J: Database of mRNA gene expression profiles of multiple human organs. *Genome Res* 2005, 15:443–450
8. Pizzo PA, Poplack DG: *Principles and Practice of Pediatric Oncology*, ed 5. Philadelphia, Lippincott Williams & Wilkins, 2005
9. Strehl S, Konig M, Mann G, Haas OA: Multiplex reverse transcriptase-polymerase chain reaction screening in childhood acute myeloblastic leukemia. *Blood* 2001, 97:805–808
10. Vansant G, Pezzoli P, Saiz R, Birch A, Duffy C, Ferre F, Monforte J: Gene expression analysis of troglitazone reveals its impact on multiple pathways in cell culture: a case for in vitro platforms combined with gene expression analysis for early (idiosyncratic) toxicity screening. *Int J Toxicol* 2006, 25:85–94
11. Wittig R, Salowsky R, Blaich S, Lyer S, Maa JS, Muller O, Mollenhauer J, Poustka A: Multiplex reverse transcription-polymerase chain reaction combined with on-chip electrophoresis as a rapid screening tool for candidate gene sets. *Electrophoresis* 2005, 26:1687–1691