[CANCER RESEARCH 64, 5539–5545, August 15, 2004]

Advances in Brief

Gene Expression Signatures Identify Rhabdomyosarcoma Subtypes and Detect a Novel t(2;2)(q35;p23) Translocation Fusing PAX3 to NCOA1

Marco Wachtel,[1] Marcel Dettling,[2] Eva Koscielniak,[3] Sabine Stegmaier,[3] Jörn Treuner,[3] Katja Simon-Klingenstein,[3] Peter Bühlmann,[2] Felix K. Niggli,[1] and Beat W. Schäfer[1]

[1]University Children's Hospital, Division of Oncology, Zurich, Switzerland; [2]Seminar for Statistics, Swiss Federal Institute of Technology, Zurich, Switzerland; and [3]Olga Hospital, Department of Pediatric Oncology/Hematology, Stuttgart, Germany

Abstract

Rhabdomyosarcoma is a pediatric tumor type, which is classified based on histological criteria into two major subgroups, namely embryonal rhabdomyosarcoma and alveolar rhabdomyosarcoma. The majority, but not all, alveolar rhabdomyosarcoma carry the specific *PAX3(7)/FKHR*-translocation, whereas there is no consistent genetic abnormality recognized in embryonal rhabdomyosarcoma. To gain additional insight into the genetic characteristics of these subtypes, we used oligonucleotide microarrays to measure the expression profiles of a group of 29 rhabdomyosarcoma biopsy samples (15 embryonal rhabdomyosarcoma, and 10 translocation-positive and 4 translocation-negative alveolar rhabdomyosarcoma). Hierarchical clustering revealed expression signatures clearly discriminating all three of the subgroups. Differentially expressed genes included several tyrosine kinases and G protein-coupled receptors, which might be amenable to pharmacological intervention. In addition, the alveolar rhabdomyosarcoma signature was used to classify an additional alveolar rhabdomyosarcoma case lacking any known PAX3 or PAX7 fusion as belonging to the translocation-positive group, leading to the identification of a novel translocation t(2;2)(q35;p23), which generates a fusion protein composed of PAX3 and the nuclear receptor coactivator NCOA1, having similar transactivation properties as PAX3/FKHR. These experiments demonstrate for the first time that gene expression profiling is capable of identifying novel chromosomal translocations.

Introduction

Rhabdomyosarcoma is the most common soft tissue sarcoma in childhood, representing 5–8% of all malignancies in children (1). On the basis of histological criteria, rhabdomyosarcoma tumors are classified into two major subgroups, namely the more frequent embryonal rhabdomyosarcoma (60%) and the rarer alveolar rhabdomyosarcoma (20%). These subtypes are associated with distinct clinical behavior whereby embryonal rhabdomyosarcoma are associated generally with a more favorable prognosis (2).

Whereas no consistent and unique genetic alterations have been identified in embryonal rhabdomyosarcoma, recent molecular findings allow additional subdivision of the alveolar rhabdomyosarcoma group. A great majority of the alveolar rhabdomyosarcoma tumors are associated with unique chromosomal translocations (fusion-positive alveolar rhabdomyosarcoma). Of these, ~75% are t(2;13)(q35;q14), and ~25% are t(1;13)(p36;q14), leading to the fusion of either the transcription factor PAX3 in the former case or the related transcription factor PAX7 in the latter case to another transcription factor named FKHR (FOXO1A; reviewed in Ref. 3). The resulting fusion products are composed of the NH_2-terminal DNA-binding domain of PAX3 or PAX7 and the COOH-terminal transactivation domain of FKHR. These fusion proteins are more potent transcriptional activators than the wild-type PAX proteins, suggesting that PAX-specific target genes may be deregulated and thereby involved in the oncogenic transformation of the fusion-positive alveolar rhabdomyosarcoma cells.

The remaining 20% of alveolar rhabdomyosarcoma are translocation-negative (fusion-negative alveolar rhabdomyosarcoma) and form a more heterogenous, unexplored group, of which the unambiguous classification and discrimination from embryonal rhabdomyosarcoma based on classical methods such as histology still remains challenging due to the lack of immunohistochemical or ultrastructural markers. Furthermore, this group may also include cases with cryptic PAX3(7) fusions or alternative fusions with other members of the forkhead family of transcription factors (4) that cannot be identified by classical PCR diagnostics. However, clinical outcome of this subtype is as unfavorable as the one from fusion-positive alveolar rhabdomyosarcoma (5) and needs increased treatment intensity compared with embryonal rhabdomyosarcoma. Therefore, novel insight into the underlying genetics of this and the other rhabdomyosarcoma subgroups is clearly needed.

Because gene expression profiling has been shown to reliably classify a range of tumors (6–8), we hypothesized that it should allow us to improve the molecular characterization of rhabdomyosarcoma, thereby facilitating future diagnostic classification. Hence, we performed a microarray study on a cohort of rhabdomyosarcoma biopsy samples ($n = 29$) and show that embryonal rhabdomyosarcoma as well as fusion-positive alveolar rhabdomyosarcoma and fusion-negative alveolar rhabdomyosarcoma have transcriptional signatures that allow identification of these distinct subtypes. Interestingly, the fusion-positive alveolar rhabdomyosarcoma-specific signature led to the identification of a novel variant translocation, which would not have been detected by specific PCR screening.

Materials and Methods

Patient Material

Rhabdomyosarcoma biopsy samples were collected in the course of the CWS studies (German Cooperative Soft Tissue Sarcoma Study; CWS-91 and CWS-96). Fresh tumor tissue samples were obtained at surgery, snap frozen in liquid nitrogen, and stored at −80°C. Morphological assessment and histological classification was done by the CWS study reference laboratories. In all of the samples >80% of tumor cells were present. Clinical and pathological details of the samples are available in the Supplementary Table 1.

Received 3/9/04; revised 6/1/04; accepted 7/7/04.

Grant support: Swiss Cancer League, the Cancer League of the Kanton Zug, and the Stiftung zur Krebsbekämpfung.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article can be found at Cancer Research Online (http://cancerres.aacrjournals.org). DNA sequence reported here is available from genbank, accession no. AY633656. Original array data are available from array express (www.ebi.ac.uk/arrayexpress) under accession no. E-MEXP-121.

Requests for reprints: Beat Schäfer, Division of Oncology, University Children's Hospital, Steinwiesstrasse 75, CH-8032 Zürich, Switzerland. E-mail: beat.schaefer@kispi.unizh.ch.

RNA Purification and Target cRNA Preparation

To prepare total RNA, 50–100 mg of frozen tissue sections were directly homogenized in Trizol (Invitrogen, Groningen, the Netherlands). RNA precipitates were resuspended in diethyl pyrocarbonate-treated water and stored at −80°C. Only RNAs with rRNA ratios > 1.2 as measured by Bioanalyzer (Agilent, Palo Alto, CA) were used.

RNA amplification and labeling were done according to the GeneChip eukaryotic small sample target labeling technical note (Affymetrix, Santa Clara, CA). Briefly, 250 ng of total RNA was used for first and second strand cDNA synthesis with oligo-dT primers containing a T7 RNA polymerase promotor site (Microsynth, Balgach, Switzerland) using the SuperScript cDNA synthesis customer kit (Invitrogen). cRNA was prepared by *in vitro* transcription using the BioArray High Yield RNA Transcript Labeling kit (Enzo, Farmingdale, NY). Of the resulting cRNA, 250 ng was used for a second round of first-strand cDNA synthesis with random hexamer primers (Promega), followed by second strand cDNA synthesis with the T7-oligo-dT primers. Labeled target cRNA was then prepared in a second round of *in vitro* transcription using biotin-labeled CTP and UTP (Enzo).

Array Hybridization and Scanning

Fifteen μg of labeled cRNA was fragmented at 94°C for 35 min in a buffer containing 40 mM Tris-acetate (pH 8.1), 100 mM potassium acetate, and 30 mM magnesium acetate. Fragmented cRNA was mixed with a mixture containing 0.1 M 4-morpholinepropanesulfonic acid (pH 6.5–6.7), 1 M NaCl, 20 mM EDTA, 0.01% Tween-20, 0.5 mg/ml acetylated bovine serum albumin, 0.1 mg/ml herring sperm DNA, 1× eukaryotic hybridization control mix, and 50 pM B2 control oligo (Affymetrix). This mixture was heated to 95°C for 5 min, followed by incubation at 45°C for 5 min and centrifugation for 5 min at maximum speed. The probe array was incubated at 45°C for 16 h at constant rotation (60 rpm). Subsequent washings and stainings were done in an Affymetrix GeneChip Fluidics Station using the protocol EukGen W2V4. Stained arrays were scanned in a Gene Array Scanner (Agilent), and expression values for each gene were calculated by Affymetrix GeneChip software.

Statistical Analysis

Clustering. Unsupervised two-way hierarchical clustering was done with the software D-chip (9). Expression values were base 2 logarithmized and normalized. Before clustering, genes were filtered using a variation coefficient (SD/mean) of at least 0.058 yielding 4168 nonredundant genes, which were used for an average-linkage hierarchical clustering based on Pearson correlation coefficients. Other variation coefficients gave similar results.

Gene Selection. Supervised testing for genes with differential expression was done for all three of the groups simultaneously by computing between-group to within-group sums of squares, based on logarithmized and normalized data. Raw *P* values were computed from an F-distribution with 2 and 26 degrees of freedom, and they were then adjusted by the Benjamini-Hochberg false discovery rate, according to the implementation in the statistical software bundle R (package, multitest). For dimension reduction and visualization of this gene set, a principal component analysis was conducted. In binary problems, gene selection was based on Student's *t* test statistic. Raw *P* values were obtained from t-distributions with corresponding degrees of freedom before the Benjamini-Hochberg false discovery rate adjustment was performed.

Classification. In the absence of genuine test sets, the predictive potential for the classification of all three of the groups simultaneously, as well as for the binary translocation status, was evaluated by leave-one-out cross-validation. In each cross-validation run, we identified the 10, 25, 50, 100, 200, and 500 most predictive genes according to their t- or F-statistic, and used them as input for three established methods in microarray classification: a support vector machine with radial basis kernel, the 1-nearest-neighbor rule, and diagonal linear discriminant analysis. We observed the best prediction results, however, by applying an alternative method called Pelora (10), a classifier that, in a forward search, constructs features (so-called supervised clusters, which are metagenes that summarize the predictively most informative individual genes) and uses them for prediction with a penalized form of logistic regression.

3′-Race and Sequencing

3′-Race was performed with the GeneRacer kit (Invitrogen). Briefly, first-strand cDNA was reverse transcribed from 1 μg of total RNA using SuperScript II RNase H^- reverse transcription (Invitrogen) and the GeneRacer Oligo-dT Primer and then amplified using a PAX3-specific forward primer (5′-AACGGCAGGCCGCTGCCCAACC-3′) and the GeneRacer 3′ Primer, followed by a nested PCR reaction using a second PAX3-specific forward primer (5′-ATCGTGGAGATGGCCCACCACG-3′) and the GeneRacer 3′ nested reverse primer. In the first reaction an initial denaturation at 94°C for 5 min was followed by 5 cycles at 94°C for 30 s and at 72°C for 6 min, by 5 cycles at 94°C for 30 s, at 71°C for 30 s, and at 72°C for 6 min and by 25 cycles at 94°C for 30 s, at 70°C for 30 s, and at 72°C for 6 min, followed by a final elongation at 72°C for 10 min. In the second reaction the same scheme was applied, but annealing temperatures were 70°C, 68°C, and 66°C. PCR products were cloned into the pCR-XL-TOPO cloning vector (Invitrogen) and sequenced (Dye Terminator Sequencing kit; Applied Biosystems, Applera Europe B.V., Rotkreuz, Switzerland).

Transactivation Assay

The full-length cDNA of PAX3/NCOA1 was amplified with the expand long template PCR system (Roche Applied Science, Penzberg, Germany). The cDNA was cloned into the pcDNA3 expression vector. Similar constructs were generated for PAX3, PAX3/FKHR, and PAX3 without the transactivation domain. Expression constructs were cotransfected together with a reporter plasmid containing the luciferase gene downstream of multimerized PAX3 DNA-binding sites (6xCD19 DNA-binding sites; Ref. 11) and a plasmid containing the lacZ gene into 293T cells. Forty-eight h after transfection, cells were lysed in reporter lysis buffer (Promega, Madison, WI). β-Galactosidase and luciferase activities were determined with the corresponding assay systems (Promega). Luciferase activity values were normalized with the β-galactosidase activity values. Experiments were performed at least twice, each in triplicate.

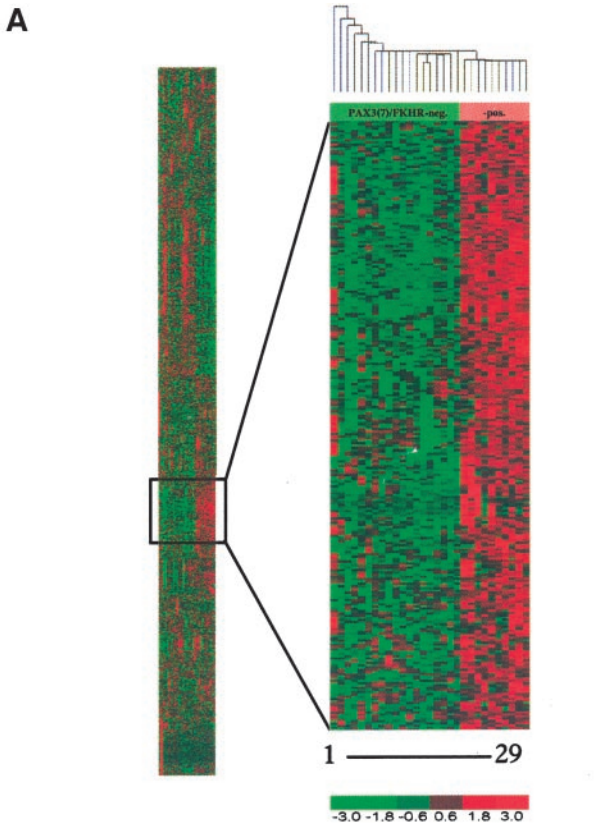
Subcellular Localization of PAX3 Proteins

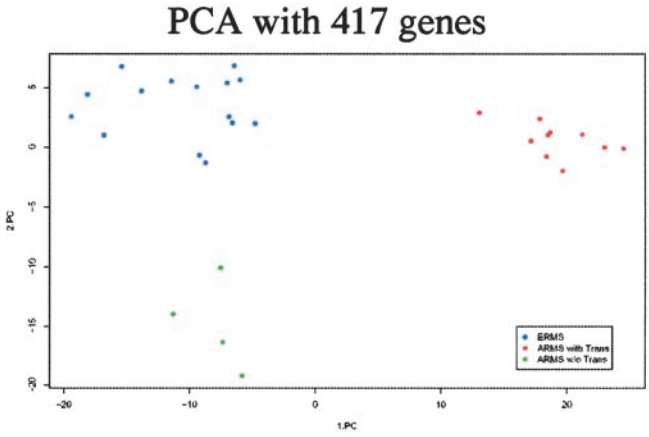
Constructs were transfected into 293T cells as described above and nuclear and cytosolic extracts prepared 48 h after transfection (12). Extracts were analyzed on Western blots using an anti-PAX3/7 antibody (clone N-19; Santa Cruz Biotechnology, Santa Cruz, CA) and visualized with the enhanced chemiluminescence system (Amersham Biosciences, Piscataway, NJ).

Results

Identification of Expression Signatures for Rhabdomyosarcoma Subgroups. Initially, we monitored expression profiles of 29 rhabdomyosarcoma samples with Affymetrix oligonucleotide microarrays (HG-U133A; see Supplemental Data for the complete data set). The expression data were validated by comparing microarray expression levels of four selected genes (*TFAP2B*, *OLIG2*, *BAIAP1*, and *DUSP6*) in eight tumor samples with results obtained by quantitative real-time PCR analysis. Results of the two methods were in good agreement (see Fig. 5 in the Supplemental Data).

To get an overview of the variability and similarity within the transcriptomes of these 29 rhabdomyosarcoma samples, we first performed an unsupervised two-way hierarchical clustering. Genes that did not show variability in expression among the rhabdomyosarcoma samples were excluded from the clustering (for filtering criteria see "Materials and Methods"). 4168 nonredundant probe sets passed the applied filtering procedure and were used for the analysis. In the resulting hierarchical clustering, the 10 translocation-positive rhabdomyosarcoma samples are clearly separated from the 19 translocation-negative ones (Fig. 1*A*). We observed a part of several hundred genes, which are consistently up-regulated in the translocation-positive samples when compared to the translocation-negative ones (see magnification of a stretch with 271 consistently differentially





B

Classifier	Leave-one-out cross validation error rates [%]				
	q=2	q=4	q=6	q=8	q=10
Pelora	0	0	0	0	0
	p=25	p=50	p=100	p=200	p=500
SVM	0	0	0	0	0
DLDA	0	0	0	0	0
NNR	0	0	0	0	0

D

Classifier	Leave-one-out cross validation error rates [%]				
	q=2	q=4	q=6	q=8	q=10
Pelora	10.3	10.3	3.4	3.4	3.4
	p=25	p=50	p=100	p=200	p=500
SVM	10.3	13.8	13.8	13.8	13.8
DLDA	6.9	6.9	6.9	10.3	10.3
NNR	13.8	10.3	13.8	13.8	13.8

Fig. 1. Identification of rhabdomyosarcoma subgroups by gene expression profiling. *A*, unsupervised two-way hierarchical clustering of 4′168 filtered genes across 29 rhabdomyosarcoma biopsy samples. A part of 271 genes, which discriminate significantly between PAX3/FKHR-positive and -negative samples, is framed with a box and magnified on the right side. Scale bar below the figure indicates fold changes in gene expression. The colorgram depicts high (*red*) and low (*green*) relative levels of gene expression. *B*, leave-one-out cross validation error rates for discrimination of translocation-positive and -negative rhabdomyosarcoma samples with various classifiers (*Pelora*, supervised clustering approach based on penalized logistic regression; *SVM*, support vector machine; *DLDA*, diagonal linear discriminant analysis; *NNR*, 1-nearest-neighbor-rule), different number of gene clusters q, which were identified by Pelora and of which the mean expression was used for classification or gene subsets of size p. *C*, principal component analysis (*PCA*) of the rhabdomyosarcoma samples using 417 genes that significantly discriminate among embryonal rhabdomyosarcoma, translocation-positive ARMS and translocation-negative ARMS. The samples are colored according to their subgroup classification as indicated in the inset. (D) Leave-one-out cross validation error rates for simultaneous discrimination of fusion-positive alveolar rhabdomyosarcoma, fusion-negative alveolar rhabdomyosarcoma, and embryonal rhabdomyosarcoma with various classifiers (*Pelora*, supervised clustering approach based on penalized logistic regression; *SVM*, support vector machine; *DLDA*, diagonal linear discriminant analysis; *NNR*, 1-nearest-neighbor rule), different number of gene clusters q, which were identified by Pelora and of which the mean expression was used for classification or gene subsets of size p. An error rate of 3.4% corresponds to one misclassification.

expressed genes) and which produce a very dominant signal. Indeed, a leave-one-out cross validation study for estimating the misclassification risk in out-of-sample classification yielded the benchmark of 0% error-rate with a wide range of genes applying four different statistical methods including a newly developed method based on gene groups called Pelora (Ref. 10; Fig. 1*B*). This demonstrates that fusion-positive alveolar rhabdomyosarcoma are characterized by an expression signature, which separates this group with very high accuracy from other rhabdomyosarcoma. Within the translocation-negative group, the fusion-negative alveolar rhabdomyosarcoma group together with the embryonal rhabdomyosarcoma.

To determine whether gene expression signatures can characterize all three of the rhabdomyosarcoma subgroups we identified differentially expressed genes by computing between-group to within-group sums of squares. After calculating the raw *P* values from an F-distribution and correcting for the multiplicity problem by the Benjamini-Hochberg false discovery rate, we obtained a set of 417 genes with significant differential expression at the 1% level. A projection of all 29 of the biopsy samples into the space of the first two principal components of these 417 genes yielded a clear separation of all of the three rhabdomyosarcoma subtypes (Fig. 1*C*). The first principal component discriminates between translocation-positive and -negative samples, and the second principal component between fusion-negative alveolar rhabdomyosarcoma and all of the other samples. Performing supervised classification in a leave-one-out cross validation study resulted in a maximal accuracy of 96.5%, *i.e.*, only one single fusion-negative alveolar rhabdomyosarcoma patient was erroneously predicted as an embryonal rhabdomyosarcoma (Fig. 1*D*). These experiments demonstrate that accurate classification of the rhabdomyosarcoma subtypes is possible with several different classifiers and gene sets of variable size.

To define the gene signatures that characterize each subgroup in more detail, genes that discriminate most significantly were selected by pair-wise comparison of different tumor subgroups with different *P* values. The number of genes that discriminate most significantly are summarized in Fig. 2*A*, and a selection of 30 genes is illustrated in

A

Test	Number of discriminating genes		
	p<0.001	*p*<0.01	*p*<0.05
Anova, all classes simultaneously	121	**417**	1246
t-test, pARMS/rest	**299**	943	2184
t-test, nARMS/ERMS	3	**28**	191
t-test, ARMS/ERMS	6	**117**	755

B

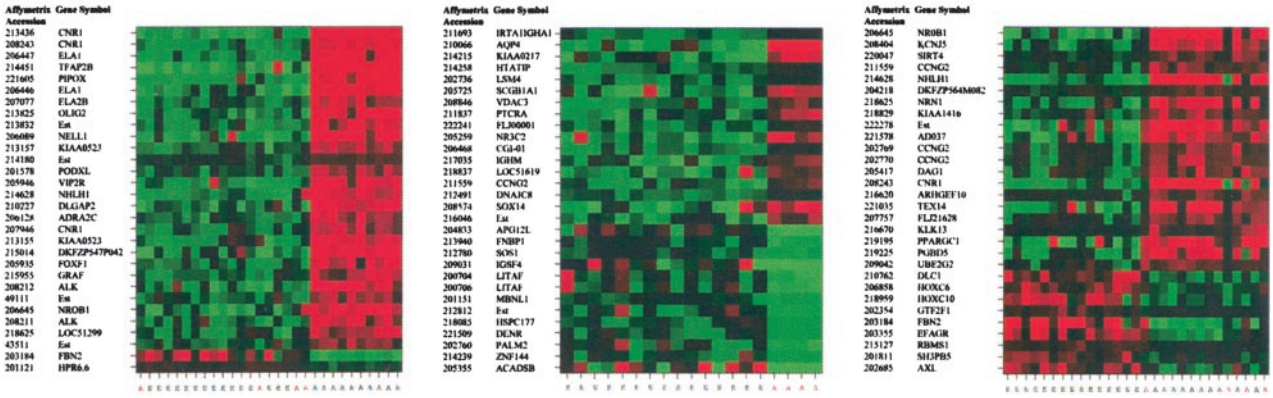


Fig. 2. Identification of subgroup-specific gene signatures. *A,* number of discriminating genes calculated for all three rhabdomyosarcoma (*RMS*) subgroups simultaneously (Anova) or by pair-wise comparison of RMS subgroups (*t* test) for different *P* values. Complete lists of the genes indicated in *bold* can be found in Supplemental Tables 2–5. *B,* subtype-specific signatures: 30 genes which discriminate between translocation-positive and -negative RMS (left panel), between embryonal RMS and translocation-negative RMS (middle panel), and between embryonal RMS and alveolar RMS (right panel) are shown. *A,* alveolar RMS; *E,* embryonal RMS. Labels for translocation-negative ARMS are in *red*. Genes are identified by their Affymetrix accession number and symbol. Colors depict high (*red*) and low (*green*) relative levels of gene expression.

Fig. 2*B*. The strongest signature was found for genes discriminating translocation-positive from translocation-negative samples with 299 genes at $P < 0.001$. The weakest signature discriminates fusion-negative alveolar rhabdomyosarcoma from embryonal rhabdomyosarcoma samples with only 28 genes at $P < 0.01$. The complete gene lists for these signatures can be found in Supplemental Tables 2–5. Taken together, these findings extend the results from the hierarchical clustering and demonstrate that rhabdomyosarcoma subgroups can be characterized by distinct gene expression signatures.

Identification of a Novel Translocation in Alveolar Rhabdomyosarcoma. During the course of these investigations, another sample of the fusion-negative alveolar rhabdomyosarcoma group (A20) was obtained and its expression profile measured as described before. Hierarchical clustering revealed that A20 clearly segregated within the branch of fusion-positive alveolar rhabdomyosarcoma tumors (Fig. 2*A*). Hence, we reinvestigated the three known translocation products PAX3/FKHR, PAX3/AFX (4), or PAX7/FKHR by specific PCR; however, no expressed transcript was detected. Therefore, we calculated the probability for the presence of a translocation involving PAX3(7) based on 100 genes of the translocation-specific signature. Whereas this probability is 0 for all of the translocation-negative samples and 1 for samples expressing a translocation, we found a similar probability of 1 for sample A20 (Fig. 2*B*). Hence, we hypothesized that an alternative translocation involving the DNA-binding domains of PAX3 or PAX7 must be responsible for the generation of the translocation-specific expression signature in AZO.

To test this hypothesis, we performed a series of 3′- rapid amplification of cDNA ends experiments. Using PAX3- and PAX7-specific forward primers and universal reverse primers in two rounds of nested PCR, we obtained a PCR product of ~3 kbp when using PAX3-specific forward primers. Sequence analysis of this product showed that there was an in-frame fusion transcript present consisting of PAX3 and the nuclear receptor coactivator NCOA1. The translocation break point of the fusion is located in intron 6 of *PAX3* and in intron 12 of *NCAO1* (splice variant 2) leading to a fusion of the first 6 exons of *PAX3* to the last 9 exons of the *NCOA1* (splice variant 2; Fig. 3*C*). Hence, the fusion protein lacks exon 7 of *PAX3*, which is always present in *PAX3(7)/FKHR* translocation products. The resulting fusion protein is 913 AA long and is composed of the 338 NH_2-terminal AA of PAX3 and the 557 COOH-terminal AA of NCOA1 (AA 843-1399) plus a unique AA at the break point. Therefore, as predicted originally, the chimeric protein includes both DNA-binding domains of PAX3, paired-box and homeodomain, and a long COOH-terminal part of NCOA1 presumably acting as transactivation domain (Fig. 3*D*).

PAX3/NCOA1 Acts as Transcriptional Activator. The domain structure of PAX3/NCOA1 suggests that it too can act as a chimeric transcription factor. To test this notion, we first confirmed the expression of the full-length PAX3/NCOA1 message by specific PCR using primers from the 5′ end of PAX3 and the 3′ end of NCOA1 (Fig. 4*A*). In contrast, no message could be detected that would code for the reverse NCOA1/PAX3 product suggesting that, indeed, PAX3/NCOA1 is the relevant molecule resulting from the chromosomal translocation. Next, we verified that the cloned PAX3/NCOA1 message can give rise to the expected protein and enter the nucleus. Toward this end, wild-type PAX3, PAX3 without the transactivation domain (PAX3-TA), PAX3/FKHR, and the novel PAX3/NCOA1 were cloned into an expression vector and transfected into human kidney carcinoma cells (293T). Nuclear and cytosolic extracts were used in a Western blot analysis with a PAX3-specific antibody and revealed bands of the expected size for all of the proteins in the nuclear extracts only (Fig. 4*B*). This analysis suggests that PAX3/NCOA1 can be synthesized *in vivo* and enters the nucleus with a similar efficiency than PAX3/FKHR. In addition, all of the proteins were expressed at similar levels.

Finally, to compare the transactivation potency of PAX3/NCOA1 with wild-type PAX3 and PAX3/FKHR, all of the constructs were cotransfected into 293T cells with a luciferase reporter gene downstream of multimerized PAX3 DNA binding-sites (6xCD19 reporter plasmid; Ref. 11). For both PAX3/FKHR and PAX3/NCOA1 we observed an ~10-fold increase in transactivation over PAX3 (Fig. 4*C*) and ~18.5 fold over PAX3-TA. Because similar transactivation efficiencies have been obtained for PAX3/

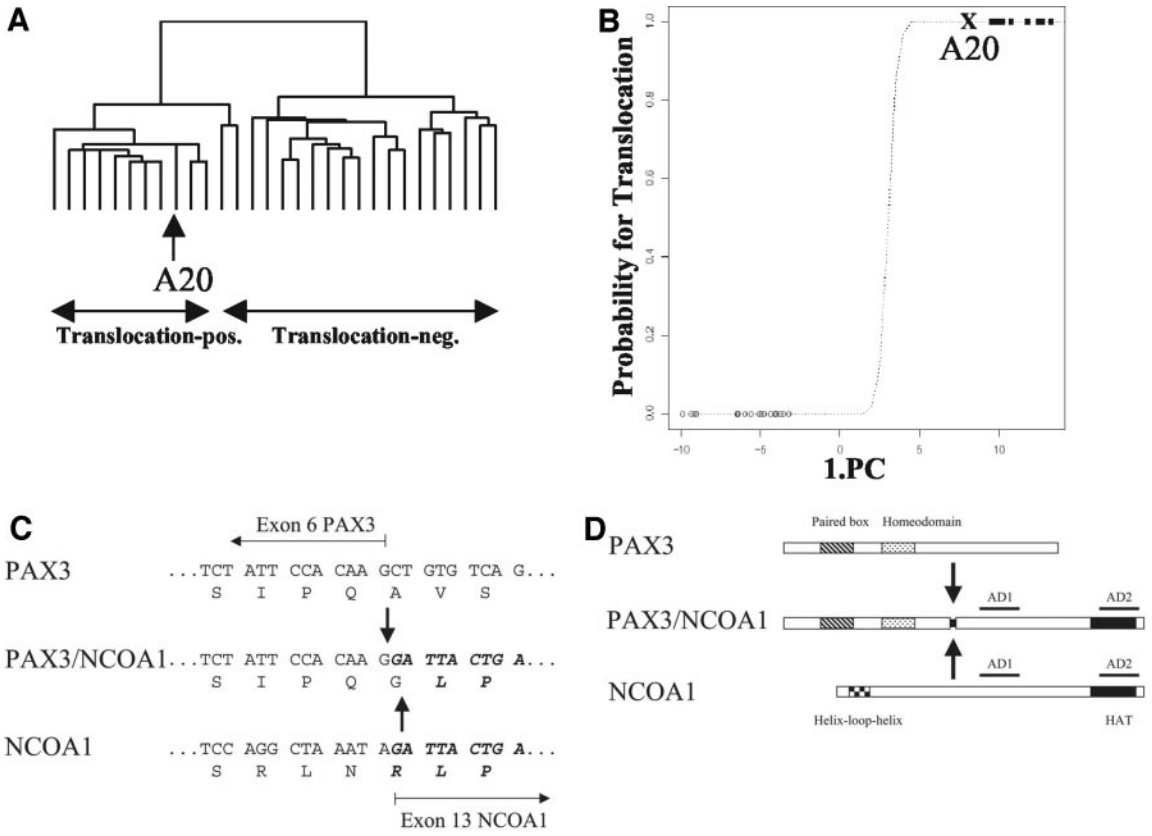


Fig. 3. Characterization of a novel chromosomal translocation. *A*, dendrogram of an unsupervised two-way hierarchical clustering including the PAX3(7)/FKHR-negative alveolar rhabdomyosarcoma sample A20 using 4′654 filtered genes. A20 is indicated by an *arrow*. *B*, probability estimate for the presence of the PAX3(7)/FKHR translocation in rhabdomyosarcoma biopsy samples. Calculation is based on 100 genes from the translocation associated signature. Individual samples are indicated; sample A20 is marked by an x. *C*, sequence alignment of the PAX3/NCOA1 breakpoint region. *Arrows* depict the fusion point. Single letter amino acid code is depicted *below* the DNA sequences. *D*, domain structure of the novel PAX3/NCOA1 in comparison to PAX3 and NCOA1. *Arrows* indicate the fusion point. *AD1, AD2*, transcriptional activation domain; *HAT*, histone acetyltransferase domain.

FKHR using the same reporter sequences in a previous study (13), we conclude that PAX3/NCOA1 can act as transcription factor on PAX3-regulated genes and thereby is involved, directly or indirectly, in the generation of the translocation-specific expression profile.

Discussion

Gene expression profiling has been shown in different studies to be capable of (sub)classifying a range of tumors like leukemia (8) and malignant melanomas (14), as well as different types of solid tumors (7, 15). Hence, to find subtype-specific expression patterns and to better understand the genetic basis of pediatric rhabdomyosarcoma, we performed gene expression profiling of 29 rhabdomyosarcoma biopsy samples with Affymetrix oligonucleotide microarrays. Due to histological criteria, 15 of the tumors in the study were classified as embryonal rhabdomyosarcoma and 14 as alveolar rhabdomyosarcoma; 10 of the latter were positive for the PAX3(7)/FKHR-translocation as detected by classical PCR amplification of the translocation product and 4 negative.

In our analysis, strongest and most consistent differences in the transcriptional pattern were found to discriminate between translocation-positive and -negative tumors. This signature is made up of several hundred genes, which are consistently up-regulated in the translocation-positive samples. This robust gene signature suggests that at least some of the genes in this pattern may be direct or indirect targets of PAX3(7)/FKHR. Moreover, the pattern is present in all of the translocation-positive samples irrespective of the PAX gene involved in the translocation, suggesting a similar influence of the two translocation products on the transcriptome. This is in accordance with the finding that PAX3 and PAX7 bind to the same DNA sequences (11) and, therefore, might have an overlapping spectra of target genes.

In an attempt to additionally define these target genes, transient or stable ectopic expression of PAX3/FKHR was carried out in a range of cell lines (HeLa, Ruch3, 293T, and Rh1; data not shown) but failed to induce the expression of these genes. These results suggest that PAX3/FKHR expression alone is not sufficient to induce the gene expression pattern identified in fusion-positive alveolar rhabdomyosarcoma biopsy samples. Cellular context such as the differentiation state or cell culture effects may be important parameters influencing the activation of target genes by PAX3/FKHR.

Up to now, few genes have been suggested to constitute targets for the PAX3/FKHR translocation product, among them NCAM (16), PDGFR-α (17), c-met (18), bcl-xl (19), as well as MYCN and arginino succinate synthetase (20). It is interesting to note that from these NCAM, MYCN, and argininosuccinate synthetase were also part of our expression signature. In contrast, no consistently higher expression levels in translocation-positive samples were found for PDGFR-α, c-met, and bcl-xl.

Our signature contains several genes thus far not associated with the development of rhabdomyosarcoma. Among them are numerous transcriptional regulators with either enhanced (AP2beta and enhancer-of-split) or reduced (HoxC6, HoxC10, ZFH4, and ZFP36) expression in the translocation-positive samples. Translocation-positive samples also displayed prominent expression of two proteases (ADAM10 and elastase), which could play a role in

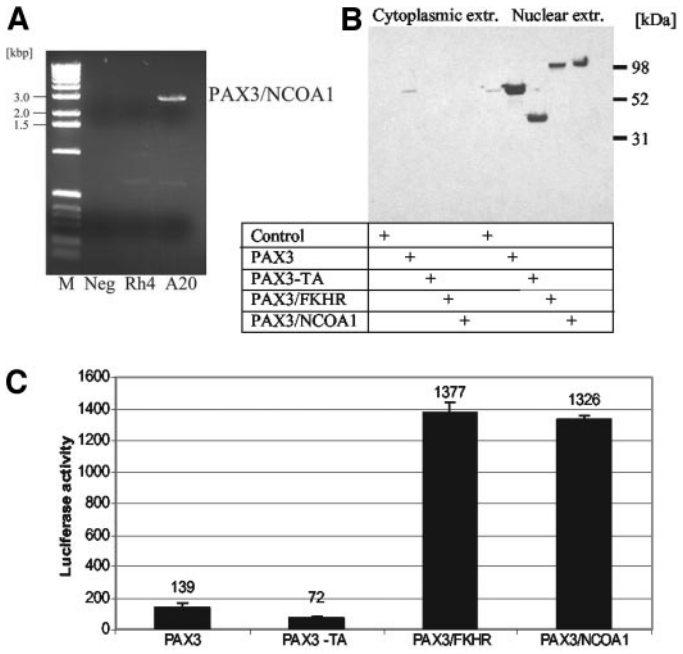


Fig. 4. PAX3/NCOA1 is a transcriptional activator. *A*, agarose gel electrophoresis of a PAX3/NCOA1-specific PCR using primers from the 5′ and 3′ ends of PAX3 and NCOA1, respectively. *M*, marker; *Neg*, control without input RNA. *B*, Western blot analysis of extracts from 293T cells transfected with the cDNA constructs indicated and stained with a PAX3-specific antibody. Control means transfected with empty vector. *C*, luciferase activity measured from extracts of 293T cells cotransfected with the indicated PAX expression constructs and a reporter plasmid containing the luciferase gene downstream of a PAX3 DNA-binding site (6 × CD-19). All values have been corrected for transfection efficiencies through measurement of β-galactosidase activity from a cotransfected β-galactosidase plasmid. Transfection experiments were done at least twice in triplicate, the result of one representative experiment is shown; *bars*, ±SD.

the enhanced metastatic potential of these cells. However, from a therapeutic point of view undoubtedly the most interesting genes with differential expression were found in the group of receptor proteins. In translocation-positive samples we found up-regulation by a factor 10–100-fold of cannabinoid receptor 1, FGFR2, and FGFR4 (see also Ref. 7), α-2 adrenergic receptor, as well as the acetylcholine receptor β3. In contrast, high expression in translocation-negative samples was found for the epidermal growth factor receptor. None of these genes are expressed in normal muscle at significant levels (data not shown). For at least two of these receptors (cannabinoid receptor 1 and epidermal growth factor receptor) experimental drugs are available suggesting novel treatment strategies for rhabdomyosarcoma. This is especially remarkable because our data derive from tumor biopsies, unlike the few other expression profiling studies for rhabdomyosarcoma that are available in the literature and that are mostly restricted to rhabdomyosarcoma cell lines. In the only other study involving tumor biopsies, rhabdomyosarcoma samples were profiled against different small, round blue cell tumors (7), and no attempt was made to discriminate rhabdomyosarcoma subtypes.

The significance of the identified translocation-positive signature was additionally underscored with the identification of a novel PAX3 partner gene, namely *NCOA1*. *NCOA1* is a member of the p160 steroid receptor coactivator gene family, which is composed of the three members *NCOA1* (*SRC-1*), *NCOA2* (*SRC-2*), and *SRC-3*. These coactivators interact with ligand-bound nuclear receptors to recruit histone acetyltransferases and methyltransferases to specific enhancer/promotor regions, which facilitates chromatin remodelling, assembly of general transcription factors, and transcription of target genes. In the COOH-terminal part, present in the fusion protein, two transcriptional activation domains, TAD1 and TAD2, are located. They are responsible for interaction with general transcriptional cointegrators such as p300 and CBP (TAD1), as well as with histone methyltransferases, coactivator-associated arginine methyltransferase 1 and PMRT1 (TAD2). In our assays, PAX3/NCOA1 proved to be a transcriptional activator comparable with PAX3/FKHR additionally supporting the notion that deregulated PAX3 target genes are important for tumor development.

Because the relevant expression signature was not identified in the remaining fusion-negative alveolar rhabdomyosarcoma samples, we conclude that these samples are indeed free of cryptic translocations involving PAX3(7) as found in some alveolar rhabdomyosarcoma initially classified as fusion-negative alveolar rhabdomyosarcoma (4) and that the fusion-negative alveolar rhabdomyosarcoma might be biologically more related to embryonal rhabdomyosarcoma than to fusion-positive alveolar rhabdomyosarcoma. However, our finding of a novel PAX3 translocation in an alveolar rhabdomyosarcoma sample suggests that other PAX3(7) translocations may exist in alveolar rhabdomyosarcoma. Therefore, expression profiling might be superior to specific PCR diagnostics because it also allows the identification and recognition of variant translocations. The identification of subtype-specific gene expression signatures from *in vivo* rhabdomyosarcoma biopsies may allow a faster and more precise molecular classification of rhabdomyosarcoma samples in the future. Certainly, it will assist in classifying cases of the translocation-negative alveolar subtype where histopathological criteria are ambiguous. Moreover, the genes that form the expression signatures displaying differential expression do also hint at novel pathways involved in rhabdomyosarcoma development and, hence, suggest new treatment strategies that could help to improve current treatment regimens.

Acknowledgments

The gene expression profiling was performed at the Functional Genomics Center of the University and ETH Zurich (FGCZ). We especially thank Andrea Patrignani from the FGCZ for his technical support and Claus Heizmann for his support in the beginning of the experiments.

References

1. Pappo AS. Rhabdomyosarcoma and other soft tissue sarcomas of childhood. Curr Opin Oncol 1995;7:361–6.
2. Koscielniak E, Harms D, Henze G, et al. Results of treatment for soft tissue sarcoma in childhood and adolescence: a final report of the German Cooperative Soft Tissue Sarcoma Study CWS-86. J Clin Oncol 1999;17:3706–19.
3. Barr FG. Gene fusions involving PAX and FOX family members in alveolar rhabdomyosarcoma. Oncogene 2001;20:5736–46.
4. Barr FG, Qualman SJ, Macris M, et al. Genetic heterogeneity in the alveolar rhabdomyosarcoma subset without typical gene fusions. Cancer Res 2002;62:4704–10.
5. Sorensen PH, Lynch JC, Qualman SJ, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. J Clin Oncol 2002;20:2672–9.
6. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7.
7. Khan J, Wie JS, Ringner M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat Med 2001;7:673–9.
8. Yeoh E-J, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1:133–43.
9. Li C, Wong WH. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. Proc Natl Acad Sci USA 2001;98:31–6.
10. Dettling M, Buhlmann P. Finding predictive gene groups from microarray data. J Multivariate Anal 2004;90:106–31.
11. Schafer BW, Czerny T, Bernasconi M, Genini M, Busslinger M. Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes. Nucleic Acids Res 1994;22:4574–82.

12. Schreiber E, Matthias P, Muller MM, Schaffner W. Rapid detection of octamer binding proteins with 'mini-extracts', prepared from a small number of cells. Nucleic Acids Res 1989;17:6419.
13. Bennicelli JL, Edwards RH, Barr FG. Mechanism for transcriptional gain of function resulting from chromosomal translocation in alveolar rhabdomyosarcoma. Proc Natl Acad Sci USA 1996;93:5455–9.
14. Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000;406:536–40.
15. Dyrskjot L, Thykjaer T, Kruhoffer M, et al. Identifying distinct classes of bladder carcinoma using microarrays. Nat Gen 2003;33:90–6.
16. Edelman GM, Jones FS. Developmental control of N-CAM expression by Hox and Pax gene products. Philos Trans R Soc Lond B Biol Sci 1995;349:305–12.
17. Epstein JA, Song B, Lakkis M, Wang C. Tumor-specific PAX3-FKHR transcription factor, but not PAX3, activates the platelet-derived growth factor alpha receptor. Mol Cell Biol 1998;18:4118–30.
18. Relaix F, Polimeni M, Rocancourt D, Ponzetto C, Schafer BW, Buckingham M. The transcriptional activator PAX3-FKHR rescues the defects of Pax3 mutant mice but induces a myogenic gain-of-function phenotype with ligand-independent activation of Met signaling in vivo. Genes Dev 2003;17:2950–65.
19. Margue CM, Bernasconi M, Barr FG, Schafer BW. Transcriptional modulation of the anti-apoptotic protein BCL-XL by the paired box transcription factors PAX3 and PAX3/FKHR. Oncogene 2000;19:2921–9.
20. Khan J, Simon R, Bittner M, et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. Cancer Res 1998;58:5009–13.