Proteomic identification of oncogenic chromosomal translocation partners encoding chimeric anaplastic lymphoma kinase fusion proteins

Kojo S. J. Elenitoba-Johnson*†‡, David K. Crockett†, Jonathan A. Schumacher†, Stephen D. Jenson*, Cheryl M. Coffin*, Alan L. Rockwood†, and Megan S. Lim*†‡

*Department of Pathology, University of Utah School of Medicine, Salt Lake City, UT 84132; and †ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, UT 84108

Edited by John R. Yates, The Scripps Research Institute, La Jolla, CA, and accepted by the Editorial Board March 21, 2006 (received for review July 29, 2005)

The anaplastic lymphoma kinase (ALK) on 2p23 is a tyrosine kinase that forms chimeric fusions with numerous translocation partners. We describe a mass spectrometry-based approach for the identification of ALK fusion partners. This approach accurately identified the nucleophosmin (NPM)–ALK fusion protein in an anaplastic large cell lymphoma (ALCL)-derived cell line carrying the t(2;5)(p23;q35), and the TPM3–ALK in a clinical biopsy of inflammatory myofibroblastic tumor (IMT) carrying the t(1;2)(q21;p23). This study shows the ability of mass spectrometry to identify oncogenic chimeric proteins resulting from chromosomal rearrangements. This strategy can be adapted for the identification of known and unknown translocation partners of chimeric ALK fusion proteins involved in oncogenesis.

immunoprecipitation | tandem mass spectrometry | peptide mapping | multiple enzyme digestion

Chromosomal translocations are among the most frequent category of genetic alterations identified in cancer (1–3). Indeed, chromosomal translocation is a common mechanism of oncogenic activation (1, 2, 4). Chromosomal translocation-mediated deregulation of cancer genes occurs either by the aberrant juxtaposition of an oncogene to another gene or locus or by the generation of a chimeric transcript that encodes a fusion protein with enhanced oncogenic activity (3). A significant proportion of these translocations are recurrent and characteristic of different types of cancer, thus providing a basis for elucidation of the pathogenesis and diagnosis of those tumors (1, 2, 4, 5). Recently, knowledge of genes participating in chromosomal translocations has permitted the development of targeted therapies that inhibit a partner of the chimeric fusion (6, 7). Thus, strategies that facilitate the identification of oncogenic proteins and their translocation partners are crucial to the understanding of tumor biology, the identification of tumor biomarkers, and the development of disease-specific therapeutic approaches.

Existing methods for the identification of chromosomal translocation partners are nucleic acid based. These methods include direct cloning and sequencing after screening of cDNA libraries using probes complementary to the known translocation partner and PCR-based technologies such as rapid amplification of cDNA ends (8–10), inverse PCR (11), inverse panhandle PCR (12, 13), and long-distance PCR (14). The development of mass spectrometry-based methods is desirable and would complement nucleic acid-based methods because of the ability of a peptide sequencing-based approach to distinguish fusion peptides from splice variants.

Tandem mass spectrometry (MS/MS) has emerged as a robust technique for peptide sequencing and identification of proteins. In contrast to Edman degradation sequencing, where sequence determination is achieved by sequential chemical cleavage of amino acids from the amino terminus, MS/MS identifies amino acid sequences by analysis of ionized peptides subjected to fragmentation. This technique provides primary structural information of amino acid composition and sequence (15). In addition, MS/MS is readily amenable to the analysis of proteins with modified amino termini or other modifications that would prohibit Edman sequencing (15). MS/MS is highly sensitive and permits analysis of proteins from a variety of sources (e.g., gel-purified, affinity-purified, and complex mixtures) and does not require purification to homogeneity (16–18). Mass spectrometry-based proteomic approaches could facilitate the identification of translocation partners and would be applicable to the analysis of samples that are limited in quantity, such as clinical tissue biopsies.

The anaplastic lymphoma kinase (*ALK*) gene at 2p23 is frequently involved in chromosomal translocations with an increasing list of partners with formation of oncogenic fusion proteins (19, 20). Native ALK is a receptor tyrosine kinase that belongs to the insulin receptor superfamily and is most similar to leukocyte tyrosine kinase (21). The full-length WT ALK protein is a single-chain 1,620-aa transmembrane protein that is physiologically expressed in cells of neural derivation (21, 22). Chromosomal translocations involving the *ALK* gene are characteristic of two neoplasms: anaplastic large cell lymphoma (ALCL) and inflammatory myofibroblastic tumor (IMT) (19, 20, 23–25). The chimeric proteins consist, at the carboxyl terminus, of the intracytoplasmic portion of the ALK protein, which includes the catalytically active kinase domain and a heterologous fusion partner at the amino terminus. In most instances, the fusion partners encode ubiquitously expressed proteins with active promoters that promote elevated transcription of the chimeric ALK fusion transcript and protein. In addition, the ALK fusion partner proteins frequently contain amino-terminal oligomerization motifs that result in the autophosphorylation and constitutive activation of the ALK tyrosine kinase, the activity of which is critical for ALK-mediated oncogenesis (21, 26, 27). Interestingly, whereas native full-length ALK is localized to the cytoplasmic membrane, chimeric ALK fusion proteins display varying subcellular localizations depending on the partner protein. Thus, the nucleophosmin (NPM)-ALK fusion exhibits nuclear and cytoplasmic distribution reflecting the nucleolar shuttling protein function of NPM. By comparison, tropomyosin 3 (TPM3)-ALK shows a cytoplasmic distribution reflecting the cytoplasmic localization of TPM3, a nonmuscular tropomyosin (21, 28).

This study presents a proteomic strategy to identify translocation partners encoding oncogenic chimeric fusion proteins, using the human ALK oncogene product as a model system. We performed immunoaffinity enrichment of chimeric ALK fusion proteins from total cell lysates of an ALCL-derived cell line and a primary tissue

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office. J.R.Y. is a guest editor invited by the Editorial Board.

Abbreviations: ALCL, anaplastic large cell lymphoma; ALK, anaplastic lymphoma kinase; IMT, inflammatory myofibroblastic tumor; MS/MS, tandem mass spectrometry; NPM, nucleophosmin; TPM3, tropomyosin 3.

‡To whom correspondence may be addressed. E-mail: kojo.elenitobaj@path.utah.edu or megan.lim@path.utah.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0506514103

biopsy of IMT. These tumor samples exhibited the chromosomal aberrations t(2;5)(p23;q35) and t(1;2)q25;2p23), respectively, by conventional cytogenetics and aberrant expression of the ALK protein by tissue immunohistochemistry. The fusion proteins were identified by molecular masses that were different from that of full-length ALK (176 kDa) on immunoblots probed with anti-ALK antibody. The candidate proteins were visualized by silver staining, excised from the SDS/PAGE gel, and subjected to parallel digestion by using four enzymes with different proteolytic cleavage specificities (Fig. 4, which is published as supporting information on the PNAS web site). MS/MS revealed multiple overlapping peptides, identifying the NPM and ALK proteins in the ALCL cell line, and the TPM3 and ALK proteins in the case of the IMT. In each case, we identified fusion peptides representing hybrid sequences from both translocation partners that conclusively established the presence of aberrant fusions of NPM and TPM3 to the truncated ALK protein.

Results

Expression of the Chimeric ALK Fusion Protein in ALCL Cell Line and IMT Tissues. Immunohistochemical studies revealed positive nuclear and cytoplasmic reactivity for the ALK protein in the ALCL cell line and an exclusively cytoplasmic signal in the IMT sample (Fig. 1*A*). The benign lymphocytes and background stromal tissues appropriately do not show ALK protein expression. We also performed Western blot analysis of lysates from the ALCL cell line and IMT tissue using an anti-ALK antibody (Fig. 1*B*). The 80-kDa ALK fusion protein was seen in the ALCL cell line, and a 104-kDa ALK fusion band was seen in the IMT tissue lysate. Detection of ALK protein expression in cells that do not normally express this protein suggests oncogenic alteration of the protein. Additionally, the preferential subcellular localization of the aberrantly expressed protein in the cytoplasmic location in the IMT and nuclear and cytoplasmic distribution in the ALCL cell line suggested chimeric fusion of ALK with a cytoplasmic protein in the clinical IMT sample and a nuclear shuttle protein in the ALCL-derived cell line SUDHL-1.

Immunoprecipitation with Anti-ALK Antibody. To enrich for the ALK fusion protein, we performed immunoprecipitation using an anti-ALK antibody on total lysate obtained from the tissue sample. As the positive immunoprecipitation control, similar experiments were also performed using anti-ALK antibody on SUDHL-1 cell lysate, which contains the NPM–ALK fusion protein. Fig. 1*C* demonstrates separation of the ALK immunocomplex by SDS/PAGE as visualized with silver staining of the separated proteins. Immunoprecipitation using a control IgG antibody was also performed. As shown in Fig. 1*C*, immunoprecipitation of the tissue lysate with the control IgG antibody revealed two predominant bands, one corresponding to the IgG heavy chain (≈50 kDa) and another corresponding to the IgG light chain (≈23 kDa). In comparison, the anti-ALK immunoprecipitate revealed unique bands at ≈80 kDa in the ALCL-derived cell line and 104 kDa in the IMT sample. Importantly, only the 80-kDa and 104-kDa bands were immunoreactive by anti-ALK immunoblotting analysis in the ALCL cell line and IMT sample, respectively (Fig. 1*B*). Corresponding bands on parallel silver-stained 1D SDS/PAGE gels were excised and further processed for analysis by MS/MS.

Fusion Protein Coverage Map by MS/MS. Analysis of the 80-kDa band by MS/MS identified a total of 51 peptides, with peptides ranging in length from 7 to 26 amino acid residues. The top protein database search hits from SEQUEST were summarized by using BIOWORKS BROWSER (ThermoElectron Corporation, San Jose, CA), with peptides identified from the 80-kDa SUDHL-1 gel band matching to UniProt database entries for both full-length ALK (38 peptides) and full-length NPM (10 peptides). In addition, we performed *de novo* sequencing of selected MS/MS spectra, which returned low Xcorr values or did not match the database, to identify candidate fusion peptides (see *Supporting Text*, which is published as supporting information on the PNAS web site). We aligned the *de novo* peptide sequences to the ALK database sequence and

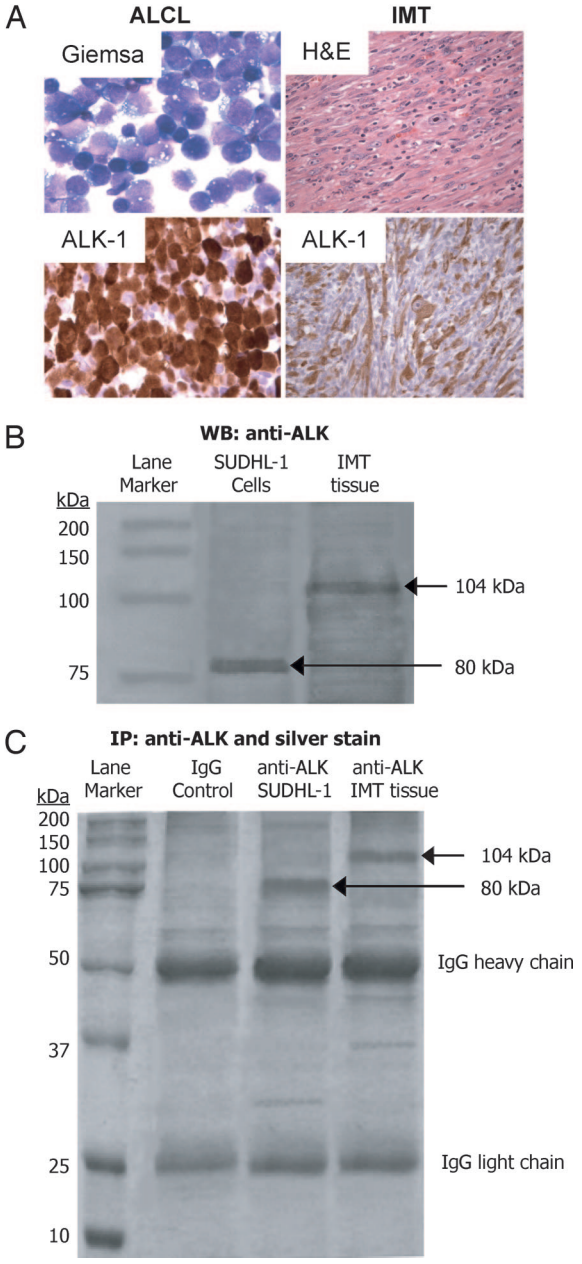


Fig. 1. Analysis of ALK protein expression in ALCL and IMT. (*A*) (*Upper Left*) Giemsa stain of ALCL cell line showing large neoplastic cells with large nuclei and abundant pale cytoplasm. (*Lower Left*) Immunocytochemistry for ALK in ALCL showing nuclear and cytoplasmic distribution of ALK protein. (*Upper Right*) Hematoxylin/eosin (H&E) of IMT tissue biopsy showing a spindle cell neoplasm. (*Lower Right*) Exclusive cytoplasmic reactivity for ALK in the neoplastic cells of the IMT tissue sample. (*B*) Immunoblot analysis of SUDHL-1 cell lysate and IMT tissue lysate using the anti-ALK antibody. The reported NPM–ALK protein was seen in the SUDHL-1 control cell line at 80 kDa, whereas an unknown ALK fusion protein was seen in the IMT tissue lysate at 104 kDa. (*C*) Immunoprecipitation of SUDHL-1 cell lysate and IMT tissue lysate using anti-ALK antibody. Lysates were incubated with ALK antibody, and the antibody selected proteins were resolved by SDS/PAGE. After silver staining, the ALK fusion protein was seen in the IMT lysate pull-down lane at 104 kDa. The IgG control lane and the SUDHL-1 lane containing the 80-kDa NPM–ALK are shown for comparison.

Fig. 2. Fusion mapping of the NPM–ALK protein by MS/MS. The strategy of combining peptide sequences generated by multiple enzyme digests yields enhanced coverage of the partner proteins involved in the chimeric fusion (82.6%). Blue, trypsin; green, lysine-C; orange, glutamic-C; pink, proteinase K.

identified segments of the fusion peptides matching ALK using both our sequence alignment cross correlation tool (Fig. 5, which is published as supporting information on the PNAS web site) and by visual inspection. The non-ALK matching segments of the peptides were then independently searched to reveal the ALK fusion partner. We thus identified three independent and overlapping NPM–ALK fusion peptides: R.LKCGSGPVHISGQHLV-VYR (trypsin), K.CGSGPVHISGQHLV-VYRRK (lysine-C), and G.PVHISGQHLV-VYRRKHQELQ (proteinase K) (Fig. 2 and Table 2, which is published as supporting information on the PNAS web site). The hyphen indicates the fusion site. By combining the peptides identified in all four independent digests, a protein coverage map of the NPM–ALK fusion protein was constructed (Fig. 2). The combined peptide coverage for the NPM–ALK fusion protein was 82.6% [Table 1 (NPM–ALK)]. Fig. 3 shows the mass spectra of a fusion peptide representing the NPM–ALK fusion protein.

Analysis of the 104-kDa ALK immunoreactive band obtained from the anti-ALK antibody pull-down from the IMT tissue identified 60 peptides, with peptides ranging in length from 7 to 29 amino acid residues. Similarly, top database search results were summarized, and peptides identified from the 104-kDa IMT tissue ALK fusion were matched to UniProt database entries for both full-length ALK (33 peptides) and full-length TPM3 (23 peptides). Using the approaches described for the identification of NPM–ALK and detailed in *Supporting Text*, we identified four independent and overlapping TPM3–ALK fusion peptides: K.LEKTIDDLE-VYR (trypsin), K. LEKTIDDLE-VYRRK (lysine-C), E.RSVAKLEKTIDDLE-VYRRKHQE (glutamic-C), and S.VAKLEKTIDDLE-VYRRKHQEL. (proteinase K) (Fig. 6, which is published as supporting information on the PNAS web site, and Table 2). The hyphen indicates the fusion site. By combining the peptides identified in all four independent digests, a protein coverage map of the TPM3–ALK fusion protein was constructed (Fig. 6). The combined peptide coverage for the TPM3–ALK fusion protein was 84.8% [Table 1 (TPM3–ALK)]. The mass spectra of a fusion peptide representing the TPM3–ALK fusion is shown in Fig. 7, which is published as supporting information on the PNAS web site. Identification of fusion peptides for TPM3–ALK using our cross-correlation sequence alignment tool is displayed in Fig. 8, which is published as supporting information on the PNAS web site.

Confirmation of NPM–ALK and TPM3–ALK Fusion Transcripts by RT-PCR. We confirmed the expression of mRNA fusion transcripts by RT-PCR using cDNA from the t(2;5)-positive SUDHL-1 cell line. NPM–ALK PCR yielded the expected product of 467 bp. Similarly, cDNA from the clinical sample with a diagnosis of IMT was subjected to RT-PCR specific for the TPM3–ALK fusion transcript. The TPM3–ALK PCR yielded a 147-bp band

Table 1. Multiple enzyme digest yields improved peptide coverage of protein

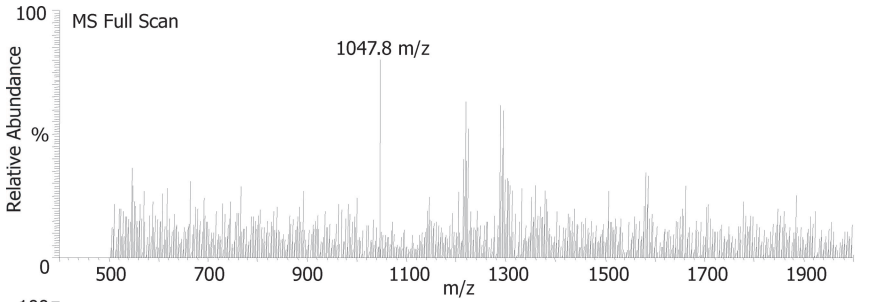
Chimeric fusion	Enzyme digest	Peptide coverage	Avg. peptide length	Combined coverage	Fusion peptide identified
NPM–ALK	Trypsin	29.9%	15	29.9%	Yes
	Lysine-C	26.2%	18	45.6%	Yes
	Glutamic-C	35.7%	17	70.7%	No
	Proteinase K	27.9%	19	82.6%	Yes
TPM3–ALK	Trypsin	34.4%	14	34.4%	Yes
	Lysine-C	39.9%	16	55.8%	Yes
	Glutamic-C	40.8%	15	76.0%	Yes
	Proteinase K	24.5%	17	84.8%	Yes

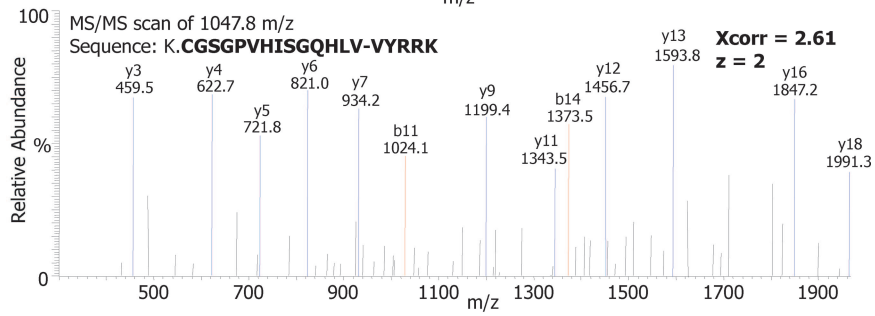
indicative of the presence of TPM3–ALK in the IMT sample (Fig. 9*A*, which is published as supporting information on the PNAS web site).

DNA Sequencing Confirms the NPM–ALK and TPM3–ALK Fusion. The presence of the NPM–ALK fusion in the ALCL cell line was confirmed by bidirectional DNA sequencing. Similarly, the TPM3–ALK fusion was confirmed in the IMT sample. Fig. 9 *B* and *C* demonstrates short segments of the nucleotide sequences obtained from the NPM–ALK and TPM3–ALK amplicons, respectively.

Discussion

We demonstrate the utilization of peptide sequencing by MS/MS for the identification of translocation partners encoding chimeric fusion proteins in primary human cancers. Both the ALCL cell line and IMT samples were shown by conventional cytogenetics to contain rearrangements involving the ALK locus at band 2p23. Immunohistochemistry demonstrated aberrant ALK expression in both tumors, with distinct subcellular localization patterns suggesting different ALK translocation partners in the ALCL and IMT. For identification of the ALK translocation partners, our study used anti-ALK antibody immunoprecipitation to identify unique 80-kDa





	AA	B ions	Y ions	
1	C	104.14	2094.44	19
2	G	161.19	1991.30	18
3	S	248.27	1934.25	17
4	G	305.32	1847.17	16
5	P	402.44	1790.12	15
6	V	501.57	1693.00	14
7	H	638.71	1593.87	13
8	I	751.87	1456.73	12
9	S	838.95	1343.57	11
10	G	896.00	1256.49	10
11	Q	1024.13	1199.44	9
12	H	1161.28	1071.31	8
13	L	1274.43	934.17	7
14	V	1373.57	821.01	6
15	V	1472.70	721.88	5
16	Y	1635.88	622.74	4
17	R	1792.06	459.57	3
18	R	1948.25	303.38	2
19	K	2076.43	147.19	1

fusion ---------- fusion

PROTEIN: NPM-ALK fusion protein

MEDSMDMDMSPLRPQNYLFGCELKADKDYHFKVDNDENEHQLSLRTVSLGAGAKDELHIVEAEAMNYEGSPIKVTLATLK MSVQPTV
SLGGFEITPPVVLRLK**CGSGPVHISGQHLV-VYRRK**HQELQAMQMELQSPEYKLSKLRTSTIMTDYNPNYCFAGKTSSISDLKEVP
RKNITLIRGLGHGAFGEVYEGQVSGMPNDPSPLQVAVKTLPEVCSEQDELDFLMEALIISKFNHQNIVRCIGVSLQSLPRFILLELMAGG
DLKSFLRETRPRPSQPSSLAMLDLLHVARDIACGCQYLEENHFIHRDIAARNCLLTCPGPGRVAKIGDFGMARDIYRASYYRKGGCAML
PVKWMPPEAFMEGIFTSKTDTWSFGVLLWEIFSLGYMPYPSKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAI
ILERIEYCTQDPDVINTALPIEYGPLVEEEEKVPVRPKDPEGVPPLLVSQQAKREEERSPAAPPPLPTTSSGKAAKKPTAAEVSVRVPRGP
AVEGGHVNMAFSQSNPPSELHKVHGSRNKPTSLWNPTYGSWFTEKPTKKNNPIAKKEPHDRGNLGLEGSCTVPPNVATGRLPGASLLL
EPSSLTANMKEVPLFRLRHFPCG NVNYGYQQQGLPLEAATAPGAGHYEDTILKSKNSMNQPGP

Fig. 3. Mass spectra of tryptic peptide K.CGSGPVHISGQHLV-VYRRK identifying the NPM–ALK fusion.

and 104-kDa bands. Immunoblotting with the same antibody demonstrated immunoreactive bands at the same molecular masses. The respective unique bands were excised and subjected to a parallel multiple enzyme digestion strategy that included three enzymes with specific but distinct cleavage activities: i.e., trypsin (K, R), endoproteinase lysine-C (K), endoproteinase glutamic-C (E), and an enzyme with nonspecific cleavage specificity (proteinase K). This digest method generated complementary and overlapping peptides and yielded >80% coverage of the chimeric fusion protein. More importantly, this approach produced several peptides that spanned the translocation point and confirmed the chimeric fusion of the respective partner genes. In both the ALCL and the IMT, the identity of the translocation partners was also confirmed by RT-PCR studies demonstrating the expression of corresponding fusion transcripts and by DNA sequencing that demonstrated in-frame fusion of the translocation partners at the DNA level.

Previous studies have used multiple proteolytic enzymes for the detection of single nucleotide polymorphisms (29) and posttranslational modifications (17). However, both studies involved a combination of the different digests into a single sample tube before analysis by liquid chromatography (LC)-MS/MS, which yields a complex peptide mixture for MS/MS analysis (17, 29). In the current study, we have used a multiple and parallel enzymatic digestion approach (Fig. 4) for the detection of chromosomal translocation-encoded chimeric fusion proteins. For each sample, aliquots from each specific enzyme digestion were processed and subjected to microelectrospray LC-MS/MS separately, with the redundant data from each analysis superimposed to generate a composite sequence map of the chimeric fusion proteins (Figs. 2 and 6). This approach, wherein data from the multiple digests are combined after data acquisition, is advantageous in that it produces simpler peptide mixtures than combinations of multiple enzyme digests in one tube. Thus, whereas individual digests yielded between 25% and 41% coverage for both NPM–ALK and TPM3–ALK, the combination of all four enzymes produced an overall coverage of 82.6% and 84.8%, respectively. Importantly, overlapping fusion peptide sequences were identified in independent endoproteinase digests, thereby providing reliable sequence confirmation of the fusion partners encoded by the hybrid gene.

The utilization of multiple enzymes with distinct cleavage specificities yields various overlapping peptides, which permits high-fidelity sequence coverage of the amino acid sequence at and beyond the junction of the chimeric fusion. The contiguous alignment of several peptides matching the two translocation partners contributes to the fidelity of identification of the gene products participating in the chimeric fusion. Conceivably, this approach can be simplified with the development of high-resolution, online preanalytical techniques that obviate the requirement for gel electrophoresis before analysis by mass spectrometry. The development of such techniques would overcome the drawback of the low-throughput nature of this approach.

The proteomic strategy described in this article is readily applicable to the identification of the participating members of any fusion protein encoded from chromosomal translocation, wherein one of the partners is known and suitable antibodies are available. When immunoprecipitation and Western blotting using an antibody against a known translocation partner yields a protein band with a molecular weight distinct from that of the known protein, this band shift from its expected size raises the possibility of the presence of a chimeric fusion protein. Parallel multiple enzyme digestion and MS/MS substantially increase the coverage of the protein sequence for identification of fusion peptides to confirm the juxtaposition of heterologous sequences from different genes. The strategy used in this article will permit the rapid identification of protein partners encoded by chromosomal translocations and represents a potential opportunity for a discovery method for identification of fusion proteins in malignant neoplasms.

Materials and Methods

Sample Preparation. The SUDHL-1 cell line was obtained from American Type Culture Collection and was maintained in RPMI medium 1640 supplemented with 10% heat-inactivated FCS and antibiotic mixture (GIBCO/BRL).

A frozen tissue sample (1.4 × 1.0 × 2.0 cm) with the pathologic diagnosis of IMT was obtained from Primary Children's Medical Center, Salt Lake City. The diagnosis of IMT was made by three pathologists (M.S.L., K.S.J.E-J., and C.M.C.). All studies using samples from human subjects were approved by the Institutional Board of Review (IRB 11849) of the University of Utah Health Sciences Center.

Tissues were ground in liquid nitrogen, and lysates were prepared by sonication using lysis buffer (1% Triton X-100/10 mM NaCl/40 mM Tris·HCl, pH 8.6) and 0.1% protease inhibitor mixture (Sigma). Cell membranes and other undissolved components were removed by microcentrifugation. Protein concentrations were estimated by using the Bradford colorimetric method against known concentrations of BSA.

Immunohistochemistry. Immunohistochemical studies were performed on 5-μm sections obtained from fixed paraffin-embedded cell blocks or tissues by using the ALK-1 antibody (DAKO), an avidin-biotin peroxidase technique, and an automated immunostainer (Ventana, Tucson, AZ).

Immunoprecipitation and Western Blot Analysis. For immunoprecipitation, lysate containing 1 mg of total protein was precleared by using the appropriate isotype IgG antibody, mixed with 10 μg of anti-ALK antibody (Santa Cruz Biotechnology), and then incubated with gentle shaking overnight at 4°C. Protein G agarose was added to each tube, and the samples were incubated again with gentle shaking at 4°C. The immunocomplex was washed with cold PBS buffer, and the antibody-selected proteins were eluted from the agarose beads by boiling in SDS loading buffer (0.1 M Tris·HCl/10% glycerol/2% SDS/0.05% bromophenol blue/0.1 M DTT) for 5 min. Each sample was resolved by SDS/PAGE and visualized with mass spectrometry compatible silver staining (Invitrogen) or transferred to nitrocellulose membrane for subsequent immunoblot analysis. Similar conditions with appropriate IgG antibody (Santa Cruz Biotechnology) were used for the control lane of each gel. All samples were prepared in replicates of four.

For immunoblotting, lysates containing 50 μg of total protein (per lane) were resolved in a 10% SDS/PAGE gel. Separated proteins were then transferred to a nitrocellulose membrane filter (Millipore) by using semidry transfer and blocked overnight at 4°C. The membrane was incubated at room temperature with the appropriate dilution of primary antibody, and then washed and incubated with horseradish peroxidase-conjugated secondary antibodies. The ALK C-19 antibody (Santa Cruz Biotechnology) was used at 1:1,000 dilution for immunoblotting analysis. Immunoreactive protein bands were visualized by using chemiluminescence (ECL Plus kit; Amersham Pharmacia).

Sample Digestion and Mass Spectrometric Analysis. After staining the gel, the bands of interest were excised, destained and washed, and chopped and dried, then rehydrated in ammonium bicarbonate buffer. To maximize the peptide profiling across the sequence of the fusion protein, replicates of each sample were independently digested by using freshly prepared trypsin, endoproteinase lysine-C, endoproteinase glutamic-C (Princeton Separations, Adelphia, NJ), and proteinase K (Sigma–Aldrich). Sequencing grade endopeptidases were added (1:50), and the tubes were incubated at 37°C. Tubes containing proteinase K were removed after 4 h, and the remaining samples were left to digest overnight. All digested peptides were extracted from the gel pieces by using 50% acetonitrile with 0.1% trifluoroacetic acid (TFA) and were reduced to a

final volume of 30 μl. Each sample was subjected to nanoflow reverse-phase LC/MS by using the LCQ Deca XP ion trap mass spectrometer (ThermoElectron Corporation). Digested peptides were analyzed by using an acetonitrile gradient [0–40% B in 80 min; A = 5% acetonitrile with 0.4% acetic acid and 0.005% heptafluorobutyric acid (HFBA); B = 95% acetonitrile 0.4% acetic acid and 0.005% HFBA] through a reverse-phase column (75-μm ID fused silica packed in-house with 12 cm of 5-μm C18 particles) to elute the peptides at a flow rate of ≈150 nl/min into the mass spectrometer. An electrospray voltage of 1.8 kV was used with the ion transfer tube temperature set to 200°C. Peptide analysis was performed by using data-dependent acquisition of one MS scan (600–2,000 m/z) followed by MS/MS scans of the three most abundant ions in each MS scan. Normalized collision energy was set to 35%, with a 30-ms activation time and isolation width of 1.7 amu. Dynamic exclusion was set to a repeat count of 3, with the exclusion duration of 3 min. For reproducibility and to accommodate the multiple enzyme digest strategy, anti-ALK immunoprecipitation experiments were performed in four replicates.

Data Analysis. Acquired MS/MS spectra were searched with SEQUEST against amino acid sequences in the UniProt protein database (10.15.2004 download). The peptide-matching criteria of a cross-correlation score (X_{corr}) > 1.5 for +1 peptides, > 2.5 for +2 peptides, and > 3.5 for +3 peptides, and a delta correlation score (ΔC_n) > 0.100 was used as a threshold of acceptance. Finally, all SEQUEST results from replicate experiments (trypsin, lysine-C, glutamic-C, and proteinase K digests) were combined to create a coverage map across the fusion protein.

For identification of fusion peptides containing potential ALK and partner sequences, acquired MS/MS spectra that passed the following thresholds of total ion signal intensity (precursor ion count > 2×10^5), acceptable signal-to-noise ratio (S/N > 5:1), and charge state determination (+1, +2, or +3), but with low charge state dependent SEQUEST Xcorr scores (i.e., +1 < 1.5, +2 < 2.5, +3 < 3.5), were selected for peptide sequence interpretation by using the *de novo* algorithm in BIOWORKS (ThermoElectron Corporation) and independently confirmed by visual/manual inspection according to previously described rules (15). The MS/MS spectra that returned low scores or that were unmatched to proteins in the database represent a "candidate pool" of possible ALK fusion peptides. Accordingly, each peptide candidate was used as an input sequence to query against the full-length ALK amino acid sequence provided in the UniProt protein database. Alignment of each of these peptides to the ALK sequence was performed by using both our sequence cross-correlation alignment tool (30) (Figs. 5 and 7) and visual inspection to identify peptide sequence segments that matched ALK. Varying lengths of amino acid residues in the fusion peptides matched ALK (Table 2). The nonmatching amino acid residue segments that were contiguous with the ALK matching segment of the peptide represented the heterologous binding partner of ALK. These nonmatching sequences were then independently subjected to a full protein database search to identify the ALK fusion partners (*Supporting Text* and Table 2).

RNA Extraction and RT-PCR. Total RNA was extracted from the SUDHL-1 cell line and IMT tissue sample by using TRIzol reagent (Life Technologies, Gaithersburg, MD). Integrity of sample RNA was assessed by RT-PCR for the ubiquitously expressed ABL gene. One microgram of RNA was reverse-transcribed by using random primers following the manufacturer's protocol (Invitrogen), and PCR was performed by using ABL forward bp 481 (5′-CCCAACCTTTTCGTTGCACTGT-3′) and ABL reverse bp 866 (5′-CGGCTCTCGGAGGAGACGTAGA-3′) with the following cycling protocol: 95°C for 1 s, 53°C for 5 s, and 72°C for 25 s for 40 cycles on the Rapid Cycler (Idaho Technology, Salt Lake City). Expected product size was 486 bp.

RT-PCR Confirmation Studies. Sample was assessed for the NPM–ALK fusion gene by using a standard probe-based PCR protocol on the Rapid Cycler. PCR was performed with NPM forward bp 296 (5′-AATTACGAAGGCAGTCCA-3′) and ALK reverse primers bp 3485 (5′-TCGTCCTGTTCAGAGC-3′) at 95°C for 1 s, 53°C for 5 s, and 72°C for 25 s for 40 cycles on the Rapid Cycler. Expected product size was 467 bp. Samples were assessed for TPM3–ALK fusion transcripts by using a nested PCR protocol as described (23). All PCR products were subjected to electrophoresis on 2% agarose gels and visualized by UV transillumination.

DNA Sequencing for Confirmation of Fusion Transcripts. Ten microliters of the PCR products were purified (Qiagen, Valencia, CA) and bidirectionally sequenced on ABI-Prism 3100 (Applied Biosystems) using the NPM forward bp 296, the TPM3 forward bp 269, and ALK reverse bp 3399 to confirm the identity of the chimeric fusions as NPM–ALK and TPM3–ALK, respectively.

1. Rabbitts, T. H. (1994) *Nature* **372,** 143–149.
2. Look, A. T. (1997) *Science* **278,** 1059–1064.
3. Futreal, P. A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N. & Stratton, M. R. (2004) *Nat. Rev. Cancer* **4,** 177–183.
4. Rowley, J. D. (2001) *Nat. Rev. Cancer* **1,** 245–250.
5. Helman, L. J. & Meltzer, P. (2003) *Nat. Rev. Cancer* **3,** 685–694.
6. Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J. & Lydon, N. B. (1996) *Nat. Med.* **2,** 561–566.
7. Gschwind, A., Fischer, O. M. & Ullrich, A. (2004) *Nat. Rev. Cancer* **4,** 361–370.
8. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.
9. Frohman, M. A. (1994) *PCR Methods Appl.* **4,** S40–58.
10. Nisson, P. E., Watkins, P. C. & Sacchi, N. (1992) *Cancer Genet. Cytogenet.* **63,** 81–88.
11. Ochman, H., Gerber, A. S. & Hartl, D. L. (1988) *Genetics* **120,** 621–623.
12. Jones, D. H. & Winistorfer, S. C. (1992) *Nucleic Acids Res.* **20,** 595–600.
13. Megonigal, M. D., Rappaport, E. F., Jones, D. H., Kim, C. S., Nowell, P. C., Lange, B. J. & Felix, C. A. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 11583–11588.
14. Willis, T. G., Jadayel, D. M., Coignet, L. J., Abdul-Rauf, M., Treleaven, J. G., Catovsky, D. & Dyer, M. J. (1997) *Blood* **90,** 2456–2464.
15. Steen, H. & Mann, M. (2004) *Nat. Rev. Mol. Cell. Biol.* **5,** 699–711.
16. Wolters, D. A., Washburn, M. P. & Yates, J. R., III (2001) *Anal. Chem.* **73,** 5683–5690.
17. MacCoss, M. J., McDonald, W. H., Saraf, A., Sadygov, R., Clark, J. M., Tasto, J. J., Gould, K. L., Wolters, D., Washburn, M., Weiss, A., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 7900–7905.
18. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M. & Yates, J. R., III (1999) *Nat. Biotechnol.* **17,** 676–682.
19. Morris, S. W., Kirstein, M. N., Valentine, M. B., Dittmer, K. G., Shapiro, D. N., Saltman, D. L. & Look, A. T. (1994) *Science* **263,** 1281–1284.
20. Pulford, K., Morris, S. W. & Mason, D. Y. (2001) *Curr. Opin. Hematol.* **8,** 231–236.
21. Morris, S. W., Naeve, C., Mathew, P., James, P. L., Kirstein, M. N., Cui, X. & Witte, D. P. (1997) *Oncogene* **14,** 2175–2188.
22. Iwahara, T., Fujimoto, J., Wen, D., Cupples, R., Bucay, N., Arakawa, T., Mori, S., Ratzkin, B. & Yamamoto, T. (1997) *Oncogene* **14,** 439–449.
23. Lawrence, B., Perez-Atayde, A., Hibbard, M. K., Rubin, B. P., Dal Cin, P., Pinkus, J. L., Pinkus, G. S., Xiao, S., Yi, E. S., Fletcher, C. D. & Fletcher, J. A. (2000) *Am. J. Pathol.* **157,** 377–384.
24. Griffin, C. A., Hawkins, A. L., Dvorak, C., Henkle, C., Ellingham, T. & Perlman, E. J. (1999) *Cancer Res.* **59,** 2776–2780.
25. Coffin, C. M., Patel, A., Perkins, S., Elenitoba-Johnson, K. S., Perlman, E. & Griffin, C. A. (2001) *Mod. Pathol.* **14,** 569–576.
26. Bischof, D., Pulford, K., Mason, D. Y. & Morris, S. W. (1997) *Mol. Cell. Biol.* **17,** 2312–2325.
27. Mason, D. Y., Pulford, K. A., Bischof, D., Kuefer, M. U., Butler, L. H., Lamant, L., Delsol, G. & Morris, S. W. (1998) *Cancer Res.* **58,** 1057–1062.
28. Pulford, K., Morris, S. W. & Turturro, F. (2004) *J. Cell Physiol.* **199,** 330–358.
29. Gatlin, C. L., Eng, J. K., Cross, S. T., Detter, J. C. & Yates, J. R., III (2000) *Anal. Chem.* **72,** 757–763.
30. Rockwood, A. L., Crockett, D. K., Oliphant, J. R. & Elenitoba-Johnson, K. S. J. (2005) *J. Biomol. Tech.* **16,** 453–458.