

# Gene Expression in Gastrointestinal Stromal Tumors Is Distinguished by *KIT* Genotype and Anatomic Site

Cristina R. Antonescu,<sup>1,7</sup> Agnes Viale,<sup>5</sup>  
 Lisa Sarran,<sup>1</sup> Sylvia J. Tschernyavsky,<sup>1</sup>  
 Mithat Gonen,<sup>2</sup> Neil H. Segal,<sup>3</sup> Robert G. Maki,<sup>4</sup>  
 Nicholas D. Socci,<sup>6</sup> Ronald P. DeMatteo,<sup>3</sup> and  
 Peter Besmer<sup>7</sup>

Departments of <sup>1</sup>Pathology, <sup>2</sup>Biostatistics and Epidemiology, <sup>3</sup>Surgery, and <sup>4</sup>Medicine, Memorial Sloan-Kettering Cancer Center, and <sup>5</sup>Molecular Biology, <sup>6</sup>Computational Biology Center, and <sup>7</sup>Developmental Biology Programs, Sloan-Kettering Institute, New York, New York

## ABSTRACT

**Purpose:** Gastrointestinal stromal tumors (GISTs) are specific *KIT* expressing and *KIT*-signaling driven mesenchymal tumors of the human digestive tract, many of which have *KIT*-activating mutations. Previous studies have found a relatively homogeneous gene expression profile in GIST, as compared with other histological types of sarcomas. Transcriptional heterogeneity within clinically or molecularly defined subsets of GISTs has not been previously reported. We tested the hypothesis that the gene expression profile in GISTs might be related to *KIT* genotype and possibly to other clinicopathological factors.

**Experimental Design:** An HG-U133A Affymetrix chip (22,000 genes) platform was used to determine the variability of gene expression in 28 *KIT*-expressing GIST samples from 24 patients. A control group of six intra-abdominal leiomyosarcomas was also included for comparison. Statistical analyses (*t* tests) were performed to identify discriminatory gene lists among various GIST subgroups. The levels of expression of various GIST subsets were also linked to a modified version of the growth factor/*KIT* signaling pathway to analyze differences at various steps in signal transduction.

**Results:** Genes involved in *KIT* signaling were differentially expressed among wild-type and mutant GISTs. High gene expression of potential drug targets, such as *VEGF*,

*MCSF*, and *BCL2* in the wild-type group, and *Mesothelin* in exon 9 GISTs were found. There was a striking difference in gene expression between stomach and small bowel GISTs. This finding was validated in four separate tumors, two gastric and two intestinal, from a patient with familial GIST with a germ-line *KIT* W557R substitution.

**Conclusions:** GISTs have heterogeneous gene expression depending on *KIT* genotype and tumor location, which is seen at both the genomic level and the *KIT* signaling pathway in particular. These findings may explain their variable clinical behavior and response to therapy.

## INTRODUCTION

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal neoplasms of the digestive tract. The stomach is the most frequent site of origin, followed by the small intestine. Most, if not all, GISTs express the *KIT* receptor, which is known to have diverse roles in several major cell systems during embryogenesis and in the postnatal organism, including hematopoiesis, the pigmentary system, in gametogenesis, and in intestinal pacemaker cells (1–3). *KIT* ligand, *KITL*, is a membrane growth factor, which is the only known agonist of the *KIT* receptor tyrosine kinase (1, 2). Oncogenic activation of *KIT* receptor tyrosine kinase is a central event in GIST pathogenesis and is generally the result of mutations involving either the extracellular or cytoplasmic domains of *KIT* (4). More recently, Heinrich *et al.* (5) described activating mutations in *PDGFRA* in one-third of *KIT* wild-type GISTs. The presence of germ-line gain of function *KIT* mutations in familial GIST syndrome, as well as somatic mutations in morphologically “benign” or incidentally diagnosed GISTs suggest that these mutations play a fundamental role in early GIST development, but it is possible that other, as yet undefined, molecular mechanisms are necessary for malignant progression (6–8).

In comparison with other types of sarcomas, GISTs have been found to have distinctly homogeneous gene expression profiles (9–11). On the basis of a restricted number of diagnostic-specific genes, such as *KIT*, *G protein-coupled receptor (GPR20)*, and *PIK3CG*, GISTs can be easily distinguished from other soft tissue tumors. Transcriptional heterogeneity within clinically or molecularly defined subsets of GISTs has not been previously reported. The purpose of this study was to analyze the consequences of *KIT* genotype and other pathological factors on gene expression profiles in a cohort of well-characterized GISTs. A better understanding of variable expression within different GIST subsets might provide insight into GIST pathogenesis and direct therapy with specific tyrosine kinase inhibitors.

## MATERIALS AND METHODS

*KIT*-positive GIST samples with available frozen tissue were retrieved from the Memorial Sloan-Kettering Cancer Cen-

Received 12/11/03; accepted 2/10/04.

**Grant support:** C. Antonescu was supported by Program Project Grant PO1 CA 47179-10A1 and American Cancer Society grant MRSR. R. DeMatteo was supported by American College of Surgeons Oncology Group and by National Cancer Institute Grant CA94503. P. Besmer was supported by NIH Grant HL/DK55748.

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.

**Requests for reprints:** Cristina R. Antonescu, Department of Pathology, Memorial Sloan-Kettering Cancer Center, 1275 York Ave, New York, NY 10021. Phone: (212) 639-5721; Fax (212) 717-3203; E-mail: antonesc@mskcc.org.

ter tumor bank under an Institutional Review Board (IRB)-approved tissue procurement protocol (IRB no. 00-032). Twenty-eight samples had high-quality RNA suitable for expression profiling experiments and were included in this study. In addition, we included six cases of intra-abdominal/retroperitoneal leiomyosarcomas (LMSs) as a control group.

### Pathological Review

The histology of all 28 GIST samples from 24 patients was reviewed. The following parameters were recorded for each sample: primary tumor location, tumor type (primary, intra-abdominal recurrence, liver metastasis), morphological type (spindle *versus* epithelioid), tumor size, and number of mitoses/50 high power fields (HPF). All 6 of the intra-abdominal/retroperitoneal LMSs used as controls were histologically high grade of the spindle cell type.

### Immunohistochemistry

KIT (CD117) immunohistochemistry was performed in all cases (GIST and LMS) with a polyclonal rabbit antibody (DAKO Corp., Carpinteria, CA), at a 1:500 dilution in citrate buffer. Endogenous mast cells or interstitial cells of Cajal (ICC) from the myenteric plexus were used as internal positive controls. All 28 GIST samples showed strong cytoplasmic KIT immunoreactivity. The six LMS were negative for KIT, but positive for desmin.

### KIT and PDGFR-A Mutation Analyses

DNA was isolated from snap-frozen tumor tissue samples stored at  $-70^{\circ}\text{C}$  using a standard phenol-chloroform organic extraction protocol. One  $\mu\text{g}$  of genomic DNA was subjected to PCR using Platinum *Taq*DNA Polymerase High Fidelity (Life Technologies, Inc. Inc., Gaithersburg, MD). Cases were subjected to PCR amplification using primers for *KIT* exons 9, 11, 13, and 17, and for *PDGFR-A*, exons 12 and 18 (5, 10). The PCR conditions were as follows: (a)  $94^{\circ}\text{C}$  for 4 min; (b)  $94^{\circ}\text{C}$  for 30 s,  $53^{\circ}\text{C}$  for 30 s,  $72^{\circ}\text{C}$  for 30 s (35 cycles); and (c)  $72^{\circ}\text{C}$  for 3 min. The PCR products were identified by agarose gel electrophoresis using a 2% MetaPhor agarose gel (BioWhittaker Applications, Rockland, ME). The expected sizes of the PCR products ranged from 200 to 500 bp in length. The PCR products were purified with the QIAquick PCR Purification kit (Qiagen, Inc., Valencia, CA) before sequencing. Each ABI sequence was compared with a National Center for Biotechnology Information (NCBI) human *KIT* gene nucleotide sequence and was screened using a NCBI Standard Nucleotide Blast Search to determine the location and type of mutation within a particular exon.

### Hybridization of Affymetrix Oligonucleotide Chips

RNA was isolated using the protocol accompanying the RNawiz RNA isolation reagent from Ambion (Austin, TX), and all of the samples were treated on the column with RNase-free DNase (Qiagen, Valencia, CA) according to the manufacturer's instructions. Twenty-five to 50 ng of total RNA were tested for quality on an RNA 6000 Nano Assay (Agilent, Palo Alto, CA) using a Bioanalyzer 2100. RNA with an  $A_{260/280}$  ratio greater than 1.8 were chosen for expression profiling experiments. Two

$\mu\text{g}$  of high-quality total RNA was then labeled according to protocols recommended by the manufacturer. Briefly, after reverse-transcription with an oligo-d<sub>T</sub>-T7 (Genset), double-stranded cDNA was generated with the superscript double-stranded cDNA synthesis custom kit (Invitrogen Life Technologies, Inc., Carlsbad, CA). In an *in vitro* transcription step with T7 RNA polymerase (MessageAmp RNA kit from Ambion), the cDNA was linearly amplified and labeled with biotinylated nucleotides (Enzo Diagnostics, Farmingdale, NY). Ten  $\mu\text{g}$  of labeled and fragmented cRNA were then hybridized onto a test array and a Human Genome U133A expression array (Affymetrix, containing 22,000 transcripts). Posthybridization staining and washing were processed according to instructions from the manufacturer (Affymetrix). Finally, chips were scanned with a Hewlett Packard argon-ion laser confocal scanner.

### Image and Data Analysis

The raw expression data were derived using Affymetrix Microarray Analysis 5.0 (MAS 5.0) software. The data were normalized using a scaling target intensity of 500 to account for differences in the global chip intensity. The expression values were transformed using the logarithm base two. To find genes that associated with different GIST subsets, we applied filtering and statistical analysis constraints to the expression data to exclude those genes that did not vary significantly between comparison groups or that were not expressed at high enough levels. A statistical group analysis was carried out to find genes that showed statistically significant differences in mean expression levels between different subsets of GIST. The log of the normalized expression data were analyzed using an unequal variance *t* test (Welch's approximation) and the addition of the Cross Gene Error Model from the Genespring 5.0 (Silicon Genetics) software. This error model adds an additional intensity-dependent term to the variance of the *t* score. For the GIST/LMS comparison, individual *t* tests were adjusted using the Benjamini-Hochberg False discovery rate, and genes with adjusted  $P_s < 0.05$  were considered significantly different. The gene lists obtained for each individual analysis were cross-referenced against both the published literature and the gene ontology consortium database (<http://www.geneontology.org/>) using NetAffx (<http://www.affymetrix.com>). In addition, two-way hierarchical clustering was performed using the Genespring software with the standard (Pearson) correlation as the similarity metric and centroid linkage clustering.

**Multidimensional Scaling.** Multidimensional scaling was used as an alternative way of visualizing the cluster structure of the data. Multidimensional scaling was performed using S-PLUS software, projecting the data into three dimensions.

**Growth Factor Signaling Pathway Analysis.** We also performed a mechanistic pathway-driven analysis to integrate the complexity of cascade events and intricate pathways involved in GIST progression. We, therefore, compared the level of expression of individual genes involved in growth factor signaling in different subsets of GISTs. A modified *KIT* pathway was adapted from Taylor and Metcalfe (12). The raw expression levels between two groups were compared for statistical significance ( $P < 0.05$ ) using a two-tailed *t* test.

**Venn Diagram.** A “negative” diagram function was used to select the nonoverlapping differentially expressed genes among gene lists generated from different expression analyses.

## RESULTS

### Tumor Samples

The 23 patients with sporadic GISTs had their primary tumors located in the stomach (9 cases, 39%), small bowel (13 cases, 57%), or rectum (1 case, 4%). The majority of the tumors were larger than 10 cm (13 cases, 57%), with only 2 tumors (9%) smaller than 5 cm. Most had a spindle cell morphology (18 cases, 78%). The tissue available for expression analysis was obtained from the primary tumor in 12 patients, an intra-abdominal recurrence in 9 patients, and a liver metastasis in 3 patients. There was one patient with familial GIST who had multiple tumors in the stomach and small bowel in the background of diffuse thickening of myenteric plexus, due to ICC hyperplasia. In this patient, each of these individual tumors was smaller than 5 cm and had spindle cell morphology. Frozen tissue was available from four separate tumors, two located in the stomach and two in the small bowel.

### *KIT* Mutation Analysis

The 24 sporadic GIST samples had the following *KIT* genotype: 5 wild type, 8 *KIT* exon 9 mutations, and 11 *KIT* exon 11 mutations. The results of the *KIT* genotype of 23 of these 24 samples were reported previously (13). In one patient, samples from two subsequent intra-abdominal recurrences were available, which showed an identical *KIT* exon 9 mutation. From the six cases with *KIT* exon 11 point mutations, four were identical V559D substitutions. From the four *KIT* exon 11 deletions, two were identical two-amino-acid deletions, WK557–558del, whereas the other two were bigger deletions (9 and 21 aa, respectively). One case showed an internal tandem duplication at the 3' end of exon 11. No mutations in *KIT* exons 13 or 17, or in *PDGFR-A* exons 12 or 18 were identified. The four different samples available from a familial case had a *KIT* exon 11 W557R germ-line substitution mutation.

### Gene Expression Analysis

To investigate the consequences of the *KIT* genotype and other pathological factors on gene expression profiles, we characterized the transcriptional levels in a cohort of 24 GISTs cases. A *t* test was performed for each gene followed by an adjustment to control the false discovery rate to find genes that showed statistically significant differences in mean expression levels between the following categories: GIST versus LMS, wild-type GIST versus LMS, wild-type versus mutant GIST, exon 11 versus exon 9 GIST, stomach versus small bowel GIST, sporadic versus familial GIST, primary versus recurrent GIST, spindle versus epithelioid GIST. Because our sample size was small and some of the tissues were collected from intra-abdominal recurrences or liver metastasis rather than the primary tumor, we did not attempt a survival analysis.

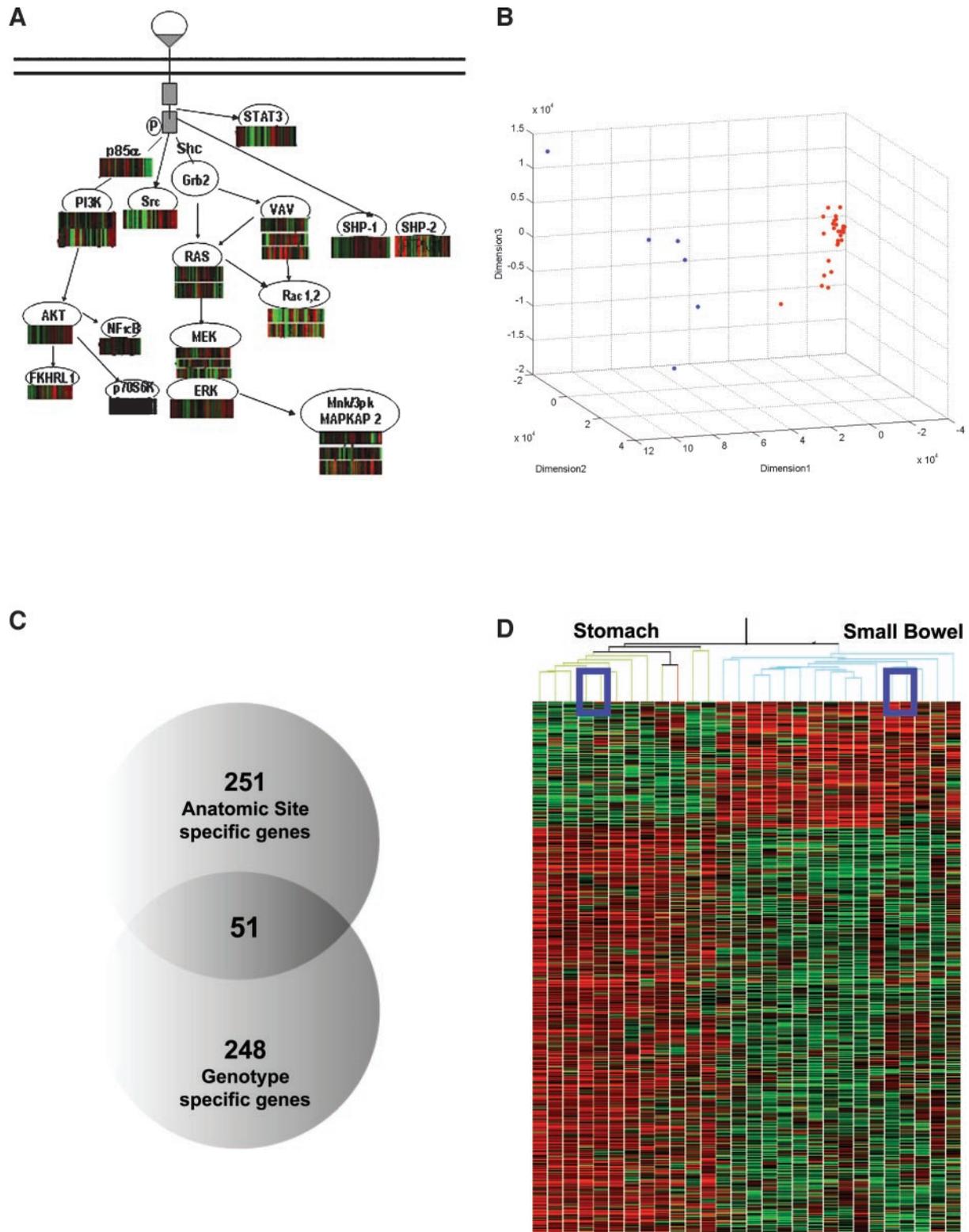
**GIST versus LMS.** We first compared the RNA expression profiles of GIST and LMS using the U133A Affymetrix platform. When comparing the transcriptional levels of genes implicated in growth factor signaling, several of these genes

were found to be differentially expressed (using the false discovery rate after a *t* test) between the two groups. High levels of the p85 phosphatidylinositol 3-kinase subunit, *PIK3R1*, the serine threonine kinase, *AKT/PKB*, the forkhead transcription factor *FKHRL1*, p70 S6 kinase, *P70S6KB*, as well as *SRC*, *RAC1*, *KRAS*, and *ERK (p38)* were identified in GISTs as compared with LMS (Fig. 1A). *KIT ligand (KITL/SCF)* expression was low in most GISTs and did not discriminate them from LMS. Furthermore, as reported previously, GIST and LMS clustered in distinct groups using all of the genes (Fig. 1B). GISTs were characterized by high expression of the receptor tyrosine kinase *KIT*, *G protein-coupled receptor (GPR20)*, and *protein kinase C  $\theta$  (PKC $\theta$ )*. As previously shown, a number of genes encoding ion channels, such as *Na<sup>+</sup>K<sup>+</sup>-ATPase  $\beta$ -1 (ATP1B1)*, *TWIK-related acid-sensitive K<sup>+</sup> channel (TASK)*, and *calcium channel  $\beta$  2 subunit (CACNB2)* were prominently expressed in GISTs. *Proenkephalin (PENK)*, a neuropeptide precursor implicated in gastrointestinal motility, was highly ranked as well. Among cell cycle regulators, *Cyclin D3 (CCND3)* was found to be differentially expressed in GISTs. Even when restricting this analysis to wild-type GIST versus LMS, *KIT* was the number one discriminatory gene, followed by *Annexin A3*, *GPR20*, and *Carbonic anhydrase II*.

**Mutant versus Wild-Type GIST.** A comparison of GISTs with and without *KIT* receptor mutations was carried out next. The growth factor signaling genes encoding the small GTPase *RAC2*, and the tyrosine phosphatase *Shp1* were markedly up-regulated in the mutant GISTs ( $P < 0.00001$ , and  $P = 0.01$ , respectively), whereas *NFkB*, *STAT3*, and *KRAS* were only marginally up-regulated. On the other hand genes involved in apoptosis: *BCL2*, glucose metabolism: *glucose transporter 1 (GLUT1)*, angiogenesis and proliferation: vascular endothelial growth factor (*VEGF*), *macrophage colony stimulating factor (MCSF)*, *interleukin-2 (IL2)* and cancer testis antigen: *MAGE1* were found to be up-regulated in wild-type GISTs.

**Exon 9 versus Exon 11 GIST.** Constitutive activation of the *KIT* receptor generally results from oncogenic mutation involving either extracellular or cytoplasmic domains of the receptors. Our hypothesis that the location of *KIT* mutation might be responsible in generating distinct expression profiles in GISTs appears to be validated by the array results. Three hundred and two genes were identified that distinguish *KIT* exon 9 and exon 11 mutated GISTs (Table 1). Among them, *Mesothelin (MPF)*,  *$\gamma$ -glutamyltransferase (GGT1)*, and genes involved in *wnt* signaling: the frizzled receptors (*FZD2* and *FZD3*) were up-regulated in exon 9 GISTs, whereas *neuregulin 2 (NTAK)*, *ephrin B2 (EphB2)*, *PDGF1*, *Schwannomin-interacting protein 1 (SCHIP-1)*, *EIF3*, *STAT3*, and  $\beta$ -*catenin (CTTNB1)* were up-regulated in exon 11 mutant GISTs.

**Familial versus Sporadic GIST.** There were 47 genes, mostly up-regulated in the familial GIST samples (Table. 1). The list of genes up-regulated in familial GIST included genes involved in synaptic transmission: *D1 dopamine receptor-interacting protein (CALCYON)* and *nitric oxide synthase 1 (NOS1)*; in neurogenesis: *protocadherin  $\beta$  12 (PCDHB12)*; in muscle contraction and development: *calcium channel  $\alpha$  1H subunit (CACNA1H)*; in signal transduction: *interleukin 1 receptor-like 1 (IL1RL1)* and *adrenergic  $\beta$ -3 receptor (ADRB3)*; in apoptosis: *APR-1 protein (MAGEH1)*; and in cell cycle regulation: *RBI*. A



*Fig. 1* A, schematic representation of expression of cytoplasmic mediators of growth factor signaling in 28 gastrointestinal stromal tumor (GIST) samples compared with 6 leiomyosarcomas (LMSs) with overimposed expression signals analyzed by Genespring 5.0. B, multidimensional scaling analysis of 34 samples. The plot displays the 28 GIST (red) and 6 LMS (blue) samples arranged in three-dimensional space. C, “negative” Venn diagram showing 51 overlapping discriminatory genes according to anatomical site and *KIT* genotype. The remaining genes are anatomical-site or genotype specific. D, hierarchical cluster analysis of the 28 GISTs samples according to anatomical site, showing two distinct genomic clusters: gastric GISTs (green) and intestinal GISTs (blue). The rectal GISTs (red) clustered together with the gastric GIST group. The bolded blue squares highlight the four familial cases: the two gastric tumors cluster with the sporadic gastric GISTs, and the two intestinal familial GISTs with the sporadic intestinal GISTs.

**Table 1** List of genes expressed differentially in mutant (MUT) versus wild type (WT), exon 9 versus exon 11, familial versus sporadic, stomach versus small bowel, epithelioid versus spindle

Selective discriminatory genes for each analysis are listed with their corresponding gene designation, GeneBank accession number, *P*, and fold change.

Title	Gene symbol	Probe set	GeneBank	<i>P</i>	Fold change	Locus
<b>WT/MUT</b>						
DNA-damage-induced apoptosis	<i>DDIT1, GADD45</i>	203725_at	NM_001924	0.04	2.5	1p31.2–p31.1
Melanoma antigen, family A, 1	<i>MAGE1</i>	207325_x_at	NM_004988	0.021	2.3	Xq28
Colony stimulating factor 1	<i>MCSF</i>	210557_x_at	M76453	0.01	2.3	1p21–p13
Vascular endothelial growth factor	<i>VEGF</i>	210513_s_at	AF091352	0.02	2.1	6p12
Bcl2	<i>NIP3</i>	201848_s_at	U15174	0.048	2.1	14q11.2–q12
Solute carrier family 2	<i>GLUT, GLUT1</i>	201250_s_at	NM_006516	0.01	1.9	1p35–p31.3
Interleukin 2	<i>IL2</i>	207849_at	NM_000586	0.042	–2.1	4q26–q27
Cyclin G1	<i>CCNG</i>	208796_s_at	BC000196	0.0311	–1.7	5q32–q34
<b>Exon 9/Exon 11</b>						
Mesothelin	<i>MPF, CAK1</i>	204885_s_at	NM_005823	1.49E–04	5.1	16p13.3
Phosphoinositide-3-kinase, CG	<i>PIK3CG</i>	206370_at	NM_002649	0.04	2.6	7q22.2
Stem cell growth factor	<i>SCGF</i>	205131_x_at	NM_002975	0.007	2.5	19q13.3
Frizzled ( <i>Drosophila</i> ) homolog 2	<i>FZD2</i>	210220_at	L37882	0.02	2.5	17q21.1
Gamma-glutamyltransferase 1	<i>GGT1</i>	209919_x_at	L20490	6.50E–04	2.3	22q11.22
Platelet-derived growth factor α	<i>PDGFI</i>	216867_s_at	X03795	0.04	1.9	7p22
Tumor protein p53-binding protein	<i>53BP1</i>	203050_at	NM_005657	0.03	–1.6	15q15–q21
Schwannomin-interacting protein 1	<i>SCHIP-1</i>	204030_s_at	NM_014575	0.04	–1.9	3q25.33
Eukaryotic translation initiation factor 3	<i>EIF3S8</i>	200647_x_at	NM_003752	0.038	–1.9	16p11.2
β-Catenin	<i>CTNNB1</i>	201533_at	NM_001904	0.04	–1.9	3p21
Neuregulin 2	<i>NTAK, DON-1</i>	206879_s_at	NM_013982	0.0083	–2.4	5q23–q33
<b>Familial/Sporadic</b>						
Calcyon	<i>CALCYON</i>	219896_at	NM_015722	2.10E–05	28	10q26.3
Calcium channel, voltage-dependent, α 1	<i>CACNA1H</i>	205845_at	NM_021098	0.009	7	16p13.3
GLI pathogenesis-related 1 (glioma)	<i>GLIPR1</i>	204222_s_at	NM_006851	0.001	4	12q21.1
Phosphatidylinositol (4,5) P 5-phosphatase	<i>PIB5PA</i>	213651_at	AI935720	1.06E–05	4	unk
Nitric oxide synthase 1	<i>NOS1</i>	207310_s_at	U31466	0.003	3	12q24.2–q24.31
APR-1 protein	<i>MAGEH1</i>	218573_at	NM_014061	0.004	2	Xp11.22
<b>Stomach/Small bowel</b>						
Cholecystokinin B receptor		210381_s_at	BC000740	2.57E–04	14.7	11p15.4
Platelet derived growth factor receptor α	<i>PDGFRA</i>	215305_at	H79306	5.43E–05	9.1	4q11–q13
Phospholipase A2, group IVB	<i>PLA2G4B</i>	222256_s_at	AK000550	9.03E–05	6.0	15q11.2–q21.3
Troponin I, skeletal muscle	<i>TNNI2</i>	206393_at	NM_003282	0.001	6.0	11p15.5
CD34 antigen	<i>CD34</i>	209543_s_at	M81104	2.87E–04	5.1	1q32
Latent transforming growth factor receptor b	<i>LTBP-4</i>	204442_x_at	NM_003573	6.56E–05	4.0	19q13.1–q13.2
Transforming growth factor β receptor III	<i>TGFBR3</i>	204731_at	NM_003243	1.55E–04	3.7	1p33–p32
Laminin, α-2	<i>LAMM</i>	205116_at	NM_000426	8.80E–04	2.7	6q22–q23
Cyclin D1 (PRAD1)	<i>CCND1</i>	214019_at	Z23022	0.03	2.7	11q13
Transforming growth factor β-stimulated	<i>pTSC22</i>	215111_s_at	AK027071	2.43E–04	2.4	13q14
Smoothelin	<i>SMTN</i>	207390_s_at	NM_006932	1.00E–02	2.1	22q12.2
Caspase 1	<i>ICE, P45</i>	211368_s_at	U13700	0.007	1.9	11q23
Frizzled ( <i>Drosophila</i> ) homolog 1	<i>FZD1</i>	204451_at	NM_003505	1.30E–02	1.9	7q21
Tropomyosin 1 (α)	<i>TPM1</i>	210987_x_at	M19267	3.00E–02	1.7	unk
Sarcoglycan, epsilon	<i>ESG</i>	204688_at	NM_003919	7.00E–03	1.5	7q21–q22
Myosin, heavy polypeptide 13	<i>MyHC-εo</i>	208208_at	NM_003802	2.00E–03	–3.7	17p13
<b>Epithelioid/Spindle</b>						
SRY-box11	<i>SOX11</i>	204915_s_at	AB028641	0.03	2.7	2p25
Cancer testis antigen 2	<i>CTAG2(CAMEL)</i>	207337_at	NM_020994	0.006	2.7	Xq28
Vascular endothelial growth factor	<i>VEGF</i>	212171_x_at	H95344	0.04	2.6	6p12
Caspase 10	<i>CASP10</i>	210708_x_at	AF111344	6.83E–04	2.6	2q33–q34
Platelet-derived growth factor α	<i>PDGFA</i>	216867_s_at	X03795	0.03	2.0	7p22
Rab geranylgeranyl transferase	<i>RABGGTB</i>	209181_s_at	U49245	0.002	1.9	1p31
Tumor Protein 73	<i>TP73</i>	211195_s_at	AF116771	0.004	1.8	3q27–q29
WAS protein family, member 3	<i>WASF3</i>	204042_at	AB020707	7.77E–04	1.7	13q12
Eukaryotic translation initiation factor 3	<i>EIF3</i>	200005_at	NM_003753	0.02	1.6	22q13.1
Keratin 1	<i>KRT1</i>	205900_at	NM_006121	0.04	1.4	12q12–q13

comparison of growth factor signaling mediators between familial and sporadic GISTs carrying an exon 11 mutation showed that, in the sporadic tumors, *KRAS* and *MKK4* (*JNKK1*) were up-regulated, whereas *ERK* (*p38*) was up-regulated in the familial tumors.

**Small Bowel versus Stomach GIST.** ICCs in the gastrointestinal tract have a common developmental origin, but the development and properties of the respective cellular environments in the esophagus, stomach, small bowel, and large bowel are distinct. It is, therefore, possible that cellular input from the

musculature at the different sites affect the gene expression profile of ICC and presumably in GISTs originating in these different sites. Our gene expression analysis confirms this prediction. A number of genes involved in muscle contraction and development was found to be differentially expressed between these two anatomical sites. *Troponin I*, *tropomyosin I*, *smoothelin*, *laminin*, and *sarcoglycan* were up-regulated in the gastric GISTs, whereas *myosin heavy-chain polypeptide* had a higher expression in the small bowel GISTs (Table 1). Furthermore, genes involved in modulating digestive enzymes and secretion, such as *Cholecystokinin B receptor* and *Phospholipase A2 (PLAG2G4B)* were up-regulated in the stomach location. The growth factor receptors *PDGFRA* and *TGFRBR3*, as well as *LTBP-4*, *TSC22*, were among highly ranked genes in the stomach GISTs. *CD34* was found to be differentially expressed as well, being up-regulated in the gastric GISTs. Among cell cycle regulators, *RBI* and *Cyclins D1 and D2* were found to be in the discriminatory gene list. In addition, when focusing on mediators of growth factor signaling, expression of the class II *PI 3-kinase C2 $\beta$  (PIK3C2B)*, *VAV2*, *Shp1*, *RAC1*, *RAC2*, and *RAC3* were up-regulated in the small bowel. Because an association between exon 9 *KIT* mutation and nongastric location had been reported previously, we wanted to exclude the possibility of identifying false-positive discriminatory genes because of this relationship. We used the negative Venn diagram strategy, which allows super-imposing gene lists obtained from the two individual analyses, based on tumor location and *KIT* genotype (Fig. 1C). As indicated in the figure, two distinct gene lists differentially expressed on each individual analysis were obtained: one list of 251 genes discriminatory in the stomach versus small bowel analysis, and a second list of 248 genes in the exon 9 versus exon 11 analysis. Fifty-one discriminatory genes found in both analyses were, therefore, excluded from further analysis (Fig. 1C). The clustering analysis by anatomical site showed separation of GISTs into two groups: stomach and small bowel (Fig. 1D), further strengthening the results on differential expression. The single rectal GIST clustered together with the gastric tumors. Furthermore, the clustering pattern of the four familial GIST samples followed the same pattern, thus strengthening our hypothesis: the two gastric tumors grouped with the other sporadic gastric GISTs, whereas the two small bowel familial GISTs clustered with the sporadic intestinal GISTs (Fig. 1D).

**Spindle versus Epithelioid GIST.** Cellular spindle and epithelioid shapes are based on distinct cytoskeletal structures. It was, therefore, anticipated that spindle shaped and epithelioid GIST exhibit distinct expression profiles based on these distinct cellular requirements. Genes involved in epithelial development, such as *TP73L* (also known as *TP63*) and *Keratin 1*, were up-regulated between the two subgroups, as were genes involved in apoptosis (*BCL2*, *BCL-G*, *Caspase 10*) and proliferation (*VEGF*, *PDGF1*). Other potential therapeutic targets included *Cancer Testis Antigen 2 (CAMEL)* and *Eukaryotic translation initiation factor 3 (EIF3)*. Four of the five epithelioid GISTs were located either in the stomach ( $n = 2$ ) or small bowel ( $n = 2$ ) and were either wild-type ( $n = 2$ ) *KIT* or had a *KIT* exon 9 mutation ( $n = 2$ ). We compared these four epithelioid GISTs by clustering algorithms using all of the genes to identify whether they cluster in relation to the *KIT* genotype or the

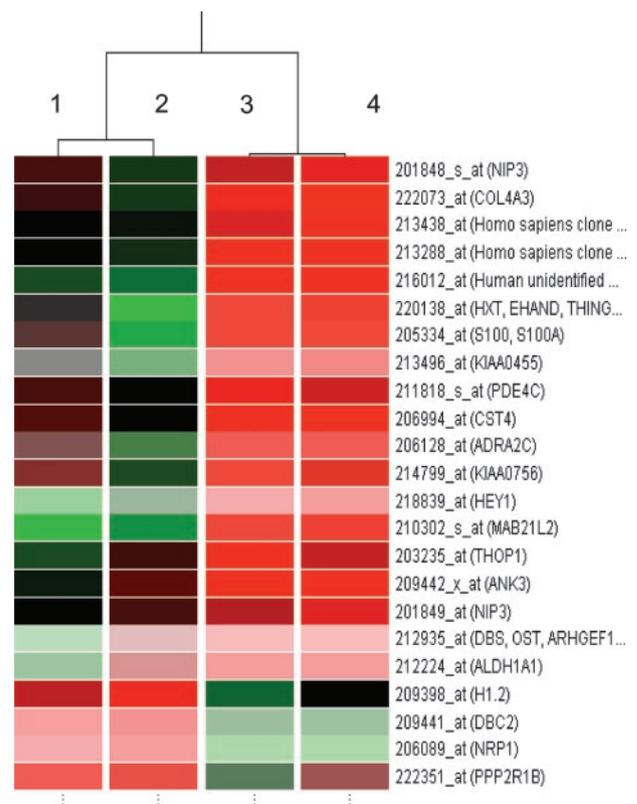


Fig. 2 Hierarchical clustering analysis using all genes of the four epithelioid gastrointestinal stromal tumors (GISTs), showing separation into two groups based on location to stomach or small intestine, rather than mutation type. Epithelioid GISTs (site, *KIT* genotype): column 1, stomach, wild-type (wt); column 2, stomach, exon 9; column 3, small bowel, exon 9; column 4, small bowel, wt.

anatomical location. As shown in Fig. 2, the tumors clustered tightly based on anatomical site, and a number of genes involved in apoptotic pathways discriminated among these two groups.

## DISCUSSION

Constitutive activation of the *KIT* receptor tyrosine kinase by mutation in ICCs is a critical early step in the development of GIST. Furthermore, the observation that *STI571* has dramatic effects on tumor maintenance implies that constitutive *KIT* signaling may be critical for cell survival and proliferation in the fully developed tumor. Therefore, a detailed understanding of the consequences of *KIT* signaling in tumor cells should be relevant to the design of new targeted therapies for GIST. The characterization of the transcriptome of a cell or tissue using DNA microarray analysis has provided a unique tool for the global characterization of cells and tissues. Distinct RNA expression profiles is the result of the unique cellular context as well as the consequence of receptor-mediated signaling cascades. Previous RNA expression profiling studies of different soft tissue sarcomas indicated that GIST expression profiles were distinct and quite homogeneous, in part because of the unique derivation of GIST from ICC (9–11). However, these studies did not dissect the RNA profiles of different pathological

or molecular subsets of GIST, including *KIT* mutation and other clinicopathological factors. In the present study, we have analyzed 28 GIST samples from 24 patients and studied their expression profile with regard to various pathological and molecular characteristics. Furthermore, we attempted to characterize expression of genes involved in *KIT* receptor-mediated signaling.

Previous analyses showed that GISTs are characterized by a distinctive transcriptional signature, which can be applied in tumor diagnosis even when compared with their closest pathological mimic, LMSs (9, 10). Our findings confirm those results in a larger group of GISTs that include a broad spectrum of *KIT* mutations. Among the most prominent discriminatory genes, high expression of tyrosine kinase receptor *KIT*, *G protein-coupled receptor*, and *protein kinase C  $\theta$*  (*PKC $\theta$* ) were the most significant, followed by genes involved in ion transport, such as *lipocortin III* (*annexin 3*), *Na/K ATPase  $\beta$ I*, *potassium channel* (*TASK-1*), and *calcium channel  $\beta$ 2 subunit* (*CACNB*). Furthermore, some genes involved in growth factor-mediated signaling were found to be differentially expressed in GISTs as compared with LMSs. *KIT* function plays a critical role in the differentiation of mesenchymal progenitor cells toward an ICC phenotype during embryonic development and presumably for the expansion of this cell compartment (14, 15). *KIT* function is required for the maintenance of functional ICC networks (14). A number of cell surface receptors and ion channels have been identified in ICCs, including neurokinin 1 (NK1) receptors, VIP, and NO synthase among others (16). Interestingly, a number of highly expressed genes in the familial GIST are involved in synaptic transmission: *D1 dopamine receptor-interacting protein* (*CALCYON*), *calcium channel  $\alpha$ -1H subunit* (*CACNA1H*), and *nitric oxide synthase 1* (*NOS1*). This finding suggested that the familial GISTs are more “differentiated” toward the ICC-lineage as compared with the sporadic counterparts. In other words, the *KIT* mutation seen in the familial cases possibly represents a weaker mutation, giving rise to a less transformed phenotype. Whereas *KIT* activation appears sufficient for ICC hyperplasia in familial GIST cases, additional oncogenic events, involving genes other than *KIT*, are needed to develop discrete GIST lesions (6). The comparative analysis between exon 11 familial versus sporadic GISTs showed that the sporadic tumors had significantly higher expression of *KRAS* and *MKK4* (*JNKK1*), genes thought to be involved in *KIT* signaling. An interesting association of up-regulated genes, such as *BCL2*, *VEGF*, *IL2*, *MCSF*, were found in the wild-type GISTs, when compared with the *KIT* mutant GISTs. In contrast, *RAC2* and *Shp1*, involved in receptor signaling, were markedly up-regulated in the mutant GISTs.

Although initial studies suggested that exon 11 *KIT* mutations are more common in “malignant” than in “benign” GISTs (7, 17), others have failed to find a significant association between *KIT* mutation status and histological grade (13, 18). The impact of *KIT* genotype on outcome seems to be limited to a small fraction of GISTs, characterized by extracellular domain of *KIT* mutations (13, 19). *KIT* exon 9 mutations define a distinct subset of GIST, because all of the cases reported to date show an identical duplication 1530 bp, encoding for Ala-Tyr, and most of them have been associated with an intestinal location and more aggressive clinical behavior (13, 19, 20). The

genomic signature of *KIT* exon 9 GISTs included high levels of *Mesothelin* (*MPF*), *MMP1*, and  $\gamma$ -*glutamyltransferase* (*GGT1*) and low levels of *neuregulin 2* (*NTAK*) and *EphB2*, as compared with exon 11 mutated tumors. A number of these genes represent potential therapeutic targets, e.g., mesothelin. Recombinant anti-mesothelin immunotoxin was recently shown to have cytotoxic effects by inducing apoptosis in lung and ovarian cancer (21, 22).

Anatomical site-specific variations in morphology, clinical outcome, and, more recently, site-specific *KIT* mutations have been reported in GIST, and the basis for these differences remains unclear (23). Anatomical site differences in ICC distribution and ultrastructural appearance have been recognized between normal human stomach and small bowel (24). Human intestinal myenteric and deep muscular plexus ICCs show more pronounced myoid features, resembling smooth muscle cells ultrastructurally, than other locations. Also, expression of embryonic smooth muscle myosin heavy chain (MHC-SMem) has been reported in *KIT*<sup>+</sup> ICC in the normal gut as well as in GISTs (25). The same authors point out that *KIT*<sup>+</sup> MHC-SMem<sup>+</sup> ICCs were also CD34 positive in the stomach and colon, although negative for CD34 in the small bowel, suggesting that the ICCs in the human gastrointestinal tract are heterogeneous (25). Thus, *CD34* as well as a number of genes involved in muscle development and contraction, such as *tropomyosin 1*, and *tropomyosin 1*, were up-regulated in the gastric-located GISTs, whereas *myosin heavy-chain polypeptide* was higher in the small bowel location. These findings might also explain the somewhat different immunoprofiles of gastric versus intestinal GIST, as a reflection of the degree of smooth muscle differentiation (26). In contrast, Allander *et al.* (10) found *CD34* to be one of the discriminatory genes between GIST and other types of sarcomas, but the exact location of the primary tumors is not evident from their study. Other site-dependent differentially expressed genes included growth factor receptors *PDGFRA*, and *TGFRBR3*, and *LTBP-4*, *TSC22*, which were all up-regulated the stomach GISTs. In addition, small bowel tumors showed high RNA expression levels of *PIK3C2B*, *VAV2*, *Shp1*, and *RAC1-3*, as compared with the gastric GISTs. Furthermore, the four samples of familial GIST and the four epithelioid GISTs in our series showed a distinctive clustering related to the anatomical site. Using a dual approach, hierarchical cluster analysis, and negative Venn diagram, we were able to demonstrate a distinct gene expression profile, independently related to both anatomical site and *KIT* genotype.

Variations in morphology with tumor location have been previously described, such as epithelioid tumors occurring far more often in the stomach, whereas spindle cell lesions of the small bowel commonly show an organoid pattern and skeinoid fibers (23). More recently, site-specific *KIT* mutations have been suggested, i.e., *KIT* exon 11 internal tandem duplication in the stomach and *KIT* exon 9 mutations predominantly seen in the small bowel location (13, 19). The incidence of exon 11 *KIT* mutations does not appear to be related to tumor site (27, 28). Further associations have been identified, such as most of the gastric epithelioid GISTs lacking *KIT* mutations (13, 29, 30), suggesting that an alternative mechanism of *KIT* activation is responsible for tumorigenesis. Epithelioid GISTs expressed

genes characteristic of the epithelial cell phenotype, such as *TP73L* and *Keratin1*. Furthermore, genes involved in apoptosis (*BCL2*, *Caspase 10*), angiogenesis, (*VEGF*), and proliferation (*PDGFR*) were up-regulated in the epithelioid as compared with the spindle cell GISTs. Potential therapeutic targets included *Cancer Testis Antigen 2 (CAMEL)* and *Eukaryotic translation initiation factor 3 (EIF3)*, which were expressed in the epithelioid tumors.

Signal transduction inhibition as cancer therapy was first tested successfully with imatinib mesylate (formerly known as STI571), a selective small-molecule tyrosine kinase inhibitor, with specificity for the Bcr-Abl, KIT, and PDGFR tyrosine kinases, in chronic myelogenous leukemia and subsequently in GIST (31). After the initial success, STI571 resistance is now being encountered not only in chronic myelogenous leukemia (32, 33), but also in patients with GIST who had an initial therapeutic response (34). There is mounting evidence that novel drug agents, with alternative or complementary mode of action to STI571, are needed to sustain response or to prevent the development of resistance in high-risk GISTs. We hypothesize that the gene expression signature can be used not only to identify potential candidate genes for alternative and novel therapeutic targeting but also to design therapeutic intervention tailored for each individual GIST subset. Other KIT tyrosine kinase inhibitors with anti-VEGF receptor inhibitor activity (e.g., PTK787, SU011248; Ref. 35) might have a greater activity than STI571 in specific groups of GISTs, such as *KIT* wild-type and/or epithelioid GISTs. A better understanding of the role of oncogenic kinase mutations in human tumorigenesis might reveal insights into selective inhibition of aberrant signal transduction or novel kinase-targeted therapies.

In this study, we have analyzed the gene expression profiles of a group of GISTs, using a genome-wide oligonucleotide platform. We identified distinct transcriptional profiles related to both *KIT* genotype and to anatomical location in GISTs, neither of which has been previously reported. These newly described genomic subsets of GISTs will provide useful information related to pathogenesis and to potential new therapeutic targets.

## REFERENCES

- Besmer P. The kit ligand encoded at the murine Steel locus: a pleiotropic growth and differentiation factor. *Curr Opin Cell Biol* 1991; 3:939–46.
- Besmer P. Kit-ligand-stem cell factor. *Colony-stimulating factors. Molecular and cell biology*. 2nd ed. New York: Marcel Dekker; 1997. p. 369–404.
- Sanders KM, Ordog T, Koh SD, Torihashi S, Ward SM. Development and plasticity of interstitial cells of Cajal. *Neurogastroenterol Motil* 1999;11:311–38.
- Hirota S, Isozaki K, Moriyama Y, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science (Wash DC)* 1998;279:577–80.
- Heinrich MC, Corless CL, Duensing A, et al. PDGFRA activating mutations in gastrointestinal stromal tumors. *Science (Wash DC)* 2003; 299:708–10.
- Heinrich MC, Rubin BP, Longley BJ, Fletcher JA. Biology and genetic aspects of gastrointestinal stromal tumors: KIT activation and cytogenetic alterations. *Hum Pathol* 2002;33:484–95.
- Taniguchi M, Nishida T, Hirota S, et al. Effect of c-kit mutation on prognosis of gastrointestinal stromal tumors. *Cancer Res* 1999;59:4297–300.
- Nishida T, Hirota S, Taniguchi M, et al. Familial gastrointestinal stromal tumors with germline mutation of the KIT gene. *Nat Genet* 1998;19:323–4.
- Nielsen TO, West RB, Linn SC, et al. Molecular characterization of soft tissue tumours: a gene expression study. *Lancet* 2002;359:1301–7.
- Allander SV, Illei PB, Chen Y, et al. Expression profiling of synovial sarcoma by cDNA microarrays: association of ERBB2, IGFBP2, and ELF3 with epithelial differentiation. *Am J Pathol* 2002;161: 1587–95.
- Segal NH, Pavlidis P, Antonescu CR, et al. Classification and subtype prediction of adult sarcoma by functional genomics. *Am J Pathol* 2003;163:691–700.
- Taylor ML, Metcalfe DD. Kit signal transduction. *Hematol Oncol Clin North Am* 2000;14:517–35.
- Antonescu CR, Sommer G, Sarraf L, et al. Association of KIT exon 9 mutations with non-gastric primary site and aggressive behavior: KIT mutation analysis and clinical correlates of 120 gastrointestinal stromal tumors. *Clin Cancer Res* 2003;9:3329–37.
- Ward SM, Burns AJ, Torihashi S, Sanders KM. Mutation of the proto-oncogene c-kit blocks development of interstitial cells and electrical rhythmicity in murine intestine. *J Physiol* 1994;480:91–7.
- Ward SM, Burns AJ, Torihashi S, Harney SC, Sanders KM. Impaired development of interstitial cells and intestinal electrical rhythmicity in steel mutants. *Am J Physiol* 1995;269:C1577–85.
- Huizinga JD, Thuneberg L, Vanderwinden J-M, Rumessen JJ. Interstitial cells of Cajal as targets for pharmacological intervention in gastrointestinal motor disorders. *Trends Pharmacol Sci* 1997;18:393–403.
- Ernst SI, Hubbs AE, Przygodzki RM, Emory TS, Sobin LH, O'Leary TJ. KIT mutation portends poor prognosis in gastrointestinal stromal/smooth muscle tumors. *Lab Invest* 1998;78:1633–6.
- Rubin BP, Singer S, Tsao C, et al. KIT activation is a ubiquitous feature of gastrointestinal stromal tumors. *Cancer Res* 2001;61:8118–21.
- Lasota J, Wozniak A, Sarlomo-Rikala M, et al. Mutations in exons 9 and 13 of KIT gene are rare events in gastrointestinal stromal tumors. A study of 200 cases. *Am J Pathol* 2000;157:1091–5.
- Lux ML, Rubin BP, Biase TL, et al. KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. *Am J Pathol* 2000; 156:791–5.
- Fan D, Yano S, Shinohara H, et al. Targeted therapy against human lung cancer in nude mice by high-affinity recombinant antimesothelin single-chain Fv immunotoxin. *Mol Cancer Ther* 2002;1:595–600.
- Hassan R, Lerner MR, Doris Benbrook, et al. Antitumor activity of SS(dsFv) PE38 and SS1(dsFv)PE38, recombinant antimesothelin immunotoxins against human gynecologic cancers grown in organotypic culture in vitro. *Clin Cancer Res* 2002;8:3520–6.
- Fletcher CD, Berman JJ, Corless C, et al. Diagnosis of gastrointestinal stromal tumors: a consensus approach. *Hum Pathol* 2002;33:459–65.
- Vanderwinden J-M, Rumessen JJ. Interstitial cells of Cajal in human gut and gastrointestinal disease. *Microsc Res Tech* 1999;47:344–60.
- Sakurai S, Fukasawa T, Chong JM, Tanaka A, Fukayama M. Embryonic form of smooth muscle myosin heavy chain (SMemb/MHC-B) in gastrointestinal stromal tumor and interstitial cells of Cajal. *Am J Pathol* 1999;154:23–8.
- Miettinen M, Sobin LH, Sarlomo-Rikala M. Immunohistochemical spectrum of GISTs at different sites and their differential diagnosis with a reference to CD117. *Mod Pathol* 2000;13:1134–42.
- Lasota J, Jasinski M, Sarlomo-Rikala M, Miettinen M. Mutations in exon 11 of c-Kit occur preferentially in malignant versus benign gastrointestinal stromal tumors and do not occur in leiomyomas or leiomyosarcomas. *Am J Pathol* 1999;154:53–60.

28. Miettinen M, Sarloma-Rikala M, Lasota J. Gastrointestinal stromal tumors: recent advances in understanding of their biology. *Hum Pathol* 1999;30:1213–20.
29. Wardelmann E, Neidt I, Bierhoff E, et al. c-kit mutations in gastrointestinal stromal tumors occur preferentially in the spindle rather than in the epithelioid cell variant. *Mod Pathol* 2002;15:125–36.
30. Debiec-Rychter M, Lasota J, Sarlomo-Rikala M, Kordek R, Miettinen M. Chromosomal aberrations in malignant gastrointestinal stromal tumors: correlation with c-KIT gene mutation. *Cancer Genet Cytogenet* 2001;128:24–30.
31. Demetri G. Identification and treatment of chemoresistant inoperable or metastatic GIST: experience with the selective tyrosine kinase inhibitor imatinib mesylate (STI571). *Eur J Cancer* 2002;38:S52–9.
32. Gorre ME, Mohammed M, Ellwood K, et al. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science (Wash DC)* 2001;293:876–80.
33. Shah NP, Nicoll NJ, Nagar B, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* 2002;2:117–25.
34. Fletcher JA, Corless CL, Dimitrijevic S, et al. Mechanisms of resistance to imatinib mesylate (IM) in advanced gastrointestinal stromal tumor (GIST) [abstract]. *Proc Am Soc Clin Oncol* 2003;22:815.
35. Heinrich MC, Blanke CD, Druker BJ, Corless CL. Inhibition of KIT tyrosine kinase activity: a novel molecular approach to the treatment of KIT-positive malignancies. *J Clin Oncol* 2002;20:1692–703.