Original Article

Integrative Genomic Characterization and a Genomic Staging System for Gastrointestinal Stromal Tumors

Antti Ylipää, MSc[1,2]; Kelly K. Hunt, MD[3]; Jilong Yang, MD, PhD[1,4]; Alexander J. F. Lazar, MD, PhD[1,5]; Keila E. Torres, MD[3]; Dina C. Lev, MD[5,6]; Matti Nykter, PhD[2]; Raphael E. Pollock, MD, PhD[3,5]; Jonathan Trent, MD, PhD[5,7]; and Wei Zhang, PhD[1]

BACKGROUND: Gastrointestinal stromal tumors (GISTs) historically were grouped with leiomyosarcomas (LMSs) based on their morphologic similarities; however, recently, GIST was established unequivocally as a distinct type of sarcoma based on its molecular features and response to imatinib treatment. **METHODS:** To gain further insight into the genomic differences between GISTs and LMSs, the authors mapped gene copy number aberrations (CNAs) in 42 GISTs and 30 LMSs and integrated the results with gene expression profiles. **RESULTS:** Distinct patterns of CNAs were revealed between GISTs and LMSs. Losses in 1p, 14q, 15q, and 22q were significantly more frequent in GISTs than in LMSs ($P < .001$); whereas losses in chromosomes 10 and 16 and gains in 1q, 14q, and 15q ($P < .001$) were more common in LMSs. By integrating CNAs with gene expression data and clinical information, the authors identified several clinically relevant CNAs that were prognostic of survival in patients with GIST. Furthermore, GISTs were categorized into 4 groups according to an accumulating pattern of genetic alterations. Many key cellular pathways were expressed differently in the 4 groups, and the patients in each group had increasingly worse prognoses as the extent of genomic alterations increased. **CONCLUSIONS:** Based on the current findings, the authors proposed a new tumor-progression genetic staging system termed genomic instability stage to complement the current prognostic predictive system based on tumor size, mitotic index, and v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) mutation. *Cancer* 2011;117:380–9. © 2010 American Cancer Society.

KEYWORDS: imatinib, copy number aberrations, gastrointestinal stromal tumors, leiomyosarcoma, array comparative genomic hybridization, survival, v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog.

Gastrointestinal stromal tumors (GISTs) previously were grouped with spindle cell and other soft-tissue sarcomas, including leiomyosarcoma (LMS).[1] However, in recent years, GIST has emerged as a distinct mesenchymal tumor type that frequently is associated with a gain-of-function mutation in the v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) gene (80%-85% of GISTs) or the platelet-derived growth factor-alpha (*PDGFRA*) gene (5%-7% of GISTs).[1-3] The presence of these mutations allows for targeted therapy using imatinib (Gleevec, STI-571; Novartis Pharmaceuticals, Basel, Switzerland), which has demonstrated efficacy in 60% to 80% of patients with GISTs.[4] Conversely, LMSs are not associated with *KIT* gene mutations or overexpression and do not benefit from imatinib therapy. The treatment of patients with LMS using contemporary cytotoxic chemotherapy has resulted in a 53% objective response rate, whereas patients with GIST who received traditional cytotoxic chemotherapy have not had a measurable response.[4,5] Although mutations in *KIT* and *PDGFRA* explain why 60% to 80% of patients with GIST initially benefit from imatinib,

the duration of benefit that patients receive from this therapy remains considerably variable. Furthermore, even rarely, some patients with *KIT* exon 11 mutations are resistant to imatinib; and secondary mutations of *KIT* reportedly have occurred in patients who initially responded to imatinib therapy.[6] Thus, robust and biologically relevant prognostic factors, especially those for predicting the survival of patients with GIST, still are needed.

Growing evidence indicates that the accumulation of specific genetic alterations ultimately leads to a highly unstable underlying genome in cancer development and progression.[7] Although some recurrent changes in GIST and LMS genomes have been investigated before, the deficiencies of early measurement technologies or small sample sizes that were used in early studies makes it necessary to accumulate additional genomic information in additional samples and to create a more refined map of the recurrent aberrations. Toward this objective, we conducted a comprehensive, high-resolution, whole-genome array comparative genomic hybridization (aCGH) analysis to map the recurrent copy number aberrations (CNAs) in GISTs and LMSs. We also investigated the clinical relevance of our results in an integrative analysis of the CNAs, gene expression profiles, and patient survival information. The results from this study led us to propose a new tumor-progression genetic staging system termed genomic instability stage (GIS) to complement the current GIST staging system, which is based on tumor size, mitotic index (MI), and c-kit mutation.

MATERIALS AND METHODS

Primary Tumors and Pathologic Evaluation

In total, 72 primary tumors, including 42 GISTs and 30 LMSs, were acquired from surgical specimens from 1989 through 2005 at The University of Texas M. D. Anderson Cancer Center under an institutional review board–approved protocol. For transcriptome analysis, high-quality RNA was acquired from 32 GISTs and 25 LMSs. For genomic profiling, we used these samples as well as 15 additional samples (10 GISTs and 5 LMSs). The diagnoses were made on the basis of clinicopathologic evaluation and molecular marker studies. The clinical information is summarized in Table 1.

Array Experiments and Preprocessing

Genomic DNA from tumors and pooled normal tissue was isolated according to standard procedure. Labeled

Table 1. Clinical Information on Patients With Gastrointestinal Stromal Tumors and Leiomyosarcomas

Category	Mean±SD or No. of Patients (%)
GIST, n=42	
Age, y[a]	62±14
Tumor size, cm	12.7±8.8
Sex	
Men	19 (45)
Women	23 (55)
Primary site	
Stomach	19 (45)
Small bowel	20 (48)
Large bowel	2 (5)
Uterus	1 (2)
Disease status	
Primary no metastasis	19 (45)
Primary with metastasis	7 (17)
Local recurrence without metastasis	6 (14)
Recurrence metastasis	10 (24)
KIT status	
Wild type	11 (27)
Exon 11 mutation	27 (64)
Exon 9 mutation	3 (7)
No data	1 (2)
Imatinib treatment	
Preoperative imatinib	25 (60)
No preoperative imatinib	17 (40)
Length of follow-up, mo[b]	61±44
LMS, n=30	
Age, y	57±12
Tumor size, cm	15.3±9.1
Sex	
Men	7 (23)
Women	23 (77)
Primary site	
Uterus	9 (30)
Retroperitoneal mass	6 (20)
Inferior vena cava	5 (17)
Other	10 (33)
Length of follow-up, mo[b]	41±27

SD indicates standard deviation; GIST, gastrointestinal stromal tumor; LMS, leiomyosarcoma; *KIT*, v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog.
[a] Age refers to the patient age at which the sample was obtained.
[b] Length of follow-up was calculated as the time between the first diagnosis and the last contact.

genomic DNA was hybridized to the Agilent Human Genome CGH Microarray 4x44 Kit according to the manufacturer's instructions (Agilent Technologies, Palo Alto, Calif). The data were extracted from microarrays with Agilent Feature Extraction software version 9.5 using the default settings and were analyzed further with MATLAB version R2007b (The MathWorks, Inc., Natick, Mass) and R statistical software (version 2.6.2; R Development Core Team, available at http://cran.r-project.org/ accessed August 26, 2010). Intensity values

were lowess-normalized to compensate for common nonlinear biases. Ratios of normalized intensity values from tumor tissues and normal tissue were transformed to $\log_2$-space. Then, log-ratio data were subjected to a circular binary segmentation algorithm[8] (R implementation DNA copy; version 1.6.0) to reduce the effect of noise. The CGHcall algorithm[9] (version 1.2.2. in R) was used to label the segments as lost, normal, or gained.

Gene expression data were measured by using whole human genome oligo arrays with 44-K 60-mer probes (Agilent Technologies) with 500 ng of total RNA starting material according to the manufacturer's protocol. Arrays were scanned with the Agilent dual laser-based scanner. Features were extracted from arrays with Agilent Feature Extraction software (version 8.0). The expression data were quantile normalized. Both aCGH and gene expression data are available at http://www.cs.tut.fi/sgn/csb/GISTLMS/ accessed August 26, 2010.

Statistical Analyses

DNA sequences were classified as recurrently aberrated if the number of aberrations in individual samples exceeded a threshold of statistical significance, as estimated using a permutation test. The 95th percentile values were chosen as the threshold of significance. By using this procedure, we estimated that similar aberrations in at least 14 samples (33%) for GISTs and 12 samples (40%) for LMSs were required for a sequence to be called recurrently aberrant. Probe average recurrence (PAR) was used to quantify the aberration rate of a recurrently aberrated DNA segment. The PAR is calculated by averaging the aberration rate over the probes in a contiguous, recurrently aberrated DNA segment. Differences in aberration frequencies between GIST and LMS were tested independently with the Fisher exact test for each probe. To account for the resulting multiple comparisons problem, the level of significance in these tests was set to .001. Differential expression between sample sets was determined with the Wilcoxon rank-sum test with a threshold of .05. In finding the subgroups within the GIST samples, hierarchical clustering with inner squared linkage was applied. The most informative genes for clustering were selected by using a 2-tailed t test. In estimating patient survival curves, Kaplan-Meier survival estimators were applied. A Mantel-Cox test also was used to determine the statistical significance of the difference of these survival estimators. A significance threshold of .05 was selected for all survival tests. A hypergeometric distribution with a significance threshold of .05 was used in computing gene set enrichments.

RESULTS

Gastrointestinal Stromal Tumors and Leiomyosarcomas Have Distinct Differences in Their Genomes

After performing comprehensive aCGH profiling experiments with 42 primary GISTs and 30 primary LMSs, we analyzed the recurrent CNAs in these tumors. Our analysis revealed several distinct loci throughout the genome that frequently were aberrant in GISTs (Fig. 1A) and in LMSs (Fig. 1B), similar to the reported data.[10] Statistical comparisons of the concluded cancer genomes revealed that losses in chromosomes 1p, 14q, 15q, and 22q were significantly more frequent in GISTs than in LMSs ($P <$ .001), whereas losses in chromosomes 10 and 16 were more common in LMSs ($P <$ .001). We not only confirmed previous CNAs, such as loss of 1p, 14q, 15q, and 22q,[10-21] but also demonstrated that the deletion of 22q was the most common recurrent deletion in GISTs (84% PAR): parts of 22q were deleted in >95% of GIST samples, a rate that was significantly higher than previously reported data.[12,21] In addition, although losses in 1p were common in both sarcoma types, many more and much larger deletions in 1p were observed in GISTs than in LMSs. In comparison, tumors from patients with LM more frequently had gains in chromosomes 1q, 14q, and 15q ($P < $ .001).

From the aberration profiles illustrated in Figure 1, we created a gene-level map of the recurrent CNAs in GISTs and LMSs. In total, 328 recurrently aberrant segments of DNA were identified in GISTs (202 gains and 126 losses), and 373 were identified in LMSs (194 gains and 179 losses) based on the PAR, which was defined as the average recurrence rate of the probes that were included in a segment. We matched CNAs with corresponding gene expression profiles and identified which genes had expression that was correlated significantly with gene dosage. Next, we investigated the effect of each dosage-sensitive gene on patient survival and identified which recurrent CNAs harbored at least 1 dosage-sensitive gene that was correlated significantly with patient survival (Table 2). These clinically relevant CNA segments and the putative target genes offer a promising starting point from which functional validations can be carried out in future studies.
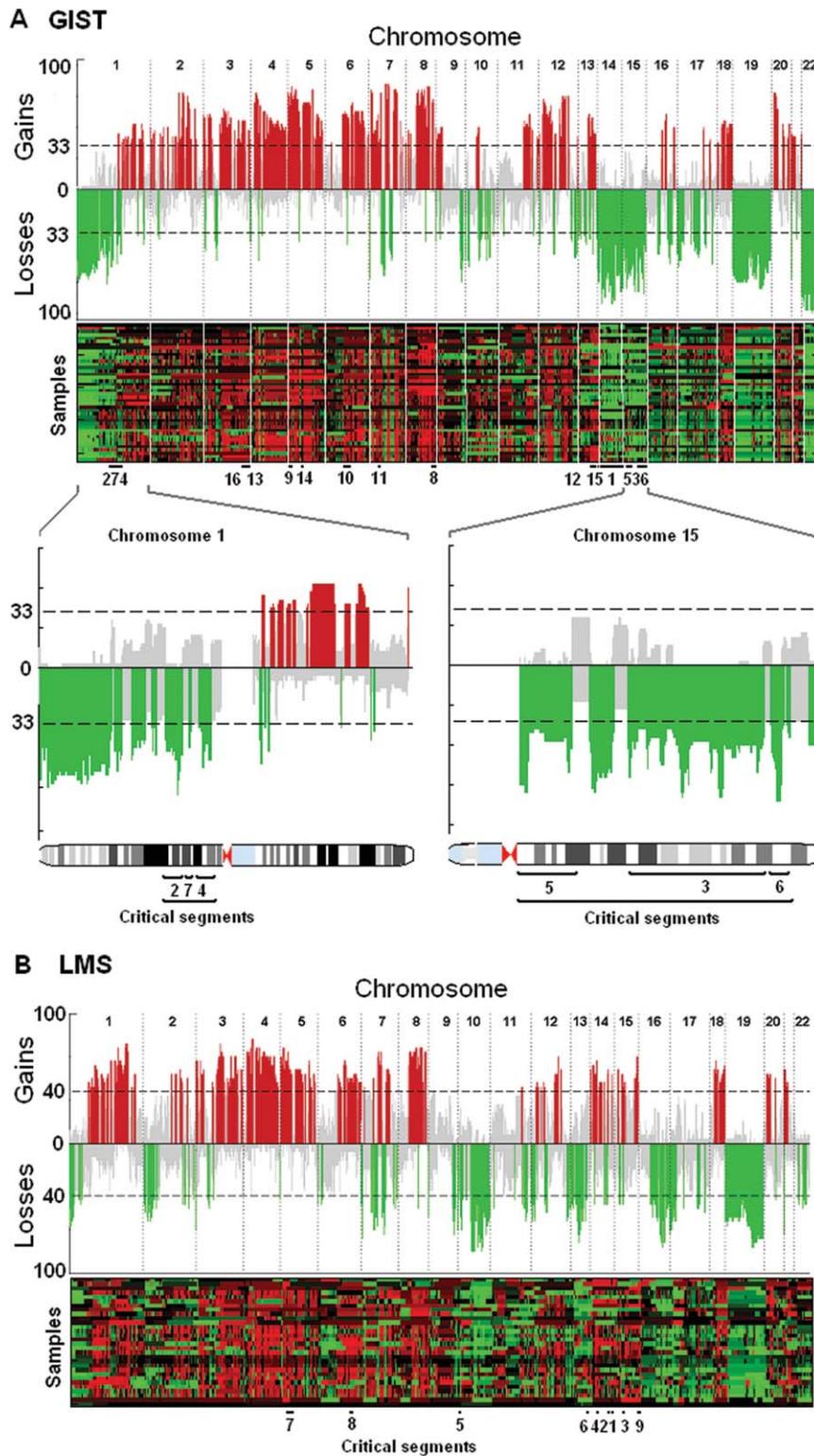
Figure 1. These are skyline plots and heat maps of signature gene copy number aberrations for gastrointestinal stromal tumors (GISTs) and leiomyosarcomas (LMSs). The recurrence of copy number aberrations throughout the genomes is shown both in skyline plots and heat maps for (A) GIST and (B) LMS. The value on the y-axis of each skyline plot is the percentage of patients who had tumors that had gains (positive axis; red) or losses (negative axis; green) in corresponding genomic loci. The probes are aligned evenly in chromosome order on the x-axis. The dashed line indicates the threshold for a significant number of patients whose tumors share the same aberration (14 patients for GIST). The significance threshold is computed with the use of a permutation test. Recurrence rates that exceed this threshold are deemed significantly recurrent and are color-coded to emphasize the locations. Gray color represents a nonsignificant amount of aberration in the locus. The lower plots in A highlight chromosomes 1 and 15, both of which contain 3 critical segments. Critical segments are contiguous genomic regions of recurrent aberration that harbor at least 1 dosage-sensitive, survival-affecting gene. Details of these segments are listed in Table 2. Losses in 1p, 14q, 15q, and 22q are markedly more common in GIST; whereas losses in chromosomes 10 and 13 are more common in LMS. Gains in chromosomes 1, 14q, and 15q are defining features of LMS.

Table 2. Copy Number Aberrations That Harbored at Least 1 Dosage-Sensitive Gene Associated With a Poor Prognosis in Patients With Gastrointestinal Stromal Tumors and Leiomyosarcomas

		Locus Boundary			Cytogenetic Band		
No.	Chromosome	Start, bp	Stop, bp	PAR, %[a]	Start	Stop	Candidate Genes
GIST losses							
1	14	19,365,081	106,329,840	65	q11.2	q32.33	OXA1L, PPP2R5E
2	1	84,129,665	95,422,234	53	p31.1	p21.3	RWDD3
3	15	48,206,991	84,693,976	52	q21.2	q25.3	AKAP13, C15orf5, GLCE, MESDC2, MTHFS, SENP8, SUHW4, USP3
4	1	106,334,457	114,964,386	49	p21.1	p13.2	AMPD1, AP4B1, BCAS2
5	15	19,109,154	33,140,043	48	q11.2	q14	APBA2, ARHGAP11A, KLF13
6	15	96,783,935	99,643,192	46	q26.3	q26.3	IGF1R
7	1	99,947,682	100,601,821	46	p21.2	p21.2	AGL, HIAT1
GIST gains							
8	8	131,377,346	140,782,724	67	q24.21	q24.3	DDEF1
9	5	1,492,821	16,643,214	65	p15.33	p15.1	DNAH5, FBXL7
10	6	44,466,155	73,840,214	54	p21.1	q13	HMGCLL1
11	7	36,220,617	39,454,216	51	p14.2	p14.1	VPS41
12	13	52,523,853	72,250,500	49	q21.1	q22.1	TDRD3
13	3	157,321,066	170,967,793	47	q25.31	q26.2	CCNL1, VEPH1
14	5	71,052,382	72,832,629	46	q13.2	q13.2	MAP1B
15	13	73,160,881	93,888,882	46	q22.1	q32.1	C13orf7
16	3	143,995,402	156,589,223	41	q23	q25.31	P2RY1, SMARCA3
LMS losses							
1	14	76,889,274	77,394,297	46	q24.3	q24.3	C14orf133, C14orf156
2	14	72,470,216	73,993,089	46	q24.2	q24.3	C14orf168, PSEN1, ZADH1, ZFYVE1, ZNF410
3	15	48,417,864	48,815,590	46	q21.2	q21.2	SPPL2A, USP8
4	14	33,942,779	34,940,779	44	q13.1	q13.2	C14orf10, C14orf24
5	10	5,878,820	6,526,558	43	p15.1	p15.1	FBXO18, PFKFB3, RBM17
6	13	97,404,021	99,434,703	43	q32.2	q32.3	CLYBL, STK24, TM9SF2
LMS gains							
7	5	17,766,514	31,520,799	69	p15.1	p13.3	CDH12
8	6	90,704,743	99,500,862	59	q15	q16.2	FBXL4
9	15	89,596,628	100,168,689	55	q26.1	q26.3	MEF2A

bp Indicates base pair; PAR, probe average recurrence; GIST, gastrointestinal stromal tumor; q, long arm; OXA1L, oxidase (cytochrome c) assembly 1-like; PPP2R5E, protein phosphatase 2, regulatory; p, short arm; RWDD3, ring finger and WD repeat domain containing 3; AKAP13, A kinase (protein kinase A) anchor protein 13; C15orf5, chromosome 15 open reading frame 5; GLCE, D-glucuronyl C5-epimerase; MESDC2, mesoderm development candidate 2; MTHFS, 5,10-methenyltetrahydrofolate synthase (5-formyltetrahydrofolate cyclo-ligase); SENP8, small ubiquitin-like modifier protein/sentrin-specific peptidase family number 8; SUHW4, suppressor of hairy wing homolog 4; USP3, ubiquitin-specific peptidase 3; AMPD1, adenosine monophosphate deaminase 1; AP4B1, adaptor-related protein complex 4, beta 1 subunit; BCAS2, breast carcinoma amplified sequence 2; APBA2, amyloid beta (A4) precursor protein-binding, family A, member 2; ARHGAP11A, rho guanosine triphosphatase-activating protein 11A; KLF13, Kruppel-like factor 13; IGF1R, insulin-like growth factor 1 receptor; AGL, amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase; HIAT1, hippocampus abundant transcript 1; DDEF1, development and differentiation enhancing factor 1; DNAH5, dynein, axonemal, heavy chain 5; FBXL7, F-box and leucine-rich repeat protein 7; HMGCLL1, 3-hydroxymethyl-3-methylglutaryl-coenzyme A lyase-like 1; VPS41, vacuolar protein-sorting 41 homolog; TDRD3, tudor domain containing 3; CCNL1, cyclin L1; VEPH1, ventricular zone-expressed PH domain homolog 1; MAP1B, microtubule-associated protein 1B; C13orf7, chromosome 13 open reading frame 7; P2RY1, purinergic receptor P2Y, G-protein coupled; LMS, leiomyosarcoma; SMARCA3, switch/sucrose nonfermentable-related, matrix-associated, actin-dependent regulator of chromatin subfamily A member 3; C14orf133, chromosome 14 open reading frame 133; C14orf156, chromosome 14 open reading frame156; C14orf168, chromosome 14 open reading frame 168; PSEN1, presenilin 1; ZADH1, zinc-binding alcohol dehydrogenase, domain containing 1; ZFYVE1, zinc finger, FYVE domain containing 1 (FYVE indicates the 4 cysteine-rich proteins: Fab1, YOTB, the vesicle transport protein Vac 1, and the early endosome antigen protein EEA1); ZNF410, zinc finger protein 410; SPPL2A, signal peptide peptidase-like 2A; USP8, ubiquitin-specific peptidase 8; C14orf10, chromosome 14 open reading frame 10; C14orf24, chromosome 14 open reading frame 24; FBXO18, F-box protein, helicase, 18; PFKFB3, 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3; RBM17, RNA binding motif protein 17; CLYBL, citrate lyase beta like; STK24, serine/threonine kinase 24; TM9SF2, transmembrane 9 superfamily member 2; CDH12, cadherin 12, type 2 (N-cadherin 2); FBXL4, F-box and leucine-rich repeat protein 4; MEF2A, myocyte enhancer factor 2A.
[a] The PAR is the mean aberration rate of all the probes that belong to that segment.

Genomic Instability Stage May Be a Valuable Prognostic System for Gastrointestinal Stromal Tumors

We performed a cluster analysis in an attempt to identify clinically relevant subgroups that were defined by chromosome-level CNAs. In contrast to LMS, which did not cluster well into clear genomic subtypes, GIST aberration profiles revealed 4 distinct groups with various degrees of genetic alterations (n1 = 12, n2 = 8, n3 = 12, n4 = 10) (Fig. 2A). A survival analysis of these groups revealed that

patient survival is increasingly worse with the presence of more and more genomic aberrations (Fig. 2B). All 4 groups feature partial losses in distal 1p, 19, and 22q, which suggests that these deletions must be early events in GIST development. The defining chromosome-scale difference between Group 1 (with the least amount of aberrations) and Group 2 (with slightly more aberrations than Group 1) is the added deletion of chromosome 14q in Group 2. Patients with tumors classified into Group 1 or 2 have significantly longer survival (Fig. 2C) than patients with Group 3 or Group 4 tumors (which feature more aberrations than the tumors in Groups 1 and 2). Group 3 harbors the same aberrations that characterize Groups 1 and 2 but also have additional deletions of chromosome 15q and the proximal part of chromosome 1p. Tumors in Group 4 are distinguished from tumors in Group 3 by the additional loss of chromosome 10. Although Group 4 retains the characteristics of the first 3 groups, it also contains a more diverse set of tumors, which is apparent in the more heterogeneous pattern of CNAs compared with the other 3 groups. This also is reflected in the survival estimate, which falls between the first 2 groups and the third group.

These results lead us to propose a new tumor-progression genetic staging system termed genomic instability stage (GIS) to complement the current prognostic staging system for GIST based on tumor size, MI, and *KIT* mutation. Although we did not have MI information for all patients in the study, we did have sequencing data on *KIT* and *PDGFRA* gene mutations (Fig. 2A). This allowed us to investigate the relation between the mutation of these genes, especially *KIT*, and genomic instability manifested by the accumulation of CNAs. The high mutation rate of *KIT* exon 11 in Groups 1 and 2 suggested that *KIT* mutation is an early event in GIST, which is consistent with its role as a driver oncogene. Increased *KIT* mutation frequency was observed in Groups 3 and 4, and this observation was consistent with reports that secondary mutations of *KIT* occur at later stages in GIST progression.[6] Imatinib-treated patients who had mutations in *KIT* exon 11 survived significantly longer (in a group-independent manner) than patients who had the same mutation but did not receive imatinib ($P = .002$) (Fig. 2D), suggesting that the differences in genomic survival estimators were not affected significantly by either imatinib treatment or *KIT* mutations. Furthermore, other common risk-assessment and clinical parameters, such as patient age, tumor size, sex, primary site, and the presence of metastases, were not correlated significantly with the groups. Because of

the lack of data on MI, we were unable to fully compare the existing risk-assessment system[4] with the genomic stages. Further prospective characterization of these genomic profiles, coupled with full risk assessment (based on 2007 National Comprehensive Cancer Network guidelines) and clinical outcome, is needed to validate this proposed model.

The incremental occurrence of the observed CNA patterns, increasing *KIT* mutations, and independence from clinical parameters other than survival suggested that the 4 GIS stages reflect the progressive accumulation of chromosome-scale genetic abnormalities during GIST progression. We further confirmed the sequential nature of the chromosome-scale events by determining which aberrations are the most prominent in each stage. That analysis clearly revealed that losses of 1p, 14q, 15q, 19, and 22q are the most distinct events in the 4 groups, although many smaller scale events also may play a critical role in GIST progression (Fig. 2E). Notably, the prominence of less aberrated chromosomes also increased from the first stage to the third, as observed in the amplification of chromosomes 3, 4, 5, and 6. In addition, the analysis revealed that dosage-sensitive genes in critical segments also changed their expression in a corresponding manner between the hypothesized GIS stages (Fig. 3A-D). From the gene expression rates, we could clearly observe copy number changes in the chromosomes that harbored the genes, such as loss of 15q for the anchor protein 13 (*AKAP13*) gene and the chromosome 15 open reading frame 5 (*C15orf5*) gene, loss of 14q for the oxidase (cytochrome c) assembly 1-like (*OXA1L*) gene, and gains in 3q for the switch/sucrose nonfermentable-related, matrix-associated, actin-dependent regulator of chromatin subfamily A member 3 (*SMARCA3*) gene. Because these are genes that significantly affect survival, these possible target genes also ultimately may be responsible for the worse outcome observed for patients in Groups 3 and 4.

Different Cellular Pathways Are Altered in Gastrointestinal Stromal Tumors With Different Genomic Instability Stages

The genes that were expressed differently between 2 adjacent GIS groups were used in an enrichment analysis with the objective of finding the biologic processes that were altered significantly during the progression from 1 stage to the next. We used the list of biologic processes in the Gene Ontology database[22] as our reference. Differences in genome level translate into several distinct cancer-related processes at the transcriptome level (Fig. 4). The

changes from GIS1 to GIS2 impaired mainly the apoptotic, DNA-repair, and damage-response pathways; whereas the progression from GIS2 to GIS3 affected the mitotic, cell cycle, and growth pathways. The final transition from GIS3 to GIS4 had substantially more differences in gene expression, most notably in the cell-cell adhesion and chromosomal organization pathways.

DISCUSSION

Recent progress in cancer genomics, highlighted by the advancement of the Cancer Genome Atlas program, has demonstrated that comprehensive genomic characterization of a large number of cancer samples is highly valuable for fully understanding the molecular basis of human cancer and for classifying cancer into clinically meaningful
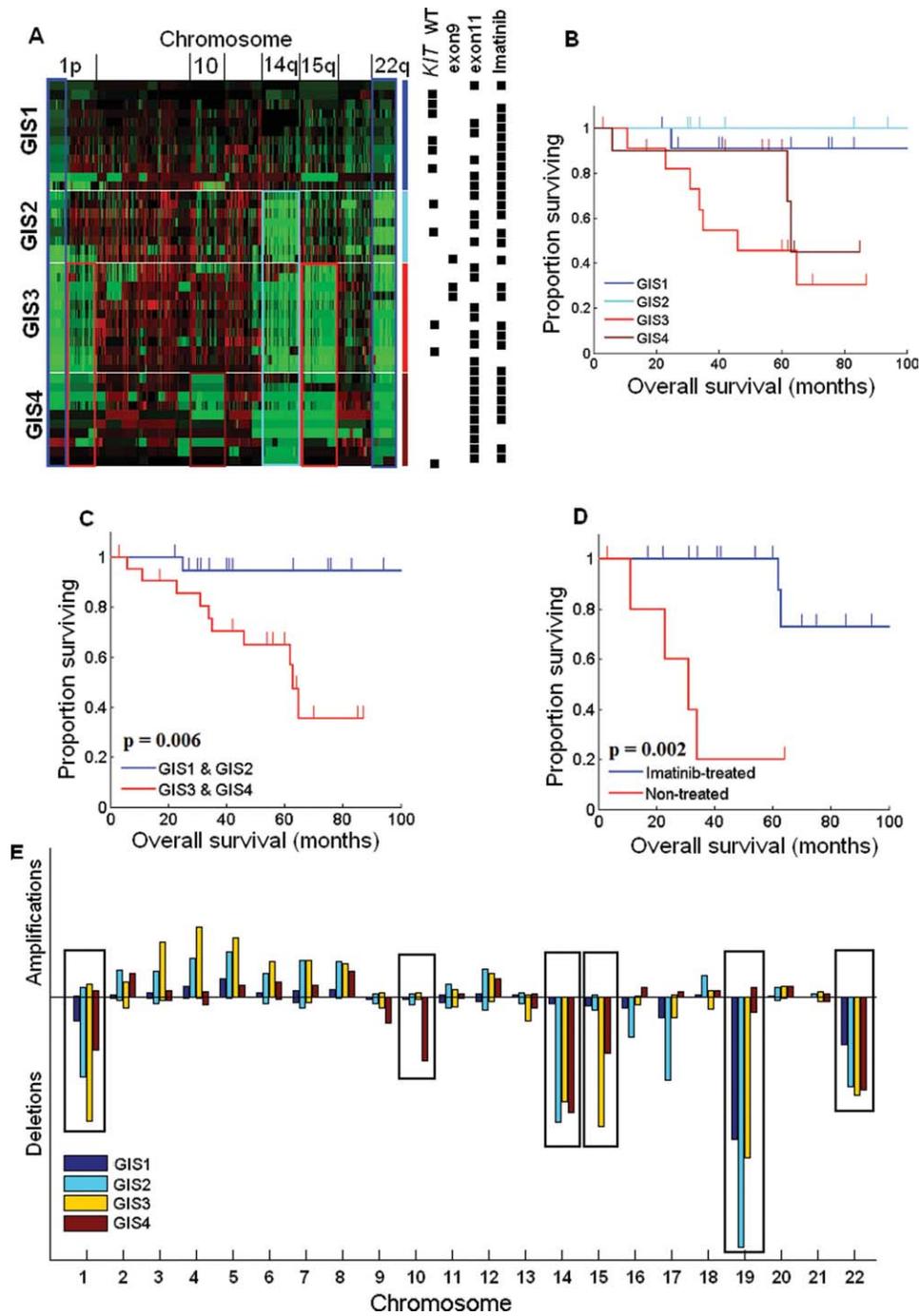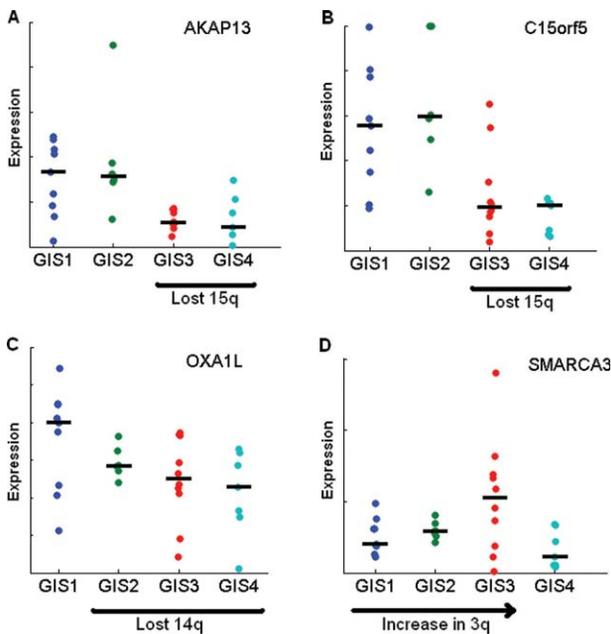

Figure 2.

Figure 3. Sequential chromosomal aberrations were correlated with gene expression in survival-critical genes. (A,B) The stepwise loss of chromosome 15q, where the A kinase (protein kinase A) anchor protein 13 gene (*AKAP13*) and the chromosome 15 open reading frame 5 gene (*C15orf5*) genes are located, also is reflected in the lower median gene expression in genomic instability stage 3 (GIS3) and GIS4. (C) The oxidase (cytochrome c) assembly 1-like gene (*OXA1L*) exhibits a clear gene-dosage effect according to loss of chromosome 14q in GIS2 through GIS4. (D) Although amplifications affected gene expression less than deletions, expression of the switch/sucrose nonfermentable-related, matrix-associated, actin-dependent regulator of chromatin subfamily A member 3 (*SMARCA3*) gene has a pattern similar to that of chromosome 3 in Figure 2E. All 4 genes are aberrated recurrently and affect survival significantly.
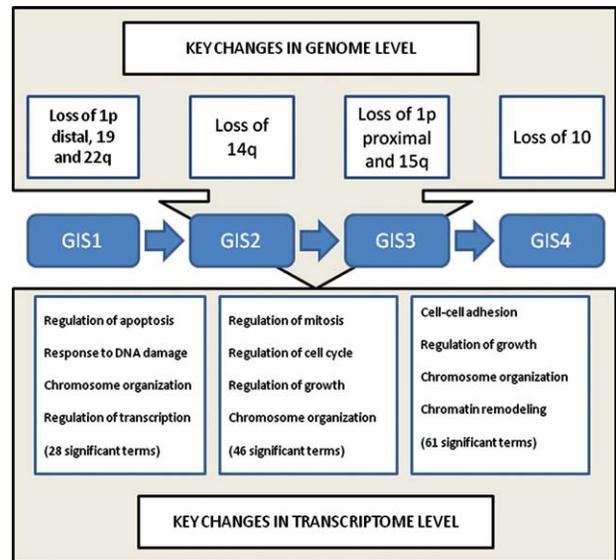

Figure 4. Key changes in genome level confer changes in transcriptome level. The most prominent copy number changes that characterize the 4 stages are shown on the top of the figure. These and smaller scale aberrations convey various tumor-promoting properties to the cell by disrupting key biochemical pathways and biologic processes. The transition from genomic instability stage 1 (GIS1) to GIS2 is characterized by altered antiapoptotic processes and DNA repair as well as chromosomal organization and regulation of transcription. Mitotic, cell cycle, and growth terms were abundant in the transition from GIS2 to GIS3. The list of altered processes in the last transition was longer and more diverse than the previous list, probably because Group 4 was the most heterogeneous. Chromosomal changes, growth, and cell-cell adhesion were among the most relevant cancer-related processes.

Figure 2. Subtypes of gastrointestinal stromal tumors (GISTs) are illustrated according to chromosomal aberrations. (A) Hierarchical clustering of copy number data reveals 4 subtypes of GIST (genomic instability stage 1 [GIS1] through GIS4). Green indicates losses, and red indicates gains. Locations of the key chromosomes are indicated on the top of the heat map. On the right, the 4 groups identified by hierarchical clustering are indicated by blue (GIS1; n = 12), light blue (GIS2; n = 8), red (GIS3; n = 12), and dark red (GIS4; n = 10). The same colors highlight the most distinct aberrations in each group that are visible in the heat map, in which the main characteristics of Group 1 are the losses of distal 1p and 22q; Group 2 has losses of distal 1p, 22q, and an additional loss of chromosome 14q; Group 3 has losses of 1p, 22q, and 14q and also features a loss of 15q; and Group 4, although more heterogeneous, is characterized by losses of 1p, 22q, 14q, and 15q and loss of chromosome 10. Black boxes on the right side of the heat map illustrate some of the known survival-affecting characteristics of individual tumors. WT indicates wild type. (B) Kaplan-Meier survival estimates are shown for each group. (C) Patients with late-stage GIST (GIS3 and GIS4; n = 22) hypothetically have a significantly worse prognosis (*P* = .006) than patients with early stage GIST (GIS1 and GIS2; n = 20). (D) Although v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*) exon 9 and 11 mutations and imatinib treatment significantly affect survival, they are distributed among all 4 groups and, thus, do not affect the survival estimate for any 1 group significantly more than another. (E) This box plot illustrates the prevalence of aberrations for each GIST subtype and chromosome. The height of each box was computed as the square of the amount of aberrant probes divided by the total amount of probes in the chromosome. The plot emphasizes both the relative size of aberrations with respect to the length of the chromosome and the total amount of aberrated loci. This procedure is needed to illustrate the importance of both completely aberrant, small chromosomes and partially aberrant, large chromosomes in the same scale. The plot illustrates the sequential nature of chromosome-scale aberrations in the 4 GIST subtypes, which is revealed best in the highlighted chromosomes and also is visible, for example, in the increased amplification of chromosomes 3, 4, 5, and 6. The overall lower prevalence of aberration in GIS4 is explained by the greater genomic heterogeneity within that group.

subtypes.[23,24] In the current study, we conducted an integrated analysis of high-resolution genomic maps, gene expression data, and clinical information on GISTs and LMSs. Our genomic analysis provided further evidence that GISTs are distinct from LMSs at the genomic level and pointed out the exact chromosomal locations of the greatest difference and similarity. However, it is most noteworthy that our analysis provides a genomic view of GIST progression and demonstrates that staging by using a specific genomic alteration may offer a clinically meaningful system for predicting the prognosis for patients with GIST, even in those who receive imatinib therapy.

Although several previous studies have profiled genomic alterations in GISTs and LMSs using different generations of technologies and relatively small sample cohorts, a key aspect of the current analysis is the correlation of genomic alterations with gene expression data and clinical information. By using this integrative approach, we were able to pinpoint clinically relevant CNAs from the vast number of biologically irrelevant aberrations. Whereas simple mapping of recurrently aberrant genes can yield hundreds or thousands of clinically irrelevant passenger genes, the clinically relevant genomic segments (critical segments) that we have uncovered provide a reasonable number of putative targets for future validation studies.

Our integrated analysis also led us to a new appreciation for the genetic basis of the progression of GISTs. Pattern recognition analysis of the genomic alterations revealed that there is an obvious incremental accumulation of gene copy number alterations in GIST. Consequently, we have proposed a new tumor-progression genetic staging system (Genomic Instability Staging or GIS) to complement the standard tumor site, size, and proliferation risk-assessment system.[4] According to the GIS staging system, deletions of distal 1p, 19, and 22q are the likely keys to early chromosome-scale events that may have triggered the transformation from normal tissue to GIS1 tumor. Whether these events occur before or after KIT mutation is not apparent from our data, because KIT mutation is a high-frequency event in every GIS stage. The most distinct event that follows these deletions is the deletion of 14q, which can be observed clearly as the defining feature in GIS2. Further key deletions of proximal 1p and 15q mark GIS3 disease. Loss of chromosome 10, which also has been associated with late stage in many solid tumors,[25] defines the final stage, GIS4. Our GIS groups are consistent with previously reported data[21,26] but provide more specific information on the key aberrant

events. The lack of significant differences in KIT and PDGFRA mutation status and in the response to imatinib for different GIS groups indicates that the GIS system may have independent prognostic value for patients with GISTs.

Our pathway analysis provides additional insight into the process of tumorigenesis in which early stage GISTs (GIS1 and GIS2) evade apoptosis, intermediate-stage GISTs (GIS2 and GIS3) undergo accelerated proliferation, and late-stage GISTs (GIS3 and GIS4) lose their dependence on cell adhesion, allowing invasion and metastasis. These different key pathway aberrations in different GIS groups validate the accumulative progressive character of GISTs. We believe that these findings are compelling; however, functionally confirming them would require a much larger study. We also must point out that, although we did not observe similar findings for LMSs, this may mean only that LMS is a more heterogeneous disease, and a larger sample size would be needed to reveal key signatures that underlie disease progression and prognosis in patients with LMS.

REFERENCES

1. Price ND, Trent J, El-Naggar AK, et al. Highly accurate 2-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proc Natl Acad Sci U S A.* 2007;104:3414-3419.
2. Heinrich MC, Corless CL, Duensing A, et al. PDGFRA activating mutations in gastrointestinal stromal tumors. *Science.* 2003;299:708-710.
3. Demetri GD, Benjamin RS, Blanke CD, et al. NCCN Task Force report: management of patients with gastrointestinal stromal tumor (GIST)—update of the NCCN clinical practice guidelines. *J Natl Compr Canc Netw.* 2007;5(suppl 2):S1-S29.
4. Joensuu H. Gastrointestinal stromal tumor (GIST). *Ann Oncol.* 2006;17(suppl 10):x280-x286.
5. Hensley ML, Maki R, Venkatraman E, et al. Gemcitabine and docetaxel in patients with unresectable leiomyosarcoma: results of a phase II trial. *J Clin Oncol.* 2002;20:2824-2831.
6. Chen LL, Trent JC, Wu EF, et al. A missense mutation in KIT kinase domain 1 correlates with imatinib resistance in

gastrointestinal stromal tumors. *Cancer Res.* 2004;64:5913-5919.

7. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell.* 1990;61:759-767.

8. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics.* 2004;5:557-572.

9. van de Wiel MA, Kim KI, Vosse SJ, van Wieringen WN, Wilting SM, Ylstra B. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics.* 2007;23:892-894.

10. Meza-Zepeda LA, Kresse SH, Barragan-Polania AH, et al. Array comparative genomic hybridization reveals distinct DNA copy number differences between gastrointestinal stromal tumors and leiomyosarcomas. *Cancer Res.* 2006;66:8984-8993.

11. Wang R, Lu YJ, Fisher C, Bridge JA, Shipley J. Characterization of chromosome aberrations associated with soft-tissue leiomyosarcomas by twenty-4-color karyotyping and comparative genomic hybridization analysis. *Genes Chromosomes Cancer.* 2001;31:54-64.

12. El-Rifai W, Sarlomo-Rikala M, Knuutila S, Miettinen M. DNA copy number changes in development and progression in leiomyosarcomas of soft tissues. *Am J Pathol.* 1998;153:985-990.

13. Otano-Joos M, Mechtersheimer G, Ohl S, et al. Detection of chromosomal imbalances in leiomyosarcoma by comparative genomic hybridization and interphase cytogenetics. *Cytogenet Cell Genet.* 2000;90:86-92.

14. Riva P, Dalpra L, Gualandri V, et al. 19p Deletion in recurring leiomyosarcoma lesions from the same patient. *Cancer Genet Cytogenet.* 2000;119:102-108.

15. Derre J, Lagace R, Nicolas A, et al. Leiomyosarcomas and most malignant fibrous histiocytomas share very similar comparative genomic hybridization imbalances: an analysis of a series of 27 leiomyosarcomas. *Lab Invest.* 2001;81:211-215.

16. Mandahl N, Fletcher CD, Dal Cin P, et al. Comparative cytogenetic study of spindle cell and pleomorphic leiomyosarcomas of soft tissues: a report from the CHAMP Study Group. *Cancer Genet Cytogenet.* 2000;116:66-73.

17. Yang J, Du X, Chen K, et al. Genetic aberrations in soft tissue leiomyosarcoma. *Cancer Lett.* 2009;275:1-8.

18. Wang R, Titley JC, Lu YJ, et al. Loss of 13q14-q21 and gain of 5p14-pter in the progression of leiomyosarcoma. *Mod Pathol.* 2003;16:778-785.

19. Hu J, Rao UN, Jasani S, Khanna V, Yaw K, Surti U. Loss of DNA copy number of 10q is associated with aggressive behavior of leiomyosarcomas: a comparative genomic hybridization study. *Cancer Genet Cytogenet.* 2005;161:20-27.

20. Larramendy ML, Kaur S, Svarvar C, Bohling T, Knuutila S. Gene copy number profiling of soft-tissue leiomyosarcomas by array-comparative genomic hybridization. *Cancer Genet Cytogenet.* 2006;169:94-101.

21. Yang J, Du X, Lazar AJ, et al. Genetic aberrations of gastrointestinal stromal tumor. *Cancer.* 2008;113:1532-1543.

22. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25:25-29.

23. Verhaak RGW, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell.* 2010;17:98-110.

24. Purow B, Shiff D. Advances in the genetics of glioblastoma: are we reaching the critical mass? *Nat Rev Neurol.* 2009;5:419-426.

25. Cavenee WK, Scrable HJ, James CD. Molecular genetics of human cancer predisposition and progression. *Mutat Res.* 1991;247:199-202.

26. El-Rifai W, Sarlomo-Rikala M, Andersson LC, Knuutila S, Miettinen M. DNA sequence copy number changes in gastrointestinal stromal tumors: tumor progression and prognostic significance. *Cancer Res.* 2000;60:3899-3903.